

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN  
OPERATIVA

ALGUNAS APLICACIONES DE LA GEOMETRÍA  
ESTOCÁSTICA EN PROCESO DIGITAL DE IMÁGENES

M<sup>a</sup> VICTORIA ZAPATER VERDUCH

UNIVERSITAT DE VALENCIA  
Servei de Publicacions  
2003

Aquesta Tesi Doctoral va ser presentada a València el dia 15 de Juliol de 2003 davant un tribunal format per:

- D. Francisco Montes Suay
- D<sup>a</sup>. María Esther de Ves Cuenca
- D<sup>a</sup>. María Ángeles López Malo
- D<sup>a</sup>. Amelia Simó Vidal
- D<sup>a</sup>. Pilar Marco Ventura

Va ser dirigida per:

D. Guillermo Ayala Gallego

©Copyright: Servei de Publicacions  
M<sup>a</sup> Victoria Zapater Verduch

---

Depòsit legal:

I.S.B.N.:84-370-5803-1

Edita: Universitat de València  
Servei de Publicacions  
C/ Artes Gráficas, 13 bajo  
46010 València  
Spain  
Telèfon: 963864115



UNIVERSITAT DE VALÈNCIA

---

# Algunas aplicaciones de la Geometría Estocástica en Proceso Digital de Imágenes

Tesis doctoral  
Departamento de Estadística e I.O.

María Victoria Zapater Verduch



**UNIVERSITAT DE VALÈNCIA**

---

# Algunas aplicaciones de la Geometría Estocástica en Proceso Digital de Imágenes

Memoria presentada para optar al grado de doctor en Matemáticas en  
el Departamento de Estadística e Investigación Operativa  
Facultad de Ciencias Matemáticas  
Universitat de València

María Victoria Zapater Verduch

Director  
Guillermo Ayala Gallego

D. Guillermo Ayala Gallego, Profesor Titular de Estadística e Investigación Operativa de la Universitat de València.

CERTIFICA: que la presente memoria, *Algunas aplicaciones de la Geometría Estocástica en Proceso Digital de Imágenes*, ha sido realizada bajo mi dirección, en el Departamento de Estadística e Investigación Operativa por Dña. M<sup>a</sup> Victoria Zapater Verduch, y constituye su tesis para optar al grado de Doctor en Matemáticas.

Y para que conste, en cumplimiento de la legislación vigente, presento ante la Facultad de Matemáticas de la Universitat de València, a 31 de marzo de 2003.

El director,

Guillermo Ayala Gallego

A mis padres y hermano

## Prólogo

La información en forma de imagen es muy frecuente en muchos campos de la ciencia. En Medicina es particularmente importante debido a que muchos diagnósticos se realizan en base a imágenes. Notar la gran cantidad de técnicas de obtención de imagen existentes en este campo. Cada vez hay más tipos de imágenes médicas y éstas son mejores. El experto en Medicina examina las imágenes y toma decisiones.

El Proceso de Imagen tiene como objetivo la mejora de la imagen eliminando las características que impiden su interpretación o bien realzando aquellas otras que la facilitan. El paso siguiente es la interpretación por medio de una descripción. Sería bueno que la Medicina, y en general cualquier otro entorno donde las imágenes cumplan un papel informativo, hiciera uso de las técnicas del proceso de imagen con el fin de asistir en la toma de decisiones.

Cualquier técnica, independientemente de su naturaleza, que consiga el objetivo anterior es una técnica de proceso de imagen. Es decir, se trata de un campo amplio donde caben puntos de vista muy distintos sobre un mismo problema.

En este trabajo se han empleado técnicas de proceso de imagen aplicadas a dos tipos de imágenes médicas, más específicamente imágenes oftalmológicas.

El primer tipo de imagen procesada es de endotelio corneal humano. El endotelio corneal es una de las capas que integra la córnea. En un endotelio sano, las células que lo forman cumplen ciertas condiciones de regularidad en área y forma. En el capítulo 2 es analizada una muestra de endotelios corneales en base a su descripción granulométrica. La granulometría es una potente herramienta descubierta por Matheron [30]. Se adoptan dos enfoques: uno global y otro local. En el enfoque global, el endotelio constituye la forma que hay que describir y en el enfoque local, primero se describen las células y después el endotelio. En el enfoque global hay una primera parte en que un oftalmólogo establece una muestra de endotelios sanos o endotelios 'control' y la salud de el resto de endotelios se evalúan comparándolos con ellos. La comparación se realiza en base a la descripción granulométrica y mediante un test gráfico. El establecimiento de endotelios control siempre lleva aparejados los riesgos de la subjetividad y es por eso que también se asume el análisis de la muestra sin establecer tal grupo de referencia realizando un análisis clúster de la muestra total.

En el capítulo 3 se propone una descripción del endotelio corneal humano mediante patrones puntuales. A partir de cada endotelio se obtiene un patrón puntual que contiene un determinado tipo de puntos característicos. En este

caso se aborda el análisis de otra muestra de endotelios corneales sin establecer endotelios de referencia previamente. Para caracterizar a los endotelios se emplean dos tipos de distancias frecuentemente utilizados en el contexto de la Teoría de Procesos Puntuales: la distancia al vecino más próximo y la distancia de punto a suceso. La comparación de endotelios se realiza por medio de la comparación de este tipo de distancias. Para ello se utilizarán algunos tests clásicos de comparación de muestras, sin embargo, se llama la atención sobre la naturaleza censurada de la muestra de las distancias. Este hecho nos permite también utilizar test de comparación de muestras provenientes de la Teoría de la Supervivencia.

En el capítulo 4 se aborda el problema de la segmentación del árbol vascular retiniano en imágenes de fondo de ojo dentro del contexto de la Teoría de Conjuntos Difusos. Se han considerado tres métodos de segmentación de esta estructura y, a partir de ellos, se han generado funciones de pertenencia a vaso en lugar de auténticas segmentaciones. Es decir, el árbol vascular ha pasado a ser un conjunto difuso y lo que se pretende es asociar al difuso un conjunto nítido (*crisp*) que sea representativo, es decir, una segmentación, en definitiva. El problema de asociar un conjunto nítido representativo a un difuso (en inglés *defuzzification*) es un problema muy debatido en el mundo de los difusos y en este capítulo se ha intentado abordar mediante el concepto de promedio de un conjunto difuso. La Teoría de Conjuntos Compactos Aleatorios aporta distintas definiciones de conjunto medio que son directamente aplicables en el contexto difuso.

Por último en el capítulo 1 se introducen los conceptos utilizados a lo largo del trabajo así como el software utilizado.

La realización de este trabajo ha sido posible gracias a la generosidad de un grupo de personas. Sobre todo, y en primer lugar, quiero destacar la de su director, Guillermo Ayala, que ha aportado su esfuerzo, su buen hacer profesional y sus excelentes ideas, junto con su gran calidad humana, a este proyecto. Gracias por todo, y nunca mejor dicho.

Lucía Martínez-Costa ha intervenido a lo largo de todo el trabajo en calidad de oftalmóloga y desde estas líneas quiero agradecerle sus interesantes comentarios clínicos imprescindibles en un trabajo aplicado como intenta ser éste.

Por supuesto, gracias también a Amelia Simó e Irene Epifanio, sin ellas no hubiera sido posible la realización del tercer capítulo de esta memoria.

A Maite León, gracias por su valiosa contribución en la elaboración del cuarto capítulo.

En todo momento he recibido el apoyo, tanto técnico como humano, del



Departamento de Estadística e Investigación Operativa de la Universitat de València al cual quiero mostrar mi profundo agradecimiento. También quiero agradecer el soporte económico y técnico recibido por parte de los proyectos TIC-2002-03494 y CTIDIA-2002-133.

Por último, a mi querida familia, gracias por el amor incondicional, el apoyo y la alegría. A Julio, gracias por la ilusión. A todos ellos, gracias por estar a mi lado.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. Imagen médica . . . . .	1
1.1.1. Imágenes de endotelio corneal . . . . .	2
1.1.2. Imágenes de fondo de ojo . . . . .	4
1.2. Herramientas . . . . .	5
1.2.1. Morfología Matemática . . . . .	7
1.2.2. Conceptos de Procesos Puntuales . . . . .	11
1.2.3. Conceptos del Análisis de Supervivencia . . . . .	13
1.2.4. Conjunto aleatorio y conjunto difuso . . . . .	15
<b>2. Análisis del endotelio corneal humano mediante granulometrías</b>	<b>19</b>
2.1. Material . . . . .	22
2.2. Metodología . . . . .	24
2.3. Análisis global del endotelio corneal basado en distribuciones granulométricas y sus momentos . . . . .	33
2.3.1. Análisis global basado en funciones de distribución granulométrica . . . . .	33
2.3.2. Análisis global basado en momentos granulométricos . . . . .	35
2.3.3. Aplicación clínica: evaluación detallada de algunos casos . . . . .	38
2.3.4. Agrupación de endotelios . . . . .	40
2.4. Análisis local del endotelio corneal basado en distribuciones granulométricas y sus momentos . . . . .	56
2.4.1. Relación de la información granulométrica con la hexagonalidad . . . . .	61
2.4.2. Aplicación clínica: evaluación detallada de algunos casos . . . . .	64
2.5. Conclusiones . . . . .	66

<b>3. Análisis del endotelio corneal mediante procesos puntuales</b>	<b>69</b>
3.1. Descripción de patrones puntuales . . . . .	71
3.2. Diferentes formas de abordar el agrupamiento de procesos puntuales . . . . .	72
3.2.1. Disimilaridad entre patrones puntuales . . . . .	74
3.2.2. Procedimientos cluster . . . . .	76
3.3. Estudio de simulación . . . . .	78
3.3.1. Validez del cluster obtenido . . . . .	78
3.3.2. Experimentos . . . . .	80
3.4. Resultados . . . . .	84
3.5. Agrupamiento de endotelios corneales humanos . . . . .	86
3.6. Conclusiones y comentarios . . . . .	96
<b>4. Segmentación del sistema vascular retiniano humano a partir de funciones de pertenencia</b>	<b>99</b>
4.1. Material . . . . .	100
4.2. Construcción de la función de pertenencia a vaso a partir de algunos métodos de segmentación . . . . .	100
4.3. Promedio de un conjunto difuso . . . . .	103
4.3.1. Promedio de Vorob'ev de un conjunto difuso . . . . .	106
4.3.2. Promedio de Baddeley-Molchanov de un conjunto difuso	109
4.4. Aplicación de la metodología . . . . .	111
4.5. Conclusiones . . . . .	116
<b>A. Agrupaciones de endotelios corneales humanos obtenidos a partir de su descripción granulométrica</b>	<b>127</b>
A.1. Grupos de endotelios corneales humanos de individuos entre 20 y 29 años . . . . .	127
A.2. Grupos de endotelios corneales humanos de individuos entre 30 y 39 años . . . . .	132

# Capítulo 1

## Introducción

El trabajo que presentamos está enmarcado dentro del campo del Proceso de Imagen. Se trata de un campo amplio tanto por la gran cantidad de aplicaciones que se derivan de él como por la multitud de técnicas desarrolladas al respecto.

El proceso o tratamiento de una imagen dada tiene por objetivo la mejora de la imagen en algún sentido o bien la obtención de información sobre algún aspecto concreto del objeto representado en ella. La técnica para conseguirlo siempre ha de estar en concordancia con la naturaleza de dicho objeto y con el tipo de información que se precise.

Este primer capítulo consiste en una presentación preliminar del tipo de información que vamos a utilizar y de las herramientas con las que contaremos para analizarla.

El conjunto de herramientas que utilizaremos para el procesamiento de nuestras imágenes es disperso. Nuestra justificación es sencilla, hemos pretendido en todo momento que se ajustaran al problema que abordábamos. Vamos a utilizar procedimientos de la Morfología Matemática en la parte de proceso de imagen. Para el análisis e interpretación utilizaremos ideas de la Geometría Estocástica y de la Estadística Espacial.

### 1.1. Imagen médica

Hoy en día muchos de los diagnósticos realizados por expertos en distintas ramas de la Medicina están basados exclusiva o parcialmente en la información contenida en algún tipo de imagen. La importancia de este tipo de información se traduce en la gran cantidad de técnicas existentes en la obtención de imagen médica. El uso que se daba a la imagen médica antes de la aparición

de los ordenadores sufrió una revolución con la llegada de éstos. Durante muchos años, la película fotográfica fue el principal medio de almacenamiento de imágenes médicas y la inspección visual el único medio de análisis. En la actualidad se han combinado sofisticados procedimientos de adquisición de imágenes con amplios medios para su manipulación y presentación. La adquisición de buenas imágenes no tendría mucho sentido si el final del proceso acabara simplemente con su visualización. La manipulación eficiente de imágenes da un paso adelante al extraer la información que se considera relevante en cada caso (dependiendo del tipo de imagen y del objetivo que se persiga con su análisis), ofreciendo de esta manera un punto de vista riguroso, general y ajeno a toda subjetividad.

El trabajo presentado en esta memoria se acerca a la manipulación o análisis de dos tipos de imágenes oftalmológicas: imágenes de endotelio corneal e imágenes retinianas de fondo de ojo.

### 1.1.1. Imágenes de endotelio corneal

La córnea es el tejido transparente que cubre la parte frontal del ojo, más concretamente cubre el iris (parte coloreada del ojo) y la pupila (ver figura 1.1). Tiene dos funciones fundamentales: una es la de barrera protectora del ojo contra gérmenes, polvo y otro material dañino y, otra, como lente.

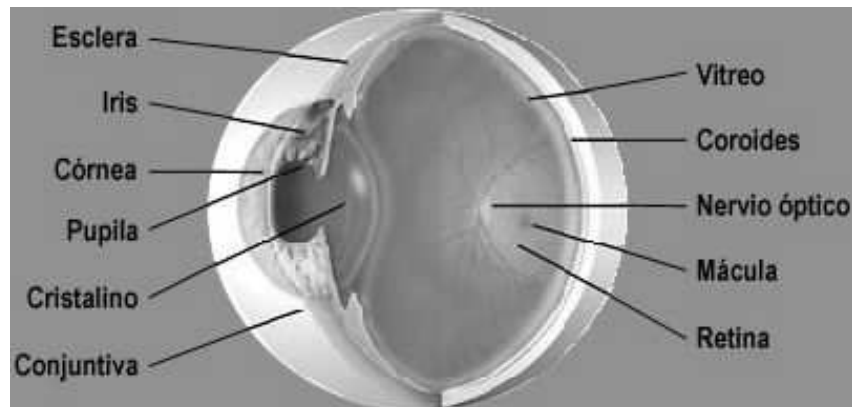
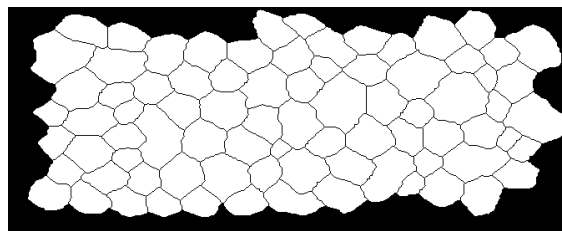


Figura 1.1: Anatomía del ojo

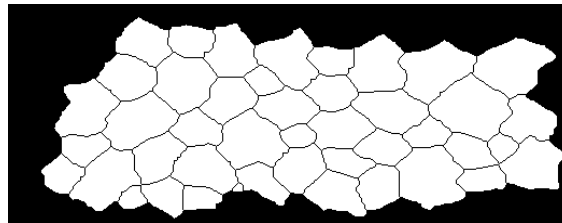
Podemos distinguir cinco capas en la córnea: el epitelio corneal (la capa más superficial), la membrana basal y membrana de Bowman, el estroma corneal,

la membrana de Descemet y el endotelio corneal (la capa más profunda) que se encuentra en contacto con el humor acuoso.

El endotelio corneal está formado por una sola capa de células las cuales tienen tamaño, forma (aproximadamente hexagonal) y apariencia uniformes cuando éste se encuentra en su estado ideal (ver figura 1.2).



(a)



(b)

Figura 1.2: Endotelios corneales pertenecientes a dos personas de 33 años

La barrera que forma el endotelio no es estanca e impermeable sino que está como 'agrietada', permitiendo no un flujo libre de agua y solutos, pero sí un cierto movimiento de intercambio. Dicho intercambio de fluidos sirve para producir la deshidratación de la córnea básica para lograr una transparencia óptima.

El endotelio corneal se ve sometido a una pérdida celular constante como consecuencia del envejecimiento, de la existencia de patologías y, extraordinariamente, a causa de lesiones. Sin embargo, las células endoteliales no regeneran en caso de pérdida, sino que son sustituidas por hipertrofia (aumento de tamaño) y migración de las células vecinas. La pérdida endotelial se manifiesta por el polimegatismo (diversidad de tamaño entre las células) y el pleomorfismo (diversidad de formas) asociados a un incremento de la permeabilidad.

Es conveniente controlar el estado del endotelio corneal en muchas situaciones: antes y después de una intervención quirúrgica en el ojo, después de

un trauma, en caso de uso prolongado de lentes de contacto, cuando existen determinadas patologías en el ojo o en la determinación de las mismas, en los trasplantes de córnea o en la implantación de lentes intraoculares.

El análisis del tejido endotelial puede realizarse en base a imágenes tomadas por un dispositivo llamado *microscopio especular* [45], [22], [8]. Como su nombre indica, el microscopio especular recoge la luz reflejada. La naturaleza de esta técnica permite su aplicación a córneas en vivo sin inflingirles ningún daño. Además es un procedimiento simple y barato por lo cual se ha convertido en la técnica más utilizada en la exploración en vivo del endotelio corneal. La última generación de microscopios proporciona muy buenas imágenes del endotelio corneal central que posteriormente son digitalizadas y analizadas. Por otra parte, los paquetes comerciales (como Imagenet de TOPCON y Bambi de Bio-Optics, Inc., Portland, Oregon) ofrecen un análisis celular muy superficial consistente en la medición de tres parámetros: la densidad (número medio de células por unidad de área), la hexagonalidad (porcentaje de células que están en contacto con otras seis células) y el coeficiente de variación (cociente entre la desviación estándar y la media) de las áreas celulares. Estos son, básicamente, los únicos parámetros a los que se hace referencia en la literatura médica. De ahora en adelante a los tres parámetros anteriores los denominaremos *parámetros clásicos*. El oftalmólogo toma decisiones apoyándose en la inspección visual de la imagen especular y en la valoración de los tres parámetros mencionados.

En este trabajo se propone el proceso de imágenes de endotelios corneales mediante granulometrías en primer lugar y mediante patrones puntuales posteriormente. La descripción granulométrica explica la morfología de las células que componen el endotelio mientras que, la descripción mediante patrones puntuales da información sobre la distribución espacial de las células.

### 1.1.2. Imágenes de fondo de ojo

Las imágenes del fondo del ojo son fotografías en color de la retina obtenidas mediante un dispositivo llamado retinógrafo. Las estructuras normales que se pueden encontrar en una fotografía de este tipo son los *vasos sanguíneos*, el *disco óptico* y la *fóvea*. El buen o mal estado de estas estructuras así como la existencia de otras estructuras 'anormales' son indicativas de algunos tipos de afecciones oculares.

En lo que sigue, el interés se va a centrar en una de las anteriores estructuras: el árbol vascular. Conocer el estado de esta parte de la anatomía ocular sirve para diagnosticar afecciones oculares que, en algunas ocasiones, son manifestaciones de otras afecciones sistémicas. Tal es el caso de la Retinopatía



Diabética (manifestación ocular de la diabetes). Notar que el fondo del ojo es la única localización donde pueden analizarse in vivo de forma incruenta pequeñas arteriolas y vénulas. Además la localización de los vasos da mucha información sobre la localización de otras estructuras como son el nervio óptico o la fóvea. Por último, la detección del sistema vascular sirve como referencia a la hora de alinear distintas imágenes de la retina de un mismo paciente obtenidas en distintos momentos con el fin de medir la evolución de una determinada patología. La estructura vascular puede entenderse, en este sentido, como una especie de huella digital.

En las imágenes de fondo de ojo en color, los vasos sanguíneos son formas curvilíneas de color rojo que pueden bifurcarse o cruzarse y cuyo diámetro decrece con la distancia al nervio óptico. En este tipo de imágenes, los vasos siempre aparecen representados más oscuros que el fondo, es decir tienen una intensidad más baja que el fondo. Las *angiografías fluoresceínicas oculares* son imágenes también del fondo del ojo de mayor calidad que las anteriores.

La figura 1.3 contiene una imagen de fondo de ojo en color.

## 1.2. Herramientas

El éxito logrado por un método de análisis de imagen depende de la muestra de imágenes de partida y de las herramientas utilizadas. A continuación se habla del segundo de los factores y se realiza un pequeño compendio, a modo de introducción, de los conceptos que más tarde manejaremos. Los métodos de análisis que se proponen están fundamentalmente enmarcados dentro de la Morfología Matemática y de la Geometría estocástica.

Una imagen es naturalmente una señal continua, es decir, está definida en un espacio continuo, sin embargo, la mayoría de los análisis son realizados sobre imágenes digitales. El proceso de digitalización de una imagen consiste en el paso del espacio continuo en que está definida a uno discreto, en particular del espacio  $\mathbb{R}^2$  al espacio  $\mathbb{Z}^2$  lo cual se consigue mediante la división de la imagen en pequeñas áreas. Esas áreas son denominadas píxeles (de la abreviatura de la denominación inglesa *pixel* de *picture element*). A cada píxel se le asigna el valor medio registrado en el área a la que representa.

Siempre vamos a considerar imágenes digitales discretas ya que los píxeles sólo van a tomar valores en un conjunto discreto. En el caso de las imágenes a niveles de gris este conjunto es  $\{0, 1, \dots, 255\}$  y en el caso de imágenes binarias los valores posibles están en el conjunto  $\{0, 1\}$ .

Más formalmente, una imagen binaria  $f$  es una aplicación de un subcon-

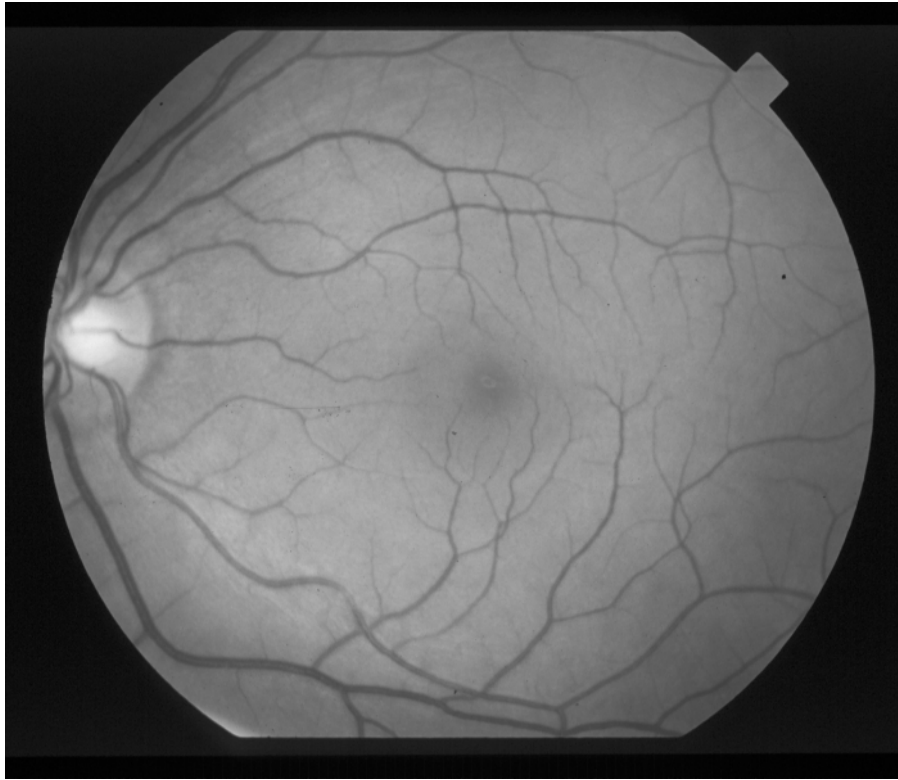


Figura 1.3: Imagen de un fondo de ojo obtenida mediante angiografía fluoresceínica

junto  $\mathcal{D}$  de  $\mathbb{Z}^2$  llamado el dominio de definición de  $f$  en el conjunto  $\{0, 1\}$ :

$$f : \mathcal{D} \subset \mathbb{Z}^2 \longrightarrow \{0, 1\}. \quad (1.1)$$

La definición formal de una imagen a niveles de gris sólo se diferencia de la anterior en que el conjunto en que toma valores la función está formado por más de dos enteros positivos. Más precisamente, una imagen  $f$  a niveles de gris es una aplicación de un subconjunto  $\mathcal{D}$  en  $\mathbb{Z}^2$ , llamado dominio de definición de  $f$ , en un conjunto acotado o una secuencia de enteros no negativos:

$$f : \mathcal{D} \subset \mathbb{Z}^2 \longrightarrow \mathbf{N}_0 = \{0, 1, \dots, t_{max}\}, \quad (1.2)$$

donde  $t_{max}$  es el valor máximo alcanzable dependiendo del tipo de datos que se utilizan para almacenar la imagen (es decir,  $2^n - 1$  si los píxeles tienen una capacidad de  $n$  bits). Las imágenes binarias son casos particulares de imágenes

a niveles de gris. En nuestro caso, el dominio siempre va a ser un subconjunto de  $\mathbb{Z}^2$  y  $t_{max}$  igual a 1 ó bien a 255.

### 1.2.1. Morfología Matemática

La *Morfología Matemática* es una teoría cuyo objetivo es el estudio de estructuras espaciales. Se basa en la teoría de conjuntos y representa una potente técnica de análisis de imagen.

La Morfología Matemática empieza a desarrollarse a mediados de los sesenta gracias a las ideas de G. Matheron [30] en relación con el estudio de la geometría de medios porosos. De esta manera surgió el concepto de *granulometría o distribución de tamaño* basado en transformaciones conjuntistas con las que se determina si una determinada estructura encaja o no en el objeto representado por la imagen.

En el capítulo 2 de esta memoria, se va a hacer uso de herramientas morfológicas aplicadas al estudio de imágenes de endotelios corneales humanos. Recordemos la definición de los operadores morfológicos en los que nos basaremos.

La Morfología Matemática considera los objetos contenidos en una imagen como *conjuntos*. En una imagen binaria la dualidad es obvia: el conjunto que representa al objeto es el formado por todos los píxeles con valor uno.

Las imágenes a niveles de gris también son consideradas como conjuntos a través del *grafo* o el *subgrafo*. El *grafo*  $G$  de una imagen  $f$  es el conjunto de puntos  $(x, t)$  tal que  $x$  pertenece a  $\mathbb{Z}^2$  y  $t = f(x)$ :

$$G(f) = \{(x, t) \in \mathbb{Z}^2 \times N_0 : t = f(x)\}. \quad (1.3)$$

El *subgrafo*,  $SG$ , de una imagen,  $f$ , es el conjunto de puntos de  $\mathbb{Z}^2 \times N_0$  que están por debajo del grafo y por encima del plano donde está definida la imagen:

$$SG(f) = \{(x, t) \in \mathbb{Z}^2 \times N_0 : 0 \leq t \leq f(x)\}. \quad (1.4)$$

Los operadores morfológicos extraen información sobre las estructuras relevantes presentes en la imagen comparándolas con una serie de conjuntos con formas predeterminadas a los que se denomina *elementos estructurantes (SE)*. La forma y el tamaño de un elemento estructurante debe elegirse en función de las propiedades geométricas de los objetos representados en la imagen sobre los que queremos información.

Antes de introducir los operadores morfológicos básicos se van a presentar los operadores de conjuntos en que estos se basan a la vez que es establecida la notación.

Los primeros operadores de conjuntos a que nos referiremos son la *unión*,  $\cup$ , y la *intersección*,  $\cap$ . A nivel funcional, la unión se traduce en el máximo punto a punto,  $\vee$ , y la intersección en el mínimo punto a punto,  $\wedge$ , de las imágenes a niveles de gris dadas por las funciones  $f$  y  $g$ :

$$(f \vee g)(x) = \max[f(x), g(x)], \quad (1.5)$$

$$(f \wedge g)(x) = \min[f(x), g(x)]. \quad (1.6)$$

A nivel conjuntista, estas operaciones pueden definirse en términos de uniones e intersecciones de subgrafos:

$$SG(f \vee g) = SG(f) \cup SG(g), \quad (1.7)$$

$$SG(f \wedge g) = SG(f) \cap SG(g). \quad (1.8)$$

La *complementación* es otro operador básico. El complemento de una imagen  $f$ , denotado por  $f^c$ , se define en cada píxel  $x$  como la diferencia entre  $t_{max}$  y el valor  $f(x)$ :

$$f^c(x) = t_{max} - f(x). \quad (1.9)$$

La *traslación* de una imagen,  $f$ , por un vector,  $b$ , se denota por  $f_b$ . El valor de la imagen trasladada en un píxel  $x$  es igual al valor de la imagen original en la posición  $x - b$ :

$$f_b(x) = f(x - b). \quad (1.10)$$

Todos los operadores morfológicos se basan en la combinación de los operadores intersección (mínimo punto a punto), unión (máximo punto a punto), complementación y traslación.

La *erosión* de un conjunto  $X$  por un elemento estructurante  $B$  es denotado por  $\varepsilon_B(X)$  y se define como el conjunto de puntos,  $x$ , tal que la versión trasladada de  $B$  por  $x$ ,  $B_x$ , está contenida en  $X$ :

$$\varepsilon_B(X) = \{x : B_x \subseteq X\}. \quad (1.11)$$

Otra posible definición del mismo concepto viene dada por la siguiente ecuación:

$$\varepsilon_B(X) = \bigcap_{b \in B} X_{-b}. \quad (1.12)$$

Esta última definición conjuntista puede directamente ser extendida al caso funcional: la erosión de una imagen  $f$  por un elemento estructurante  $B$  es

denotado por  $\varepsilon_B(f)$  y se define como el mínimo de las traslaciones de  $f$  por los vectores  $-b$  de  $B$ :

$$\varepsilon_B(f) = \bigwedge_{b \in B} f_{-b}. \quad (1.13)$$

De esta manera, el valor de la erosión en un píxel  $x$  no es otra cosa que el mínimo de la imagen en la ventana definida por el elemento estructurante trasladado por  $x$ :

$$[\varepsilon_B(f)](x) = \underset{b \in B}{\text{mín}} f(x + b). \quad (1.14)$$

La *dilatación* es el operador dual de la erosión. La dilatación de un conjunto  $X$  por un elemento estructurante  $B$  se denota  $\delta_B(X)$  y se define como el conjunto de puntos  $x$  tal que  $B_x$  interseca con  $X$ :

$$\delta_B(X) = \{x : B_x \cap X \neq \emptyset\}. \quad (1.15)$$

La anterior ecuación puede ser reescrita como la unión de conjuntos trasladados mediante puntos del elemento estructurante:

$$\delta_B(X) = \bigcup_{b \in B} X_{-b}. \quad (1.16)$$

La extensión del concepto a imágenes binarias y a niveles de gris es directa: la dilatación de una imagen a niveles de gris,  $f$ , mediante un elemento estructurante  $B$  se denota por  $\delta_B(f)$  y se define como el máximo de la traslación de  $f$  mediante los vectores  $-b$  de  $B$ :

$$\delta_B(f) = \bigvee_{b \in B} f_{-b}. \quad (1.17)$$

Dicho de otra manera, el valor dilatado en un píxel  $x$  es el valor máximo de la imagen en la ventana definida por el elemento estructurante trasladado por  $x$ :

$$[\delta_B(f)](x) = \underset{b \in B}{\text{máx}} f(x + b). \quad (1.18)$$

La *apertura*  $\gamma$  de una imagen  $f$  por un elemento estructurante  $B$  se denota por  $\gamma_B(f)$  y se define como la erosión de  $f$  por  $B$  seguida de la dilatación por  $\tilde{B}$ , donde  $\tilde{B} = \{-b : b \in B\}$ :

$$\gamma_B(f) = \delta_{\tilde{B}}[\varepsilon_B(f)]. \quad (1.19)$$

Cuando partimos de un conjunto  $X$ , su apertura es la unión de todas las versiones del elemento estructurante trasladadas que caben dentro de  $X$ :

$$\gamma_B(X) = \bigcup \{B_x : B_x \subseteq X, x \in \mathbb{R}^2\}. \quad (1.20)$$

El *cierre* de una imagen  $f$  mediante un elemento estructurante  $B$  se denota por  $\phi_B(f)$  y se define como la dilatación de  $f$  mediante un elemento estructurante  $B$  seguido por la erosión con el elemento estructurante traspuesto  $\check{B}$ :

$$\phi_B(f) = \varepsilon_{\check{B}}[\delta_B(f)]. \quad (1.21)$$

El cierre del conjunto  $X$  se define como:

$$\phi_B(X) = \left[ \bigcup \{B_x : B_x \subseteq X^c, x \in \mathbb{R}^2\} \right]^c, \quad (1.22)$$

o equivalentemente:

$$\phi_B(X) = \bigcap \{B_x^c : X \subseteq B_x^c, x \in \mathbb{R}^2\}. \quad (1.23)$$

Intuitivamente, cuando abrimos un conjunto, lo evaluamos desde dentro eliminando aquellos píxeles del objeto que no pueden ser cubiertos mediante alguna traslación del elemento estructurante. Es decir un conjunto siempre contiene a su apertura. Por el contrario, en el cierre evaluamos el conjunto complementario, añadiendo al conjunto aquellos píxeles del complementario que no están contenidos en su apertura. De esta forma, un conjunto siempre está contenido en su cierre. La anterior dualidad se expresa formalmente de la siguiente manera:

$$\gamma_B(X) = [\phi_B(X^c)]^c. \quad (1.24)$$

También es muy importante dentro de la Morfología Matemática el concepto de *granulometría* o *distribución de tamaño* [20], [38], [40], [48], [42].

Una granulometría se define como sigue: sea  $A$  una forma (un conjunto o una imagen binaria) en  $\mathbb{R}^2$  que queremos describir y sea  $\mathcal{P}(A)$  el conjunto de todos los subconjuntos de  $A$  (o simplemente una familia de subconjuntos). Entonces, una **granulometría** sobre  $\mathcal{P}(A)$  es una familia de transformaciones  $\{\Psi_\lambda : \lambda \geq 0\}$  verificando las siguientes condiciones:

1.  $\Psi_0(B) = B$  para todo  $B \in \mathcal{P}(A)$ , es decir,  $\Psi_0$  es la identidad.
2. Si  $B_1 \subset B_2$  entonces  $\Psi_\lambda(B_1) \subset \Psi_\lambda(B_2)$  para cualquier  $B_1, B_2 \in \mathcal{P}(A)$  y  $\lambda \geq 0$ , es decir,  $\Psi_\lambda$  es una aplicación creciente.
3.  $\Psi_\lambda(B) \subset B$  para cualquier  $B \in \mathcal{P}(A)$  y  $\lambda > 0$ .
4.  $\Psi_{\lambda_1} \Psi_{\lambda_2} = \Psi_{\lambda_2} \Psi_{\lambda_1} = \Psi_{\max\{\lambda_1, \lambda_2\}}$ .

El conjunto  $\Psi_\lambda$  resultante puede medirse de distintas formas dando lugar a tres tipos de curva granulométrica: número de partículas de  $\Psi_\lambda$  versus  $\lambda$ , área de  $\Psi_\lambda$  versus  $\lambda$ , y pérdida de área entre  $\Psi_\lambda$  y  $\Psi_{\lambda+1}$  versus  $\lambda$ .

Todos los cálculos asociados a la parte morfológica de este trabajo han sido realizados con MATLAB. Se trata de un software diseñado por *The MathWorks Inc* (Natick, USA) para la realización de cálculos técnicos. Más específicamente se han utilizado dos librerías: la librería *Image processing* elaborada por *The MathWorks* (<http://www.mathworks.com/products/image>) y la librería *SCD Morphology* elaborada por *SCD Information Systems* (Naperville, USA) (<http://www.mmmorph.com>).

### 1.2.2. Conceptos de Procesos Puntuales

El capítulo 3 enfoca el análisis del endotelio corneal dentro del contexto de la estadística espacial y en especial dentro de la teoría de procesos puntuales. Esta teoría surgió en respuesta a problemas planteados en el ámbito de la física, la biología y la teoría de colas, en los que había que modelizar patrones irregulares de puntos.

Un proceso puntual es un modelo estocástico que produce conjuntos de puntos localmente finitos *con probabilidad uno*, es decir, en cada conjunto de Borel acotado, hay un número finito de puntos (con probabilidad 1). Se denomina *patrón puntual* a una realización de un proceso puntual dado. Un grupo de referencias estándar sobre procesos puntuales es el constituido por [9, 43, 13, 44].

Formalmente expresado, un proceso puntual  $\Psi$  es una aplicación medible de un espacio de probabilidad  $[\Omega, \mathcal{A}, \mathbf{P}]$  en  $[\mathbb{N}, \mathcal{N}]$  donde  $\mathbb{N}$  es el conjunto de todos los conjuntos  $\psi = \{x_1, \dots, x_n\}$  de puntos de  $\mathbb{R}^d$  satisfaciendo las dos siguientes condiciones de regularidad:

1. la secuencia  $\psi$  es localmente finita (es decir, cualquier conjunto acotado de  $\mathbb{R}^d$  debe contener sólo un número finito de puntos de  $\psi$ ),
2. y, además, es simple (es decir, si  $i \neq j$ , entonces  $x_i \neq x_j$ );

y  $\mathcal{N}$  es la  $\sigma$ -álgebra más pequeña sobre  $\mathbb{N}$  de forma que todas las aplicaciones  $\psi \rightarrow \psi(B)$  son medibles, siendo  $B$  un conjunto de Borel acotado.

El análisis estadístico de un proceso puntual se obtiene a partir de una de sus realizaciones, es decir, a partir de un *patrón puntual*.

Denotemos por  $s$ , un patrón puntual observado en una ventana  $W$ , es decir, una realización de  $\Psi$ . Los puntos del patrón puntual,  $s$ , suelen denominarse *sucesos*. Dado  $B$  un subconjunto de Borel del espacio euclídeo correspondiente,

$\Psi(B)$  es el número aleatorio de puntos de  $\Psi$  en  $B$ , y por tanto una variable aleatoria. Por otra parte, cuando escribimos  $x \in \Psi$ , estamos diciendo que el punto  $x$  pertenece al conjunto de puntos aleatorio  $\Psi$ . Si en lugar de hablar del proceso puntual aleatorio pasamos a hablar de una realización del mismo, un patrón puntual  $s$ , entonces  $s(B)$  es una cantidad fija y  $x \in s$  significa que  $x$  es un elemento de  $s$ .

Vamos a asumir que los procesos puntuales que manejamos son estacionarios, es decir, su distribución de probabilidad es invariante por traslación. Los descriptores que a continuación se presentan sólo tienen sentido en este tipo de procesos puntuales.

El objetivo de la estadística espacial es cuantificar la información espacial contenida en el patrón puntual mediante estadísticos descriptivos que informen sobre la configuración de los datos (puntos).

Básicamente, hay dos tipos de estadísticos para la descripción de un patrón puntual: los basados en cuadrantes y los basados en distancias entre sucesos o entre sucesos y puntos aleatorios no pertenecientes al patrón pero pertenecientes a la región de estudio.

Los descriptores basados en cuadrantes cuentan el número de sucesos en los cuadrantes y obtienen estadísticos derivados. Un inconveniente es que no tienen en cuenta la dependencia espacial entre los sucesos y que son muy dependientes del tamaño y la forma del cuadrante. Un descriptor de este tipo es la intensidad. En cualquier proceso puntual estacionario se cumple que  $E\Psi(B) = \lambda\nu(B)$ , donde  $\nu(B)$  es el área de  $B$ , es decir, el número medio de puntos en  $B$  es proporcional a su área. La *intensidad* del proceso se define como el cociente  $E\Psi(B)/\nu(B)$ .

Los *métodos basados en distancias* proporcionan otro tipo de descripción del patrón consistente en el análisis fundamentalmente de dos tipos de distancias: las distancias entre un suceso y su vecino más próximo (distancia al vecino más próximo) o la distancia entre punto y suceso. Vamos a denotar por  $D$  la *distancia al vecino más próximo* que se define como la distancia de un punto del patrón al punto del patrón más próximo. Denotaremos por  $E$  la *distancia de punto a suceso* que es la distancia entre una localización fija arbitraria al punto más cercano del patrón. Los anteriores conjuntos de distancias obtenidas para un patrón puntual dado, son resumidas por medio de sus funciones de distribución correspondientes. La *función de distribución de las distancias al vecino más próximo* se define como

$$G(t) = P(D \leq t) = 1 - P(\Psi(B(0, t)) = 0 : 0 \in \Psi), \quad (1.25)$$

y la *función de distribución de las distancias de punto a suceso* o *función del*



*espacio vacío* se define análogamente como

$$F(t) = P(E \leq t) = 1 - P(\Psi(B(0, t)) = 0). \quad (1.26)$$

Esas funciones de distribución se estiman a partir del patrón puntual observado. Debido a que el patrón puntual se observa dentro de la ventana de muestreo, las distancias que acabamos de introducir no son conocidas en todos los puntos. Cualquier estimador de la funciones  $F$  y  $G$  debe considerar esta distorsión y corregirla para lo cual hay propuestas varias *correcciones de borde*. Los métodos basados en distancias también son muy sensibles al tamaño de la ventana  $W$ .

### 1.2.3. Conceptos del Análisis de Supervivencia

En el capítulo 3 también se aplican determinados conceptos del Análisis de Supervivencia, sobre todo, los relativos a la comparación de muestras de datos censurados. Esta parte de la estadística se dedica al análisis de datos *censurados*, es decir, datos que sólo contienen información parcial sobre la variable de interés ya que el verdadero valor de la variable es inobservable. Las variables censuradas miden típicamente tiempos de fallo de dispositivos o tiempos en que se produce la muerte de una cierta unidad biológica, sin embargo, hay otros muchos tipos de fenómenos que pueden modelizarse de esta manera. Algunos de los contenidos que siguen a continuación así como la notación han sido extraídos de [33].

Hay varios tipos de censura pero nos centraremos en la censura aleatoria. Sean  $X_1, X_2, \dots, X_n$  e  $Y_1, Y_2, \dots, Y_n$  dos conjuntos de variables independientes e idénticamente distribuidas, con funciones de distribución  $F$  y  $G$  respectivamente, donde  $Y_i$  es el tiempo de censura asociado a  $X_i$ . Por tanto, los valores observados son  $(T_1, \delta_1), \dots, (T_n, \delta_n)$  donde

$$T_i = \min(X_i, Y_i) \quad (1.27)$$

y

$$\delta_i = I(X_i \leq Y_i) = \begin{cases} 0 & \text{si } X_i \leq Y_i \text{ (es decir, } X_i \text{ es una observación no censurada)} \\ 1 & \text{si } X_i > Y_i \text{ (es decir, } X_i \text{ es una observación censurada)} \end{cases}$$

Notar que  $T_i, \dots, T_n$  son independientes e idénticamente distribuidas según una función de distribución  $H$ .

El conjunto de riesgo asociado a un valor  $t$ ,  $\mathcal{R}(t)$ , es el conjunto de observaciones en las que se ha superado ese valor (conjunto de unidades que aún no han fallado).

Veamos ahora dos métodos de comparación de muestras censuradas. Partimos de dos muestras de variables aleatorias independientes,  $(X_{11}, Y_{11}), \dots, (X_{1N_1}, Y_{1N_1})$  y  $(X_{21}, Y_{21}), \dots, (X_{2N_2}, Y_{2N_2})$ , formadas por una variable y su variable de censura asociada respectivamente.

El primero de ellos es el test de Mantel-Haenszel [29] o test log-rank. Para construir el test log-rank, para cada valor tiempo de fallo  $t_j$  observado en cualquiera de las muestras, se elabora una tabla  $2 \times 2$  en la que se clasifican los elementos que forman el conjunto de riesgo asociado a  $t_i$ ,  $\mathcal{R}(t_i)$ , según la muestra a la que pertenecen y su estado (continúa funcionando o ha fallado). Denotamos  $r_j$  al tamaño del conjunto de riesgo asociado al intervalo  $(t_{j-1}, t_j]$  cuando consideramos las dos muestras conjuntamente, y  $f_j$  es el número de fallos que se produce en dicho intervalo. Las mismas cantidades pueden calcularse en cada una de las muestras por separado en cuyo caso pasaremos a denotarlas por  $r_{ij}$  y  $f_{ij}$  con  $i = 1, 2$ . El numerador del estadístico log-rank se basa en la suma para  $j$  de las diferencias entre las ocurrencias observadas y esperadas (condicionadas en las marginales de la tabla  $2 \times 2$ ) en el primer grupo:

$$U = \sum_j (f_{1j} - g_{1j}), \quad (1.28)$$

donde  $g_{1j}$  es el número esperado de fallos en el primer grupo,  $g_{1j} = f_j * r_{1j} / r_j$ . El denominador es la suma de las varianzas del número de ocurrencias en el primer grupo dentro de cada tabla  $2 \times 2$ , las cuales se obtienen usando la distribución hipergeométrica:

$$Var(U) = \sum_j r_{1j} * f_j / r_j (1 - f_j / r_j) (r_j - r_{1j}) / (r_j - 1). \quad (1.29)$$

La distribución nula del estadístico log-rank,  $(U^2 / Var(U))$ , es aproximadamente  $\chi^2$  con un grado de libertad. El test log-rank es especialmente sensible a diferencias en la cola derecha de las distribución de supervivencia que comparamos y es más apropiado para la detección de desviaciones cuando ambas distribuciones tienen riesgos proporcionales [26].

El segundo método de comparación de muestras censuradas al que se hará mención está basado en un estadístico de tipo Kolmogorov-Smirnov. Este test es sensible en la detección de diferencias grandes en puntos particulares. Sea  $S_i(t) = P(X_{ij} \geq t)$  la función de supervivencia asociada con  $X_{ij}$  y  $\beta_i(t) = \int_0^t -\frac{d}{ds} \ln S_i(s) ds$  la función de riesgo acumulada. A partir de  $Y_{ij}$  también puede definirse  $C_i(t) = P(Y_{ij} \geq t)$  y  $\alpha_i(t) = -\ln C_i(t)$ . Entonces, sea

$$H(s) = \left( \frac{N_1 \hat{C}_1(s^-) N_2 \hat{C}_2(s^-)}{N_1 \hat{C}_1(s^-) + N_2 \hat{C}_2(s^-)} \right)^{1/2} \frac{1}{2} [\hat{S}_1(s^-)^\alpha + \hat{S}_2(s^-)^\alpha], \quad (1.30)$$

y

$$B^\alpha(t) = \int_0^t H(u) I_{[N_1(u)N_2(u)>0]} d\{\hat{\beta}_1(u) - \hat{\beta}_2(u)\}, \quad (1.31)$$

donde  $N_i(t)$  representa el número de *individuos* en la muestra  $i$  bajo observación en el instante  $t$ , es decir,

$$N_i(t) = \sum_{j=1}^{N_i} I_{[T_{ij} \geq t]} \quad (1.32)$$

donde  $I$  es la función indicatriz y, además, para cualquier función  $f$ ,  $\hat{f}$  denota su estimación y  $f(t^-) = \lim_{s \uparrow t} f(s)$  [16].

El estadístico de tipo Smirnov bilateral de dos muestras se define como

$$K^\alpha = \frac{\sup_{0 \leq t \leq \tau} |B^\alpha(t)|}{\hat{\sigma}_\alpha(\tau)}, \quad (1.33)$$

donde  $\hat{\sigma}_\alpha(\tau)$  es el estimador de  $\sigma_\alpha(\tau)$  (la desviación estándar de  $B^\alpha(\tau)$  siendo  $\tau$  un número positivo fijo). Bajo  $H_0 : S_1(t) = S_2(t)$ , con  $t \in [0, \tau]$  y  $N_1, N_2 \rightarrow \infty$  se tiene que

$$P(K^\alpha \leq y) \rightarrow \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \exp(-\pi^2(2k+1)^2/8y^2) \quad (1.34)$$

En [16] pueden encontrarse las fórmulas para el cálculo así como una explicación más detallada de este test.

Los cálculos relativos a estadística espacial y análisis de supervivencia han sido realizados mediante R, un software gratuito apropiado para la realización de cálculos estadísticos. Se han utilizado las siguientes librerías: *SplanCS*, desarrollado por Barry Rowlingson y Peter Diggle, y transvasada y sostenida por Roger Bivand para el entorno R; *Spatstat*, desarrollada por Adrian Baddeley y Rolf Turner y mantenida en R por Adrian Baddeley; *Cluster*, originalmente diseñada para S por Peter Rousseeuw, Anja Struyf y Mia Hubert, posteriormente transvasada a R por Kurt Hornik y Martin Maechler y mantenida para R por Martin Maechler y *Survival* originalmente diseñada para S por Terry Therneau y transvasada a R por Thomas Lumley. Todo este software puede obtenerse en la página <http://www.R-project.org/>.

#### 1.2.4. Conjunto aleatorio y conjunto difuso

Un problema de gran relevancia en el campo del proceso de imagen es el de la segmentación de objetos de interés en imágenes. En la literatura pueden

encontrarse un gran número de procedimientos de este tipo. En la mayoría de tales procedimientos, se deja en manos del usuario la toma de determinadas decisiones sobre los parámetros, las cuales afectan a la segmentación final. Veremos en el capítulo 4 cómo, intentar eludir ese tipo de decisiones en determinados procedimientos, sobre las que normalmente no tenemos idea previa, nos conduce a una imagen a niveles de gris que puede interpretarse como una función de pertenencia al objeto de interés. Es decir, el objeto de interés pasa a ser un conjunto difuso y lo que pretendemos ahora es obtener una especie de resumen de la función de pertenencia a dicho conjunto. Este problema, recurrente en la literatura de conjunto difusos, de representar el difuso mediante un conjunto nítido (crisp), fue abordado por Ralescu en [36] basándose en la media de conjuntos de Aumann y tomando los  $\alpha$ -cortes como conjuntos a promediar.

En el capítulo 4 va a hacerse uso de los conceptos básicos de la Teoría de Conjuntos Difusos. Esta teoría, definida por Lofti A. Zadeh [50] permite la representación de situaciones ambiguas o imprecisas en relación con clases mal definidas. En la teoría clásica de conjuntos, dados un elemento y un conjunto, sólo hay dos posibilidades, el elemento pertenece al conjunto o no. La teoría de Zadeh es una generalización de lo anterior ya que también permite que un elemento pertenezca a un conjunto en un cierto grado. Mientras que la función característica de conjunto clásico sólo toma los valores 0 ó 1, sobre un conjunto difuso esta función puede todos los posibles valores en el intervalo  $[0, 1]$ .

Un conjunto difuso  $u$  se caracteriza por una función de pertenencia  $u : X \rightarrow [0, 1]$  donde  $X$  es el universo de discurso. Para cada elemento  $x$ ,  $u(x)$  es el grado de pertenencia de  $x$  al conjunto difuso.

Otra forma de caracterizar a un conjunto difuso es mediante sus  $\alpha$ -cortes. Un  $\alpha$ -corte de  $u$ ,  $L_\alpha(u)$ , se define como

$$L_\alpha(u) = \{x \in X : u(x) \geq \alpha\}. \quad (1.35)$$

La aplicación  $Lu : [0, 1] \rightarrow \mathcal{P}(X)$  donde  $Lu(\alpha) = L_\alpha(u)$  es el *mapa de nivel* del conjunto difuso  $u$  en  $X$ .

Como ya se ha comentado, Ralescu definió en [36] la media de un conjunto difuso mediante la integral de conjuntos definida por Aumann. Hay equivalencia entre el nivel medio de un conjunto difuso propuesto en [36] y la media de Aumann del conjunto compacto aleatorio cuyas realizaciones son los  $\alpha$ -cortes del difuso. En lo que sigue va a ser recordada la definición de conjunto compacto aleatorio así como tres definiciones distintas de conjunto medio: conjunto medio de Aumann, Vorob'ev y Baddeley-Molchanov.

En primer lugar, recordemos la definición de conjunto compacto aleatorio.

Se trata de un mecanismo aleatorio que produce subconjuntos compactos de un espacio dado. Fue formalmente definido por Matheron [31]. Sean  $\mathcal{F}$  y  $\mathcal{K}$  las clases de subconjuntos cerrados y compactos, respectivamente, del espacio  $\mathbb{R}^d$ . Consideremos la  $\sigma$ -álgebra dada por  $\sigma_f = \sigma(\mathcal{F}_K)$ , donde  $K \in \mathcal{K}$ , es decir, la  $\sigma$ -álgebra generada por los conjuntos  $\mathcal{F}_K = \{F \in \mathcal{F} : F \cap K \neq \emptyset\}$ . Si  $\mathcal{P}$  denota una medida de probabilidad en  $(\mathcal{F}, \sigma_f)$  entonces  $(\mathcal{F}, \sigma_f, \mathcal{P})$  es un conjunto cerrado aleatorio. Si asumimos que la distribución está concentrada sobre los subconjuntos compactos ( $\mathcal{P}(\mathcal{F} \setminus \mathcal{K}) = 0$ ), entonces tenemos un conjunto compacto aleatorio. Sea  $\Phi$  un conjunto aleatorio compacto en  $\mathbb{R}^d$ , es decir, un conjunto cerrado aleatorio tomando valores en el espacio  $\mathcal{K}$  de todos los subconjuntos compactos de  $\mathbb{R}^d$ .

Vamos a definir en primer lugar la media de Aumann asociada a un conjunto compacto aleatorio en base a la función soporte.

La *función soporte* sobre un conjunto  $C$  en  $\mathbb{R}^d$  es una función definida sobre la circunferencia unidad  $\partial B(0, 1)$  y queda definida por

$$h(C, u) = \sup_{x \in C} \langle u, x \rangle \tag{1.36}$$

donde  $\langle u, x \rangle = u_1 x_1 + \dots + u_d x_d$  es el producto escalar de  $u = (u_1, \dots, u_d) \in \partial B(0, 1)$  y  $x = (x_1, \dots, x_d)$ . La función soporte  $h(C, \cdot)$  es continua en  $\partial B(0, 1)$ .

La función soporte asociada a  $\Phi$ ,  $h(\Phi, \cdot)$ , es un proceso aleatorio definido sobre la circunferencia unidad  $\partial B(0, 1)$ .

La *media de Aumann* de  $\Phi$ ,  $E_a \Phi$ , se define como el conjunto convexo con función soporte dada por la media o el valor esperado de la variable aleatoria  $h(\Phi, v)$

$$h(E_a \Phi, v) = E h(\Phi, v). \tag{1.37}$$

Por tanto, la esperanza o media de Aumann queda determinada por medio de la función soporte esperada de  $\Phi$ . Esta definición es equivalente a la que ya fue implícitamente proporcionada por Aumann en [2].

La *media de Vorob'ev* de  $\Phi$  se obtiene a partir de su *función de cobertura* definida en un punto  $x$  como la probabilidad que tiene ese punto de pertenecer a  $\Phi$ ,  $P(x \in \Phi)$  (ver [44], página 113). Si

$$S_p = \{x \in \Phi : P(x \in \Phi) \geq p\}, \tag{1.38}$$

entonces la media de Vorob'ev,  $E_v \Phi$ , se define como  $S_{p_0}$  donde  $p_0$  es tal que el área de  $S_{p_0}$  es igual al área media de  $\Phi$ . Si hay diferentes valores de  $p$  verificando lo anterior, se tomaría el mínimo de los valores.

Finalmente, Baddeley y Molchanov propusieron un concepto distinto de media de un conjunto aleatorio basado en su representación mediante la función distancia. Los detalles así como un buen número de ejemplos pueden

encontrarse en la referencia original [4]. Sea  $\mathcal{F}'$  el espacio de conjuntos cerrados no vacíos con topología hit-miss [31] y sea  $d : \mathbb{R}^d \times \mathcal{F}' \rightarrow \mathbb{R}$  una **función distancia generalizada**, es decir, una función semicontinua por la izquierda con respecto al primer argumento y medible respecto al segundo de forma que cumple los dos requerimientos siguientes: (i) Si  $F_1 \subset F_2$  entonces  $d(x, F_1) \geq d(x, F_2)$  para todo  $x \in \mathbb{R}^d$ , y (ii)  $F = \{x : d(x, F) \leq 0\}$ . A partir de la familia  $\mathcal{F}'$  surge el espacio  $\mathbb{F} = \{d(\cdot, F) : F \in \mathcal{F}'\}$  de todas las funciones distancia. Sea  $m$  una métrica definida sobre el espacio de funciones distancia,  $\mathbb{F}$ , y  $m_W$  la restricción de  $m$  a  $W$ , un cierto conjunto compacto (la ventana). Una vez introducidos los anteriores elementos, podemos asociar a cualquier conjunto cerrado aleatorio  $\Phi$  la función aleatoria  $d(\cdot, \Phi)$ . Supongamos que  $d(x, \Phi)$  es integrable para todo  $x \in \mathbb{R}^d$ , entonces definimos la función distancia media  $d^*(x) = Ed(x, \Phi)$ . Sea  $\Phi(\epsilon) = \{x \in W : d^*(x) \leq \epsilon\}$  con  $\epsilon \in \mathbb{R}^d$  y sea  $\epsilon^* = \operatorname{argmin}_\epsilon m_W(\Phi(\epsilon), d^*)$ . La *media de Baddeley-Molchanov* de  $\Phi$ ,  $E_{bm}\Phi$ , es el conjunto  $\Phi(\epsilon^*)$ , es decir una umbralización de la función distancia media de  $\Phi$ .

Por último los calculos relacionados con conjuntos difusos y conjuntos aleatorios han sido realizados con MATLAB.

## Capítulo 2

# Análisis del endotelio corneal humano mediante granulometrías

Las imágenes de endotelio corneal, obtenidas mediante microscopía especular son originariamente imágenes a niveles de gris. En un primer paso estas imágenes han de ser digitalizadas (ver figura 2.1 (a)).

Antes de abordar el análisis de las imágenes especulares digitalizadas, éstas son sometidas a un preproceso consistente en una umbralización global, que mantiene las membranas entre células además de algunos artefactos los cuales son eliminados más tarde mediante una esqueletización. La imagen binaria final es el negativo de la imagen binaria obtenida mediante el anterior proceso. En la imagen resultante los píxeles pertenecientes a célula tienen valor uno (blanco) y los que forman parte de la membrana entre células tienen valor cero (negro) (ver figura 2.1 (b)). Una descripción detallada de este preproceso puede encontrarse en [12].

El correcto funcionamiento del endotelio depende de ciertas condiciones de regularidad en el tamaño y la forma de las células que lo forman. Aunque la distribución de tamaños es normal en adultos jóvenes, el pleomorfismo y polimegatismo celulares aumentan y la densidad disminuye con la edad ([49], [34], [6]). Dichos cambios son normalmente simétricos entre los dos ojos. Los cambios asimétricos afectando sólo a un ojo son usualmente signo de patología.

En la literatura existente al respecto, han sido establecidos multitud de datos normativos sobre la densidad y morfología de las células endoteliales [45], [5], [27], [32], [49]. Por ejemplo, se ha comprobado que la densidad celular en endotelios normales de niños de 5 años es de alrededor de  $3500 \text{ células/mm}^2$

([34]). A la edad de 80 años se consideran densidades normales las que se encuentran en el rango 2000-2500 células/ $mm^2$  ([8]). Durante la vida adulta, la densidad celular en el endotelio central decrece en media el 0.6% por año en córneas normales, dándose un incremento gradual del polimegatismo y pleomorfismo ([6]). Parece que la densidad y la morfología celulares están influidos por factores genéticos y raciales [32], [37]. Matsuda et al. ([32]) obtienen una densidad celular de 2977 con una desviación estándar (SD) de 324 en individuos con edades comprendidas entre 20-39 años y un coeficiente de variación (CV) de las áreas de 28.5 con una SD de 5.8. En individuos de 30-39 años, la densidad celular es de 2739 con una SD de 208 y CV medio de 27.4 con SD 3.7. Estos datos corresponden a población americana. En la población japonesa detectaron densidades ligeramente más altas.

La salud del endotelio es mejor si el número de células es alto, su tamaño y forma son similares y su apariencia es aproximadamente hexagonal.

Ya hemos mencionado en la introducción (ver sección 1.1.1) que el estado del endotelio puede deteriorarse como resultado de la existencia de algún trauma o de enfermedades corneales como distrofias o degeneraciones. Por ejemplo, hay edema corneal cuando la densidad celular en el endotelio baja de 500 células/ $mm^2$ . En los casos de baja densidad (menos de 1000 células/ $mm^2$ ) y población de células irregular, la cirugía puede afectar irreversiblemente a la transparencia corneal ya que se destruyen algunas células cuyo hueco es ocupado por el resto de células (hipertrofia y migración). De esta forma se crean diferencias regionales persistentes en el patrón de distribución celular. En un endotelio traumatizado la pérdida de células continúa a un ritmo acelerado y la mal función del endotelio puede dar lugar años más tarde a una enfermedad.

En la sección 2.1 se describe la base de datos de imágenes de la que disponemos y se proporciona una pequeña descripción estadística de los parámetros clásicos en los elementos que la componen.

En las secciones 2.3 y 2.4 se aplica el concepto de granulometría y de distribución granulométrica de tamaño a la descripción del endotelio. El análisis se enfoca desde dos puntos de vista distintos: el global y el local. El análisis global (sección 2.3) comienza comparando endotelios de pacientes con endotelios control en base a la descripción granulométrica para determinar su 'normalidad'. Los endotelios control son determinados por una oftalmóloga. Posteriormente se prescinde de esa clasificación experta y se clasifica la muestra de endotelios disponible en base también a las características granulométricas mediante un análisis cluster jerárquico. Puede obtenerse una descripción granulométrica similar de las células, en ese caso, evaluaremos el endotelio en función del estado de sus células en comparación con un grupo de células control (ver sección 2.4).



Se trata del punto de vista local.

## 2.1. Material

Nuestra base de datos de imágenes de endotelio está formada por 151 imágenes especulares centrales de la córnea en diferentes ojos: 54 son controles (seleccionados por un oftalmólogo en base al historial clínico, a la inspección visual y al valor de los parámetros clásicos) y las 97 imágenes restantes pertenecen a diferentes pacientes. Los indicadores clásicos de los controles están dentro de los datos normativos publicados sobre personas en edad similar (ver la tabla 2.1).

Cuadro 2.1: Media (desviación estándar) de la densidad, hexagonalidad y coeficiente de variación

		Densidad	Hexagonalidad	Coficiente de variación
Controles	20-30 años	2765 (226.9)	65.2 (11.2)	29.9 (6.9)
	31-40 años	2739 (294.4)	62.3 (9.1)	33.2 (7.5)
	Todos	2754 (280.7)	62.6 (11.0)	31.3 (6.4)
Pacientes	20-30 años	2631 (485.1)	57.9 (12.5)	33.6 (6.1)
	31-40 años	2573 (322.9)	48.9 (11.1)	37.1 (8.2)
	Todos	2502 (536.2)	53.3 (14.3)	35.3 (7.4)

Cada imagen corresponde a un ojo distinto y fueron obtenidas con un microscopio especular (TOPCON, Tokio, *Japan*). Una cámara a color CCD (dispositivo de doble carga, XC-711P, Sony, Tokio, Japón) tomó las fotografías de las células bajo una iluminación estandarizada de 1900 lux. La salida de la cámara fue procesada con un digitalizador de vídeo PIP-512/1024 (Matrox Electronic Systems Limited, Quebec, Canada) instalado dentro de un PC. Todas las imágenes fueron obtenidas en la misma escala. Las imágenes originales a niveles de gris fueron preprocesadas como se ha comentado al comienzo del capítulo 2. La figura 2.1 (a) contiene una de las imágenes que conforman la muestra antes de ser procesada. La figura 2.1 (b) muestra la misma imagen una vez ha sido preprocesada.

Al final del preproceso, nuestra información básica es una imagen donde cada célula corresponde a una componente conexas distinta. Es decir, cada imagen preprocesada estará formada por un conjunto de  $N$  células a las que podemos denotar  $A_1, A_2, \dots, A_N$  (ver figura 1.2). El endotelio completo, denotado por  $A$ , resulta de la unión de todas las células  $A_i$ 's ( $A = \bigcup_{i=1}^N A_i$ ) y no cubre totalmente el plano ya que quedan libres los intersticios entre células (fondo negro). Los microscopios comerciales también realizan un análisis

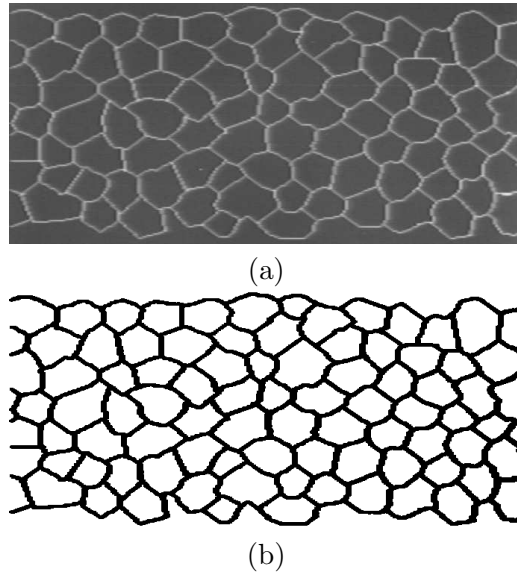


Figura 2.1: Imagen de endotelio tal y como es producida por el microscopio especular, (a), y después del preproceso que la convierte en binaria, (b)

básico de esas imágenes que incluye:

- hexagonalidad,
- densidad y
- un análisis muy superficial de las *áreas celulares*  $\{a_1, \dots, a_N\}$

La salud del endotelio depende del número de células que lo componen, del tamaño de éstas y de su forma. La densidad de células por unidad de área nos está indicando el número de células en el endotelio, la hexagonalidad nos informa sobre la regularidad en la localización de las células y las áreas celulares nos dan idea de su tamaño y hasta qué punto hay variabilidad en el mismo. Esta información es normalmente completada con un histograma de las áreas observadas. Aparte de que la anterior descripción es poco profunda, los resultados no vienen acompañados de algún tipo de intervalo de confianza que pueda servirnos de referencia.

## 2.2. Metodología

Como ya se ha mencionado, el análisis del endotelio se viene realizando en base a la densidad, la hexagonalidad y el coeficiente de variación de las áreas celulares. En primer lugar, pensamos que el análisis de las áreas celulares no proporciona mucha información per se y que no tiene mucho sentido si posteriormente no es comparado con el mismo análisis realizado en un endotelio normal o, mejor dicho, en un grupo de endotelios normales ya que, a parte de la variación en la morfología del endotelio que se produce cuando resulta dañado o con el envejecimiento, hay otro factor que es la variación entre individuos. En segundo lugar, la idea que se obtiene con esos parámetros sobre la morfología del endotelio es muy pobre. En este trabajo se proponen nuevos indicadores más sutiles en la detección de pequeñas diferencias entre endotelios.

Hay dos formas de abordar la evaluación del endotelio: la global y la local. Global, en este contexto, significa que la evaluación no está basada en valores medios de los valores obtenidos en cada célula. En la evaluación local analizamos las células del endotelio aisladamente. En este caso, en primer lugar extraeremos conclusiones sobre las células y, en base a ellas, caracterizaremos el endotelio.

En esta sección vamos a explicar la metodología que hemos seguido a lo largo del capítulo. La herramienta básica de análisis que utilizaremos es la función de distribución granulométrica y algunos de sus momentos. En principio explicaremos la metodología seguida a nivel global. La metodología aplicada a nivel local es fundamentalmente igual salvo algunas variaciones lógicas que iremos mencionando.

Todos los análisis realizados en el trabajo pasan por la comparación de endotelios corneales de pacientes con endotelios corneales considerados normales (controles). Respecto al tema de los endotelios control seleccionados por el experto y que servirán como referencia hay que decir que han sido divididos en dos grupos de edad: los de personas entre 20 y 30 años y los de personas entre 31 y 40 años. El primer grupo lo forman 29 controles y el segundo está integrado por 19. Finalmente para un pequeño número de casos de más de cuarenta años, se ha considerado como grupo de controles la unión de todos los controles anteriores, independientemente de la edad, junto con algún caso de más de 40 años. Los endotelios control pertenecen a pacientes sin patologías cuyos indicadores clásicos están dentro de los rangos establecidos por otros autores [32], [49] y que han superado una inspección cualitativa por parte de un oftalmólogo. En la tabla 2.1 aparecen los valores medios de los indicadores clásicos tanto en el grupo de pacientes como en el de controles y según el

grupo de edad. La fila con el nombre 'todos' contiene el valor medio de los parámetros clásicos en pacientes y en controles sin discriminación por edad.

Una primera comparación que podemos establecer entre el endotelio de un paciente y los endotelios control en su mismo grupo de edad es mediante la distribución empírica de las áreas celulares. Se trata de una descripción funcional de las áreas observadas y se define como sigue: si  $\{a_1, \dots, a_N\}$  son las áreas observadas de las células endoteliales, la función de distribución empírica es

$$F_n(t) = \frac{\#\{a_i : a_i \leq t\}}{n}, \quad (2.1)$$

donde  $\#$  denota el cardinal o el número de elementos en un conjunto. Por lo tanto,  $F_n$  da la proporción de células con un área menor o igual que  $t$ . Las figuras 2.2 (a), (b) y (c) muestran las células del endotelio corneal de la figura 1.2 (a) con área mayor que  $t=1500$  (a), 2000 (b) y 2500 (c). Es importante notar que la función de distribución empírica de las áreas celulares no tienen en cuenta su morfología. Se trata de una función escalonada que considera una célula si su área es menor o igual que un valor  $t$  considerado. Una célula es tenida en cuenta o no dependiendo sólo de su área.

En las figuras 2.3 (a) y 2.4 (a) se muestra la función de distribución empírica de las áreas celulares (línea de puntos) correspondientes a las imágenes en la figura 1.2 (a) y 1.2 (b) respectivamente. Son obviamente distintas, como cabía esperar. Sin embargo compararlas entre sí no sirve de mucho. La cuestión es si las áreas observadas en estos dos casos podrían ser las de una persona sin patología. La única forma de comprobarlo es comparándolas con controles en su mismo grupo de edad. En las figuras 2.3 (a) y 2.4 (a) también aparecen dos funciones correspondientes al máximo y al mínimo de las funciones de distribución empírica de las áreas calculadas para el grupo de controles adecuados en cada caso (línea continua). De ahora en adelante, a esas funciones las llamaremos envolventes superior (máximo de las funciones de distribución empírica de los controles) e inferior (mínimo de las funciones de distribución empírica de los controles). Si la función de distribución empírica de un caso está fuera de estas envolventes, dicho caso será detectado como anormal ya que es diferente a los controles con edad similar. Se trata de un test gráfico simple que completaremos con otras comparaciones.

El análisis que hemos propuesto, sólo válido a nivel global, es simple y fácil de obtener. Sin embargo, se trata de una pequeña sofisticación del análisis del área celular realizado por el software comercial y, como éste, no tiene en cuenta la morfología de las células.

En lo que sigue se va a hacer uso de una herramienta de amplio uso en la Morfología Matemática y que ya ha sido introducida en la sección 1.2.1 : la

granulometría y la distribución de tamaño granulométrica [20], [38], [40], [48].

Según la sección 1.2.1, una granulometría es una familia de transformaciones uniparamétricas,  $\{\Psi_\lambda : \lambda \geq 0\}$ , verificando una serie de condiciones allí enumeradas. Para medir el conjunto resultante de una transformación cualquiera de la familia,  $\Psi_\lambda$ , podemos utilizar su área, por ejemplo. Bajo condiciones de regularidad simple, puede demostrarse que

$$F_A(s) = 1 - \frac{\nu(\Psi_s(A))}{\nu(A)}, \quad (2.2)$$

donde  $\nu(A)$  denota el área del conjunto  $A$ , es una función de distribución de probabilidad a la cual se denomina *función de distribución granulométrica*. La distribución de probabilidad asociada a esta función de distribución se llama *distribución de tamaño granulométrica*.

Vamos a considerar un tipo especial de granulometría basado en el concepto de apertura estructural, o sea, tomamos

$$\Psi_s(A) = A \circ sT. \quad (2.3)$$

La definición de apertura estructural es como sigue: sea  $T$  un conjunto convexo y compacto conteniendo el origen, entonces

$$A \circ T = \bigcup \{T + x : T + x \subset A\}, \quad (2.4)$$

donde  $T + x = \{t + x : t \in T\}$ , es la translación de  $T$  por  $x$ , es la apertura de  $A$  por  $T$ . En el contexto de la Morfología Matemática, a  $T$  se le denomina elemento estructurante. Los elementos estructurantes más frecuentemente utilizados son los que proporcionan los paquetes disponibles sobre morfología matemática (el cuadrado, la cruz, el disco y segmentos en distintas orientaciones).

Por definición, todas las aperturas satisfacen las tres primeras propiedades de las granulometrías. La verificación de la cuarta propiedad o propiedad de *absorción* depende del elemento estructurante utilizado en la apertura. Todos los elementos estructurantes que utilizamos en este trabajo conducen a la verificación de esta última propiedad.

En este caso la imagen binaria  $A$  es la imagen del endotelio de una persona (caso global) o bien una de las células integrantes de un endotelio (caso local). Entonces el conjunto  $A \circ sT$  no es más que la parte del endotelio (célula)  $A$  más grande que  $sT$  y que puede ser recubierta mediante translaciones de dicho elemento estructurante. Pueden encontrarse definiciones más formales de los conceptos de apertura estructural y granulometría en [20], [40], [48], [42].

Con la descripción basada en granulometrías de un objeto, no sólo estamos teniendo en cuenta su tamaño sino también su forma. Básicamente la idea consiste en medir el tamaño y la forma de un objeto comparándolo con otro.

Siendo  $A$  un endotelio, el conjunto  $A \circ sT$  puede ser medido de dos formas. La primera cuenta cuántas células pueden contener una versión trasladada de  $T$  (señalar que estas células tendrán un tamaño como mínimo el tamaño de  $T$ ). Más precisamente, si  $A_1, \dots, A_n$  son las  $n$  células del endotelio  $A$  entonces  $\#\{A_i : A_i \circ sT \neq \emptyset\}$  es la cantidad mencionada, la cual decrece con el tamaño de  $T$ . Podemos considerar como *distribución granulométrica de tamaño en número* la siguiente función,

$$G(s) = 1 - \frac{\#\{A_i : A_i \circ sT \neq \emptyset\}}{n}, \quad (2.5)$$

que es una función de distribución de probabilidad acumulada. Esta forma de medir la apertura no tiene sentido cuando  $A$  es una célula.

La segunda forma de medir el conjunto  $A \circ sT$  que puede considerarse es por medio de su área. La *distribución granulométrica de tamaño en área* se define como

$$H(s) = 1 - \frac{\nu(A \circ sT)}{\nu(A)}, \quad (2.6)$$

donde  $\nu(A)$  denota el área de  $A$ .  $H(s)$  es la proporción del conjunto  $A$  cubierto por el elemento estructurante  $sT$  cuando éste está completamente contenido en  $A$ . En este caso la definición tiene sentido tanto si  $A$  es un endotelio o una simple célula. Las figuras 2.5 (a), (b) y (c) muestran el efecto de las aperturas con discos de radio 14, 18 y 22 píxeles respectivamente sobre el endotelio representado en la figura 1.2(a). Hay que destacar que todas las células pueden formar parte de este análisis, incluidas las de la corona externa de la imagen especular.

Los endotelios corneales (o las células) que son irregulares en forma y tamaño producirán funciones  $F$  y  $G$  típicas. En las figuras 2.3(b) y 2.4(b) (respectivamente 2.3(c) y 2.4(c)) podemos ver las funciones de distribución granulométrica acumulada  $G$  (respectivamente  $H$ ) para las imágenes en la figura 1.2 junto con las envolventes superior e inferior de las funciones de distribución granulométrica acumulada determinada para los controles correspondientes a cada caso. La información proporcionada por  $G$  y  $H$  es similar. Sin embargo  $H$  es más suave que  $G$  y será la que utilizemos de ahora en adelante. Con la función  $G$  podría realizarse un análisis similar también con resultados similares.

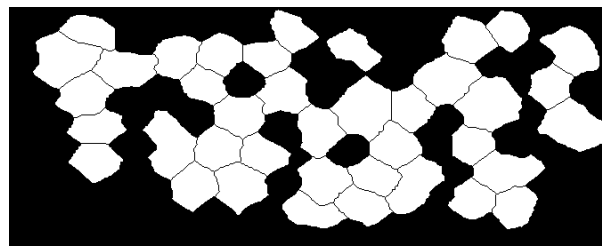
También podemos realizar el análisis utilizando un resumen de las funciones de distribución granulométrica basado en algunos de sus momentos.

Los momentos que utilizaremos son los que normalmente se utilizan en la descripción de distribuciones de probabilidad: la media, la desviación típica, la asimetría y la curtosis. Recordemos cómo se pueden calcular los momentos a partir de las funciones de distribución acumuladas. Asumimos que  $H$  ha sido estimada en el conjunto de puntos discreto  $\{\delta, 2\delta, \dots, k\delta\}$  con  $\delta$  un número positivo no demasiado grande (un múltiplo del tamaño de un píxel). Hemos tomado  $\delta$  igual a 1 píxel, el menor valor posible, y  $k$  igual a 30. A partir de  $H(\delta), \dots, H(k\delta)$  se estima la función de densidad de probabilidad  $h(i\delta) = \frac{H(i\delta) - H((i-1)\delta)}{\delta}$  con  $i = 1, \dots, k - 1$  donde  $H(0) = 0$  para después calcular el  $m$ -ésimo momento como

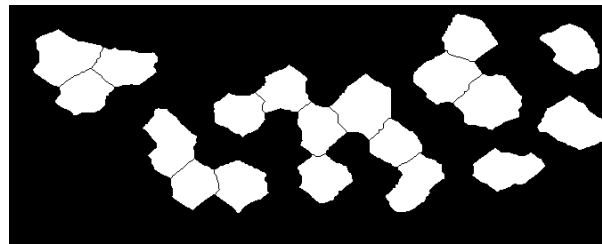
$$\mu_m = \int_0^\infty t^m h(t) dt. \quad (2.7)$$

Los momentos utilizados son la media,  $\mu_1$ ; la desviación estándar,  $\mu_2 - \mu_1^2$ ; la asimetría,  $\mu_3/(\mu_2)^{2/3}$ , y la curtosis,  $\mu_4/\mu_2^2$ . El momento  $m$ -ésimo ha sido estimado a partir de la  $h$  estimada en  $i = 1, \dots, k - 1$  usando la ecuación 2.7 y la regla de Simpson de integración numérica.





(a)

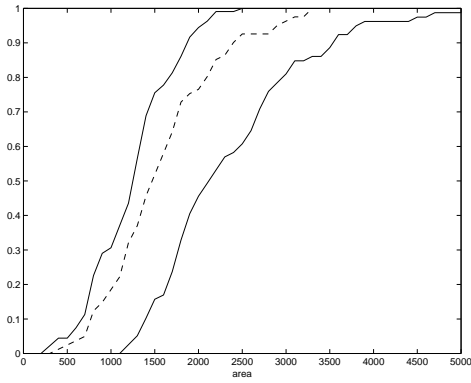


(b)

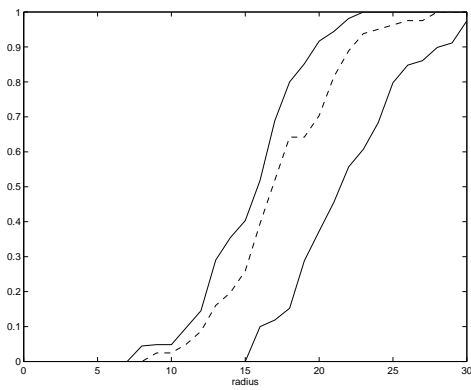


(c)

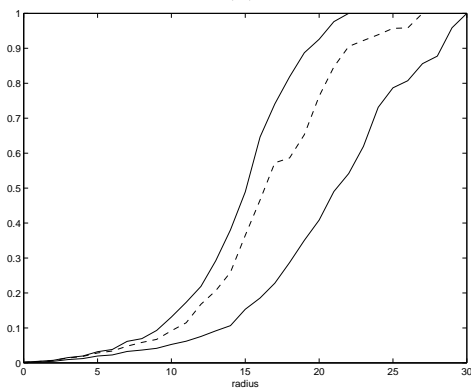
Figura 2.2: Células seleccionadas según su área en el endotelio representado en la figura 1.2 (a): las células con área superior a 1500 píxeles están representadas en (a), las que tienen área superior a 2000 píxeles en (b) y las que tienen área superior a 2500 píxeles aparecen en (c)



(a)

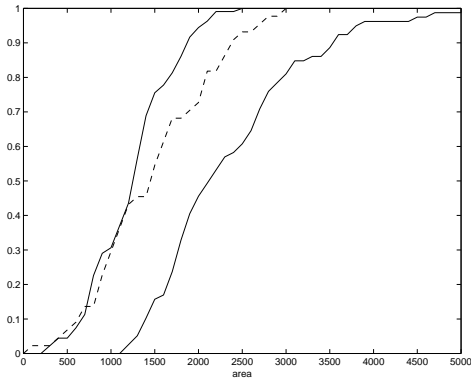


(c)

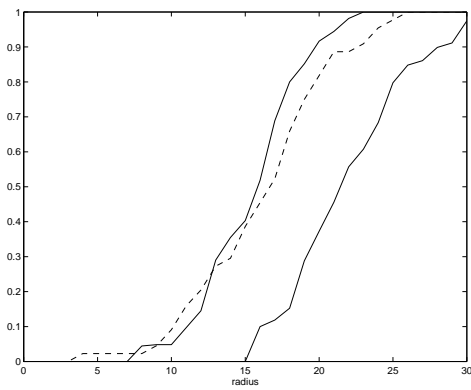


(d)

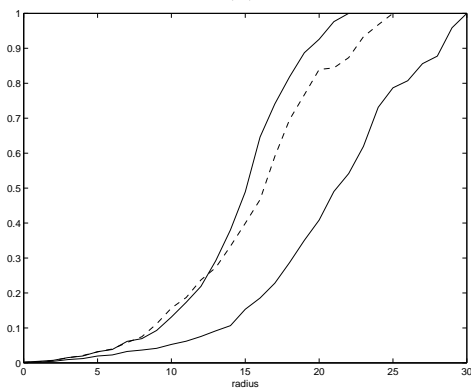
Figura 2.3: Funciones de distribución acumuladas correspondientes a la imagen de la figura 1.2 (a)



(a)

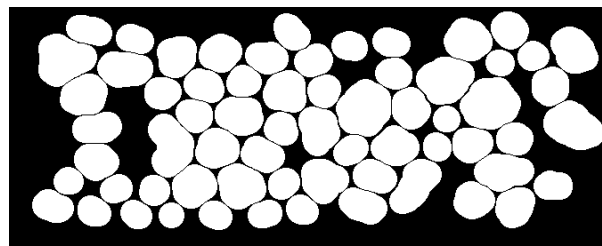


(c)

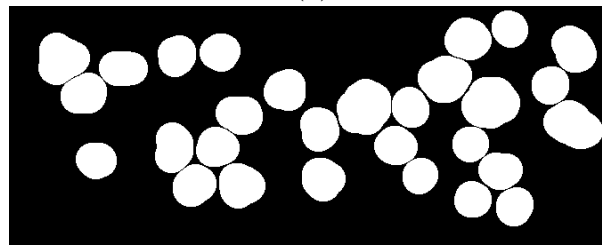


(d)

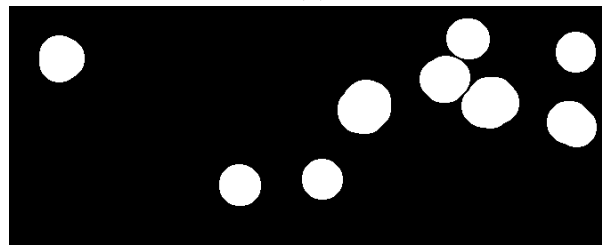
Figura 2.4: Funciones de distribución acumuladas correspondientes a la imagen de la figura 1.2 (b)



(a)



(c)



(d)

Figura 2.5: Aperturas estructurales de la figura 1.2 (a) realizadas con el disco de radio 14 píxeles (a), 18 píxeles (b) y 22 píxeles (c)

## 2.3. Análisis global del endotelio corneal basado en distribuciones granulométricas y sus momentos

Para un paciente dado, se selecciona un subconjunto de imágenes control correspondientes a imágenes normales en su mismo grupo de edad. Recordemos que los controles están divididos en dos grupo de edad distintos, de veinte e treinta años y de treinta a cuarenta.

La caracterización granulométrica de un paciente dado es comparada con la misma caracterización en los controles. Se han utilizado dos tipos de caracterizaciones: la basada en la función de distribución granulométrica y la basada en los momentos de esta función (media, desviación estándar, asimetría y curtosis).

En esta sección se aborda el análisis de endotelio corneal en base a la función de distribución granulométrica, en primer lugar, y en base a los momentos granulométricos, en segunda instancia. En ambos casos se realiza una evaluación global del endotelio en el sentido al que hemos hecho referencia anteriormente.

El elemento estructurante en base al cual se han calculado las granulometrías es el disco debido a que se simétrico y se parece al hexágono.

### 2.3.1. Análisis global basado en funciones de distribución granulométrica

Cada endotelio corneal es descrito por medio de la función de distribución granulométrica acumulada. Veamos un test gráfico para comparar  $F_0(t)$ , la función de distribución de un paciente dado, y  $F_1(t), F_2(t), \dots, F_s(t)$ , las funciones de distribución de los controles asociados con el paciente (los correspondientes a su grupo de edad). Sean  $L$  y  $U$  las funciones dadas por

$$L(t) = \underset{i=1\dots s}{\text{mín}} \{F_i(t)\}, \quad (2.8)$$

$$U(t) = \underset{i=1\dots s}{\text{máx}} \{F_i(t)\}, \quad (2.9)$$

o sea, las envolventes inferior y superior respectivamente de las funciones de distribución calculada en los controles. Si  $F_0$  está contenida entre las envolventes entonces la imagen corresponde a una imagen normal. En otro caso, estaremos observando una imagen anormal. El principal problema es la evaluación del error experimental: la proporción de imágenes normales evaluadas

como anormales. En [10], se propone un estimador empírico. Sean

$$L_{-r}(t) = \min_{i=1, \dots, r-1, r+1, \dots, s} (F_i(t)), \quad (2.10)$$

y

$$U_{-r}(t) = \max_{i=1, \dots, r-1, r+1, \dots, s} (F_i(t)), \quad (2.11)$$

las envolventes inferior y superior calculadas sobre el conjunto de controles excepto el  $r$ -ésimo. Si  $F_r$  no está contenido entre  $L_{-r}$  y  $U_{-r}$ , entonces el control es erróneamente evaluado como anormal. La proporción de controles que son evaluados como anormales según este procedimiento es el error experimental. Los valores observados se encuentran en la tabla 2.2. Es importante notar el pequeño número de controles en nuestro estudio. Una aplicación clínica necesitaría un número mayor de controles.

Cuadro 2.2: Error experimental empírico

Controles (20-30 años)	0.1034
Controles (31-40 años)	0.2941
Todos los controles	0.0556

La comparación entre el endotelio del paciente y los controles también puede realizarse por medio de un test de aleatorización. La hipótesis nula de tal test de aleatorización consideraría que nuestro paciente no puede distinguirse de los controles. Consideremos

$$d_i = \int_0^T (F_i(t) - \bar{F}_i(t))^2 dt, \quad (2.12)$$

donde  $\bar{F}_i(t) = \sum_{j=0, j \neq i}^s F_j(t)$  con  $i = 0, 1, \dots, s$ . Si realmente la imagen del endotelio del paciente no puede ser distinguida de los controles entonces cualquier permutación de  $d_0, d_1, \dots, d_s$  es equiprobable. El  $p$ -valor unilateral de tal test sería  $1 - \frac{k}{s+1}$  donde  $k$  es la posición que ocupa  $d_0$  en la serie de los  $d_i$ 's ordenados. De ahora en adelante, este  $p$ -valor será denotado por  $p_g$ . Puede encontrarse una explicación detallada de esta metodología en [13].

Si utilizamos un nivel de significatividad  $\alpha = 0.1$ , veintinueve imágenes son evaluadas como anormales ya que su  $p_g$  está por debajo de 0.1. En la tabla 2.3 se muestra un breve estudio descriptivo de los parámetros clásicos distinguiendo los casos en que el  $p_g$ -valor es menor que 0.1 (o sea, los evaluados como anormales) de los casos en que es mayor. Las distintas columnas en la tabla corresponden al mínimo, primer cuartil, mediana, tercer cuartil y

máximo de los parámetros clásicos. Observando esta tabla puede apreciarse claramente la existencia de relación entre los parámetros clásicos y el valor  $p_g$ .

### 2.3.2. Análisis global basado en momentos granulométricos

Sea  $x_0$  un vector de descriptores del endotelio compuesto por la media, la varianza, la asimetría y la curtosis de la distribución granulométrica. Similarmente tenemos los mismos valores calculados sobre los correspondientes controles,  $x_1, x_2, \dots, x_s$ . Bajo la hipótesis nula de que nuestro endotelio es un endotelio normal (o por lo menos no puede distinguirse de los endotelios control),  $x_0$  tiene que ser similar a la muestra  $\{x_1, \dots, x_s\}$ . Si  $X$  es el vector aleatorio tal que  $x_i$  con  $i = 1, \dots, s$  es una muestra aleatoria suya, puede valorarse hasta qué punto la media de  $X$  se parece a  $x_0$  mediante un test  $T^2$  de Hotelling (ver [39], pg. 63, o cualquier texto de análisis estadístico multivariante). Bajo la hipótesis de que  $X$  sigue una distribución normal,

$$F_0 = \frac{s(\bar{x} - x_0)'S^{-1}(\bar{x} - x_0)}{s-1} \frac{s-d}{d} \quad (2.13)$$

sigue una distribución  $F$  de Fisher con  $d$  y  $s-d$  grados de libertad donde  $d$  es la dimensión de  $X$  (cuatro si son usados todos los momentos granulométricos que hemos calculado),  $\bar{x} = \sum_{i=1}^s x_i/s$  y  $S$  es la matriz de covarianzas. Los valores altos de  $F_0$ , o lo que es lo mismo,  $p$ -valores pequeños del test, son asociados a endotelios corneales anormales (o, genéricamente, distintos a los controles). En ese caso lo que ocurre es que el vector de descriptores del endotelio corneal que evaluamos es demasiado distinto de la media del vector de descriptores calculado en los controles. La tabla 2.4 muestra los indicadores clásicos asociados a los cinco pacientes con los  $p$ -valores más altos. Hay que notar que los  $p$ -valores que hemos obtenido con este procedimiento son muy pequeños. Los números que están entre paréntesis son los cuantiles dentro de los controles de su grupo de edad. Por ejemplo, el paciente 1 tiene una densidad de 2673 y el 63% de los controles en su mismo grupo de edad tienen una densidad más baja que 2673. Los cuantiles que aparecen junto al coeficiente de variación nos indican que en los cinco casos son muy pequeños en comparación con los controles. Notar que los pacientes 2 y 4 tienen densidades bastante bajas (en el caso 3 no es mucho más grande), mientras que los pacientes 2 y 3 también presentan hexagonalidades bastante bajas. Sacar una conclusión es difícil. Es evidente que los  $p$ -valores observados no pueden discriminar a los pacientes de los controles.

Una alternativa robusta a este enfoque paramétrico puede ser el test de aleatorización descrito en la sección anterior. Los momentos granulométricos

Cuadro 2.3: Resumen de los parámetros clásicos teniendo en cuenta los nuevos indicadores (más detalles en el texto)

		Mín.	1 <sup>er</sup> $Q_u$ .	Mediana	Media	3 <sup>er</sup> $Q_u$ .	Max.
$p_g < 0.1$	densidad	421.7	2206	2663	2409	2985	3519
	hexagonalidad	0	45	53	54.07	61.9	88
	CV	23.7	28.3	34.55	35.27	40.17	64.5
$p_g \geq 0.1$	densidad	1501	2352	2565	2541	2730	3864
	hexagonalidad	31	42.68	52	53.11	61.25	96
	CV	21.6	31.32	34.6	35.38	39.8	52
$p_g(0.1) < 0.8$	densidad	421.7	1316	2421	2187	2657	3519
	hexagonalidad	0	43	46.5	49.33	61	88
	CV	28	33.8	40.3	39.6	43.4	64.5
$p_g(0.1) \geq 0.8$	densidad	1501	2352	2624	2576	2789	3864
	hexagonalidad	32	43.5	54	54.32	61.95	96
	CV	21.6	30.6	34.2	34.36	38.7	52
$p_m < 0.1$	densidad	421.7	1901	2356	2114	2549	3026
	hexagonalidad	0	38.5	44.5	47.98	61	88
	CV	21.6	35.7	40.4	40.37	44.7	64.5
$p_m \geq 0.1$	densidad	1501	2423	2657	2636	2876	3864
	hexagonalidad	32	45.5	56	55.38	62	96
	CV	22.7	29.9	33.9	33.63	37	52



Cuadro 2.4: Pacientes con los cinco  $p$ -valores más altos en el test  $T^2$  de Hotelling

Paciente	Densidad	Hexagonalidad	C.V.	$p$ -valor
1	2673 (63)	61 (86)	32 (34)	0.12
2	2338 (25)	35 (9)	31 (29)	0.03
3	2599 (35)	56 (37)	31 (30)	0.02
4	2494 (39)	62 (78)	31 (32)	0.02
5	2944 (84)	60 (81)	31 (24)	0.01

sustituyen ahora a las distribuciones granulométricas en el procedimiento descrito en dicha sección. El test se construye de la misma forma, sólo hay que sustituir los  $d_i$ 's de la ecuación 2.12 por

$$d_i = (\bar{x}_i - x_i)' S_i^{-1} (\bar{x}_i - x_i), \quad (2.14)$$

donde  $\bar{x}_i = \frac{\sum_{j=1, j \neq i}^n x_j}{n}$  y  $S_i^2$  es la matriz de covarianza de  $\{x_j : j = 0, \dots, i - 1, i + 1, \dots, n\}$ .

A la distancia de Mahalanobis entre los momentos granulométricos de la imagen dada y los correspondientes momentos granulométricos de los controles asociados la denotamos  $d_0$ . El test de aleatorización valora cuál es la magnitud de  $d_0$ . En la tabla 2.5 están las correlaciones entre los indicadores clásicos (densidad, hexagonalidad y coeficiente de variación de las áreas celulares),  $d_0$  y el  $p$ -valor del test de aleatorización basado en los momentos granulométricos (de ahora en adelante denotado como  $p_m$ ) calculadas utilizando el conjunto total de pacientes. Es importante notar la alta correlación existente entre el coeficiente de variación (un descriptor de forma) y las dos nuevas cantidades. Por otra parte, las correlaciones entre  $d_0$  y  $p_m$  con la densidad (un descriptor de tamaño) están muy próximas a cero.

Cuadro 2.5: Correlaciones

	Hexagonalidad	C.V.	$d_0$	$p_m$
Densidad	0.113	-0.031	-0.063	0.060
Hexagonalidad		-0.528	-0.236	0.099
C.V.			0.546	-0.293
$d_0$				-0.423

Otra forma de indagar sobre la relación de  $p_m$  con los indicadores clásicos

es dividiendo a los pacientes según el valor de  $p_m$  y describir esos indicadores en cada grupo. La tabla 2.3 contiene este estudio descriptivo en el grupo de pacientes cuyo  $p_m$ -valor es más pequeño que 0.1 y en el grupo de pacientes cuyo  $p_m$ -valor es mayor que 0.1. En este caso se aprecia una relación bastante clara y sin contradicciones: las densidades y hexagonalidades más altas junto con los coeficientes de variación más bajos pueden observarse en el grupo con  $p_m \geq 0.1$ .

### 2.3.3. Aplicación clínica: evaluación detallada de algunos casos

Con el fin de ilustrar la metodología utilizada, se ha contado con la opinión experta de una oftalmóloga para el análisis detallado de cuatro casos. Los cuatro casos corresponden a ambos ojos de dos pacientes. En la tabla 2.6 se muestran sus índices clásicos y nuevos. Para la evaluación de las cuatro imágenes seleccionadas partimos esos datos y de los tres tests gráficos.

Caso	Densidad	Hexagonalidad	CV	$p_g$	$p_g(\alpha)$	$p_m$
1	2485	36	41.1	0.88	0.81	0.66
2	2421	33	40.3	0.61	0.72	0.05
3	2513	51	45.1	0.65	0.84	0.14
4	1117	88	29.2	0	0.54	0.07

Cuadro 2.6: Índices clásicos y nuevos de los casos analizados

Los casos 1 y 2 son los que aparecen representados en las figuras 1.2(a) y (b). Corresponden a ambos ojos de un paciente de 33 años. La densidad, la hexagonalidad y el coeficiente de variación son similares en ambas (ver la tabla 2.6). Sin embargo, la inspección visual indica que las células son más irregulares en el caso 2 (figura 1.2(b)). Esta diferencia sí es captada en los tres tests gráficos. La figura 2.3 muestra que las tres funciones de distribución acumulada (la empírica de las áreas, (a); la granulométrica en número, (b) y en área (c)) calculadas para el caso 1 están dentro de las correspondientes envolventes. Sin embargo, para el caso 2, las tres funciones de distribución acumulada quedan fuera de la región delimitada por las envolventes (ver la figura 2.4). Todos los  $p$ -valores en el caso 1 son mayores que en el caso 2, como era esperable. En el caso de  $p_m$  la diferencia es mucho más acusada.

Los casos 3 y 4 se corresponden con las imágenes (a) en las figuras 2.6 y 2.7 y representan los endotelios corneales de un paciente de 33 años. Hace algún tiempo, a este paciente se le extirpó una catarata en el ojo izquierdo (figura 2.7 (a)). El caso 3 presenta una alta densidad y una baja hexagonalidad. Por el

contrario, el caso 4 tiene una baja densidad pero una alta hexagonalidad. Los descriptores clásicos no son capaces de discriminar el caso 4 como dañado. De nuevo, los tres tests gráficos sí distinguen entre ese par de casos como puede verse en las figuras 2.6(b), (c) y (d) y 2.7(b), (c) y (d). En relación con los  $p$ -valores, se observa un comportamiento similar. Esta vez, la mayor diferencia se observa en  $p_g$ .

De la observación de estos casos no puede extraerse ninguna conclusión global. Estos comentarios simplemente sirven de ilustración a la metodología expuesta.

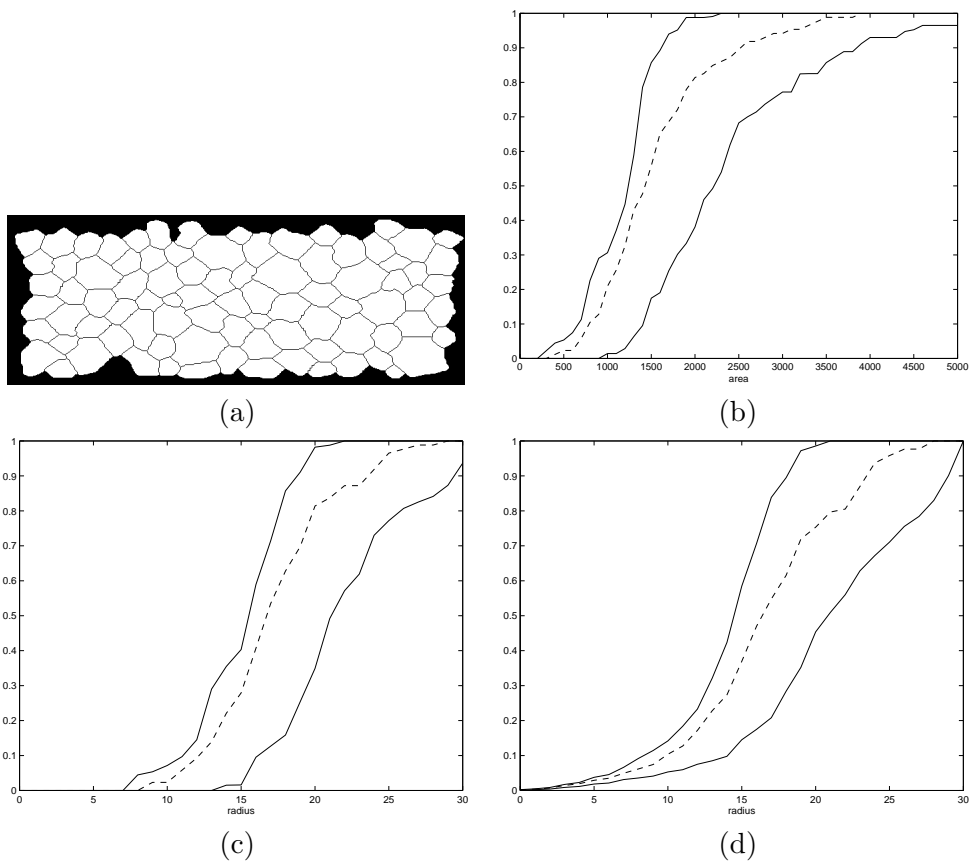


Figura 2.6: Caso 3

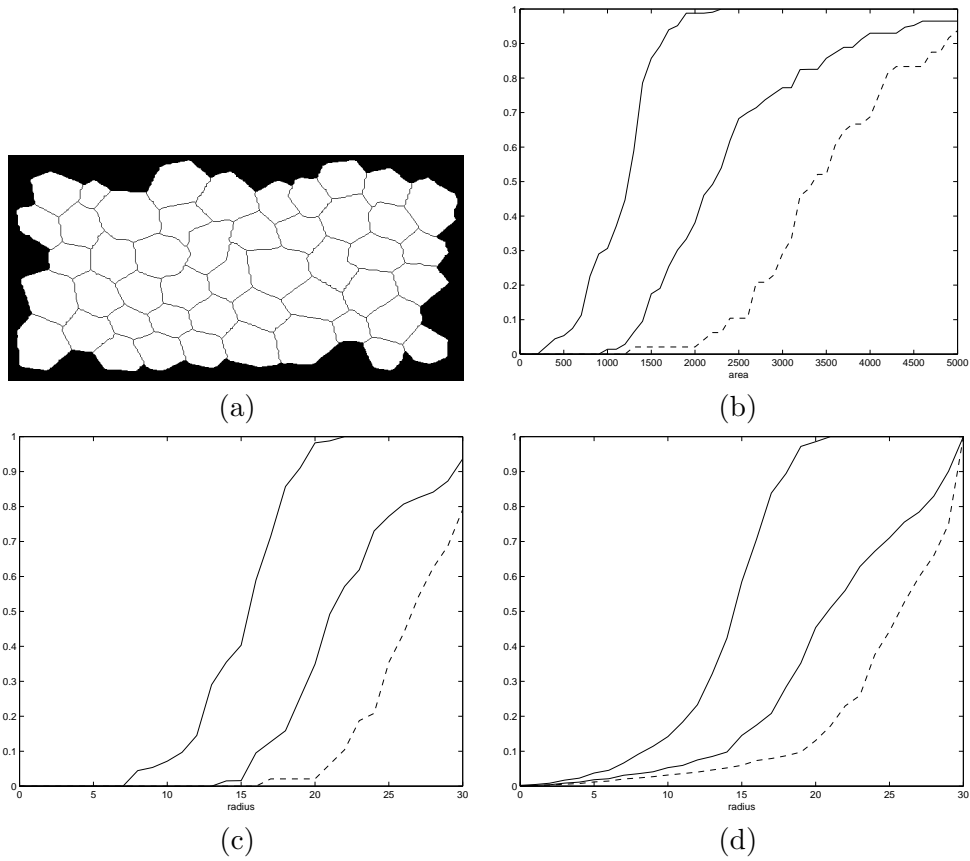


Figura 2.7: Caso 4

### 2.3.4. Agrupación de endotelios

En este punto vamos a enfocar la clasificación de endotelios bajo otro punto de vista y también con otro objetivo. Hasta ahora hemos comparado dos tipos de endotelios, los de pacientes con los controles. La aplicación clínica de esta comparación es evidente: un endotelio de paciente que se parezca a uno control será considerado normal. Ahora no establecemos esa comparación sino que partimos del conjunto total de endotelios (no distinguiendo entre endotelios pacientes y control) y realizamos un análisis cluster basado en la descripción granulométrica. El objetivo es comprobar si la descripción por medio de granulometrías puede discriminar entre tipos de endotelios diferentes.

El análisis, tal y como lo hemos abordado hasta ahora exige la selección de un grupo de endotelios control o endotelios normales. El experto puede tener

auténticos problemas en esa primera selección ya que, en ocasiones, determinar la *normalidad* de un endotelio no es un asunto sencillo. Los casos en los que esto ocurre están en el límite de lo normal. En un caso de ese tipo, el experto sabe que no se está equivocando mucho al clasificarlo como normal pero no sabe cuánto. La magnitud del error experimental empírico asociado a la muestra de endotelios normales que hemos utilizado hasta ahora viene a corroborar este razonamiento (ver tabla 2.2). La agrupación de endotelios en base a las características granulométricas dará al experto un elemento más de juicio en el que poder apoyarse.

Aunque nuestro interés está centrado en la clasificación en casos normales y no normales, el punto de vista que ahora planteamos puede ampliarla. El agrupamiento de endotelios puede no sólo aislar los casos normales de los anormales, sino también establecer tipologías de endotelios normales y tipologías de endotelios anormales. Se trata de una información complementaria que enriquece el análisis.

Hemos establecido dos rangos de edad para realizar la clasificación: por una parte hemos considerado los endotelios de personas entre 20 y 30 años y por otra los de personas entre 30 y 40 años. Cada endotelio ha sido descrito por medio de la función de distribución granulométrica tomando como elemento estructurante el disco unitario. El método de clasificación elegido ha sido el *cluster jerárquico* para lo cual hemos utilizado el procedimiento AGNES (AGlomerative NESTing) de la librería 'cluster' de R. Aunque su utilización aparece ya en este momento, la descripción más detallada así como la referencia asociada a este procedimiento se encuentra en la sección 3.2.2 del capítulo siguiente donde es usado intensivamente.

El coeficiente aglomerativo (AC) asociado al cluster jerárquico obtenido se calcula como la media de las disimilaridades entre los clusters que van uniéndose en cada paso, una vez han sido normalizadas al rango 0-1. Puede entenderse como un coeficiente de calidad de la agrupación. Su valor siempre está entre 0 y 1, y los valores próximos a 1 indican que hay un agrupamiento claro de elementos. La existencia de observaciones anómalas puede provocar valores altos de este coeficiente. Considerando los casos entre 20 y 30 años obtenemos un AC de 0.91. Se trata de un valor muy alto que podría deberse a los casos 25, 77 y 136 que, por la estructura del árbol (ver figura 2.8), parecen muy distintos al resto. El AC calculado prescindiendo de esos tres casos es igual a 0.87, lo cual demuestra que efectivamente hay una sólida estructura en el conjunto de datos.

Cuando clasificamos el conjunto de endotelios de individuos entre 30 y 40 años obtenemos un AC de 0.85. En ambos rangos de edad hemos obtenido coe-

ficientes de aglomeración bastante altos lo cual indica que hay una agrupación natural de endotelios.

Al final de un análisis cluster jerárquico hay que tomar la decisión sobre cuál es el número de grupos a considerar. En nuestro caso hemos descompuesto la muestra de endotelios, tanto la de individuos entre 20 y 30 años como la de los que están entre 30 y 40, en ocho grupos en primer lugar y en tres grupos en segundo lugar. A los grupos resultantes de la clasificación en ocho grupos los denotamos grupo 1 hasta grupo 8. A los grupos resultantes de la clasificación en tres grupos los denotaremos grupo A, B y C.

En la figura 2.8 se muestra el diagrama de árbol que representa el análisis cluster jerárquico de los endotelios entre 20 y 30 años. Como ya hemos comentado, hemos descompuesto la muestra en tres grupos en un caso y en ocho en otro. En la tabla 2.7 aparecen algunos estadísticos de los parámetros clásicos calculados en cada uno de los ocho grupos obtenidos, en concreto el máximo, el mínimo, la media y la desviación típica. En la tabla 2.8 aparecen los mismos estadísticos cuando descomponemos la muestra en tres grupos.

A pesar de que en las tablas 2.7 y 2.8 aparecen los valores medios de los parámetros clásicos, los gráficos de la figura 2.9 dan una información más completa ya que además de las medias por grupo (el grupo viene dado por el eje y) aparecen también representados los valores de los elementos que forman dicho grupo. Por ejemplo puede apreciarse que la densidad media en los grupos 1, 2 y 3 (los grupos más numerosos) son muy parecidas. Sin embargo la estructura de los grupos es muy distinta, por ejemplo en el grupo 2 hay una dispersión muy grande, es decir, la media es poco representativa de la muestra, y ello es debido a dos elementos con densidades anormalmente grandes (3476.2 y 3864.3). Sin embargo, si recurrimos al apéndice A donde están representados los endotelios de individuos entre 20 y 30 años por grupos, observamos que los dos endotelios a que hemos hecho referencia antes no se diferencian demasiado del resto del grupo. En los grupos primero y tercero los datos están concentrados en la misma zona (entre 2500 células/ $mm^2$  y 3000 células/ $mm^2$ ), excepto dos casos del primer grupo: un caso con una densidad de 1912 células/ $mm^2$  y otro con densidad igual a 3262.6 células/ $mm^2$ . Entre ambas densidades hay una diferencia de aproximadamente 1300 células/ $mm^2$  difícil de apreciar visualmente (ver el apéndice A). Puede decirse que la diferencia entre ambos grupos se encuentra en la hexagonalidad, es decir, el tercer grupo contiene endotelios con hexagonalidades más bajas y todos aquellos endotelios con hexagonalidades mayores pero que por su morfología han sido considerados similares a los primeros. En el primer grupo sin embargo no hay endotelios con densidad por debajo del 50 por ciento. Los grupos cuarto y quinto están muy bien caracterizados por sus

densidades, muy altas en el primer caso y muy bajas en el segundo. El sexto grupo está formado por un único endotelio con densidad y hexagonalidad altas. Por otra parte, todos los elementos del séptimo grupo menos uno tienen densidades dentro del rango considerado normal y además hexagonalidades altas. El único endotelio que se diferencia tiene una densidad y un hexagonalidad bastante bajas y se caracteriza por una morfología muy irregular de las células, parece que no tiene mucho en común con el resto de endotelios del grupo. Por último el octavo grupo está formado por dos endotelios con densidades y hexagonalidades relativamente normales pero observando el gráfico 2.9 (c) apreciamos que se distinguen por sus altos valores del coeficiente de variación.

Posteriormente hemos agrupado la muestra en tres conjuntos. Las representaciones gráficas de la densidad, hexagonalidad y coeficiente de variación de las áreas celulares de los elementos que componen estos conjuntos, así como las medias, aparecen en la figura 2.10 (a), (b) y (c) respectivamente. El primero de tales conjuntos, el grupo A, es el menos numeroso y es el resultado de la unión del quinto y sexto cluster de la partición en ocho conjuntos. Se han unido por tanto conjuntos muy distintos si atendemos a sus parámetros clásicos (densidades muy bajas con densidades muy altas) pero no tan distintos si lo que hacemos es observarlos en el apéndice A. A nivel visual tienen cierta semejanza y lo que resulta entonces ilógico es que el endotelio que compone el antiguo conjunto 6 tenga una densidad tan alta. El segundo conjunto, grupo B, es la unión de los antiguos conjuntos 2, 3 y 4. Se trata de conjuntos cuya representación de la densidad en la figura 2.9 (a) se encuentra aproximadamente por la misma zona. Más o menos, el mismo comentario puede hacerse del grupo C que es reunión de los antiguos conjuntos 7, 8 y 1. En esta nueva agrupación, es el tercer conjunto el que mayor concentración de hexagonalidades altas registra.

En la figura 2.11 aparece el árbol jerárquico resultado de aplicar el procedimiento de cluster jerárquico que anteriormente hemos presentado a los endotelios de individuos entre 30 y 40 años. De la misma forma que hemos hecho con los casos entre 20 y 30 años, hemos realizado el análisis separando la muestra en ocho y en tres grupos.

La tabla 2.9 contiene el máximo, el mínimo, la media y la desviación típica por ese orden de los parámetros clásicos dentro de cada uno de los ocho grupos. Sin embargo, para entrar en más detalles vamos a recurrir al mismo tipo de gráficas que en el análisis anterior. Lo primero que se aprecia es que hay muy pocos endotelios con una densidad de células realmente baja. En la figura 2.12 (a) están representadas las densidades dentro de cada grupo. Inmediatamente

se ve que, por ejemplo, el grupo 2 recoge densidades más altas que el grupo 1. Merece la pena destacar el grupo 6 que presenta una dispersión muy grande de densidades. En este grupo están contenidos los tres endotelios de la muestra con las densidades más bajas mientras que el resto de elementos presenta densidades por encima de las 2500 células/mm<sup>2</sup>. Si recurrimos a la inspección visual de los endotelios que integran el grupo seis, nos resulta difícil distinguir entre los endotelios con bajas densidades y los que tienen densidades altas. El resto de grupos son muy poco numerosos y todos sus elementos tienen densidades parecidas. Sin embargo, en el caso de los grupos 3 y 4, la diferencia puede ser establecida por sus valores del coeficiente de variación de las áreas, el cual también los diferencia, junto con una baja hexagonalidad, del resto de grupos ya que es notablemente mayor. De la observación visual también se obtiene la misma conclusión (ver el apéndice A). Respecto a los grupos 7 y 8, en los gráficos 2.12 (a), (b) y (c) no se puede establecer una diferencia clara entre ellos, a no ser por el coeficiente de variación, y tampoco entre ellos y el grupo 1. Sin embargo, de la inspección visual (ver el apéndice A) si se llega a la conclusión de que tienen una morfología bastante distinta a la de los elementos del grupo 1.

La figura 2.13 contiene gráficas del mismo estilo que en el caso de los ocho grupos pero ahora hemos formado sólo tres grupos con la totalidad de los datos. El grupo A resultante ahora es la unión de los antiguos grupos 1 y 2. El grupo B es el resultado de la unión de los antiguos 3, 4, 5 y 6 y, por último, el grupo C es la unión del 7 y el 8. En el gráfico 2.13 (a) se ve que el grupo A es bastante homogéneo, los valores de la densidad están bastante concentrados y son en general altos. En ese sentido, los endotelios que forman el grupo B son bastante heterogéneos en cuanto a la densidad de células y, aunque el rango de hexagonalidades coincide casi perfectamente con el del grupo 1, en el grupo 2 hay más valores de la hexagonalidad próximos al límite inferior. Por último, el grupo C está formado por los grupos 7 y 8 obtenidos en la clasificación en ocho grupos, como parecía natural después de observar su morfología.

En algunas ocasiones la agrupación obtenida ha resultado coherente desde el punto de vista de los parámetros clásicos, pero en otras ocasiones ha puesto en evidencia casos cuyos valores de los parámetros clásicos no describen bien la morfología del endotelio. Esto viene a avalar de nuevo el análisis granulométrico sobre el análisis clásico. Un ejemplo concreto es el endotelio e136 de la tabla A.1 del anexo el cual podría haber sido elegido como endotelio control observando los parámetros clásicos y sin embargo ha aparecido junto a los endotelios con peores densidades en el grupo de edad 20-29.

A nivel visual, en las tablas mostradas en el anexo puede observarse en ge-



neral que los grupos obtenidos son homogéneos, sobre todo, cuando atendemos a la división en ocho grupos.

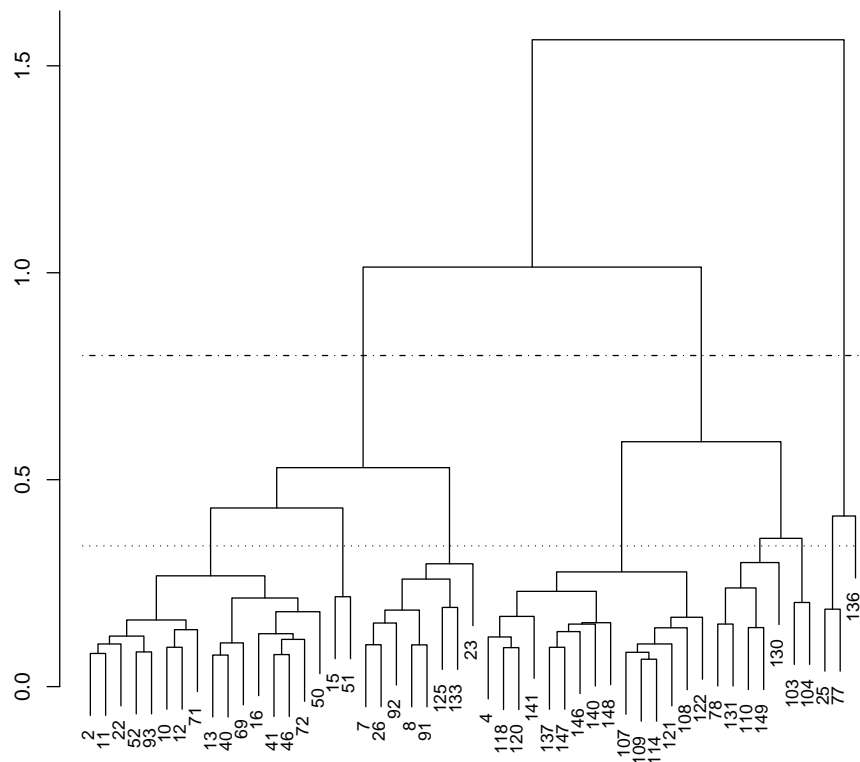


Figura 2.8: Dendrograma o diagrama de árbol resultante de la agrupación jerárquica de endotelios de pacientes entre 20 y 30 años

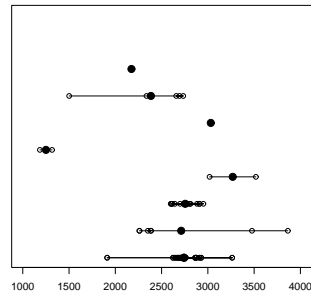
2.3 Análisis global del endotelio corneal basado en distribuciones granulométricas y sus momentos 47

Grupo	Densidad	Hexagonalidad	CV
1 (15 eltos.)	3262.6	82.0	42.4
	1912.0	54.0	22.9
	2742.1	66.0	31.7
	281.1	7.6	5.6
2 (8 eltos.)	3864.3	80.0	39.8
	2260.3	52.0	26.3
	2711.9	63.1	32.8
	666.2	9.1	4.5
3 (16 eltos.)	2951.7	70.0	38.7
	2598.8	40.0	22.7
	2755.1	56.4	31.5
	123.1	9.5	4.3
4 (2 eltos.)	3518.8	60.0	31.6
	3018.7	52.0	23.7
	3268.8	56.0	27.7
	353.6	5.7	5.6
5 (2 eltos.)	1316.1	61.0	35.3
	1185.0	44.0	28.0
	1250.6	52.5	31.7
	92.7	12.0	5.2
6 (1 elto.)	3031.7	75.0	25.8
	3031.7	75.0	25.8
	3031.7	75.0	25.8
	0.0	0.0	0.0
7 (5 eltos.)	2734.0	77.0	41.2
	1501.4	32.0	23.1
	2385.0	62.0	33.7
	518.5	17.6	7.9
8 (2 eltos.)	2182.2	64.0	41.1
	2166.6	51.0	39.5
	2174.4	57.5	40.3
	11.0	9.2	1.1

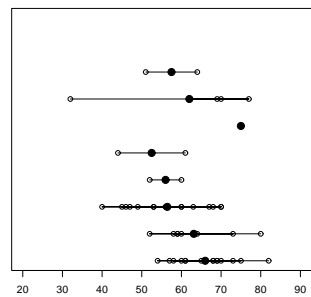
Cuadro 2.7: Resumen de los parámetros clásicos en cada uno de los ocho clusters en que se ha dividido la muestra de pacientes entre 20 y 30 años. La primera fila de cada celda contiene el máximo, la segunda el mínimo, la tercera la media y la cuarta la desviación típica

Grupo	Densidad	Hexagonalidad	CV
	3031.7	75.0	35.3
A	1185.0	44.0	25.8
(3 eltos.)	1844.3	60.0	29.7
	1030.4	15.5	5.0
	3864.3	80.0	39.8
B	2260.3	40.0	22.7
(26 eltos.)	2785.3	58.5	31.5
	391.2	9.4	4.4
	3262.6	82.0	42.4
C	1501.4	32.0	22.9
(22 eltos.)	2609.4	64.3	33.0
	382.7	10.4	6.3

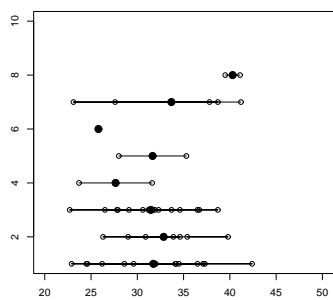
Cuadro 2.8: Resumen de los parámetros clásicos en cada uno de los tres clusters en que se ha dividido la muestra de pacientes entre 20 y 30 años. La primera fila de cada celda contiene el máximo, la segunda el mínimo, la tercera la media y la cuarta la desviación típica



(a)

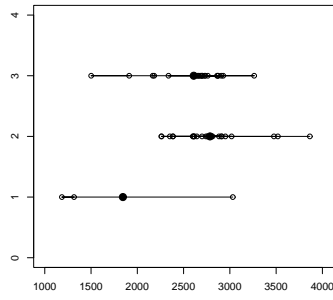


(b)

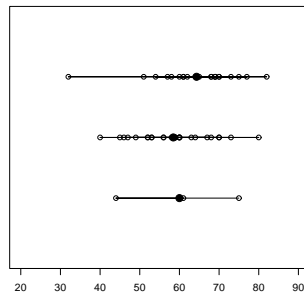


(c)

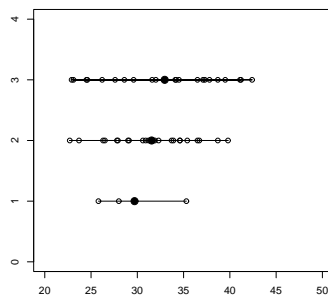
Figura 2.9: Representación de la densidad (a), hexagonalidad (b) y coeficiente de variación de las áreas (c) para cada uno de los ocho grupos obtenidos en la muestra de endotelios entre 20 y 30 años



(a)



(b)



(c)

Figura 2.10: Representación de la densidad (a), hexagonalidad (b) y coeficiente de variación de las áreas (c) para cada uno de los tres grupos obtenidos en la muestra de endotelios entre 20 y 30 años

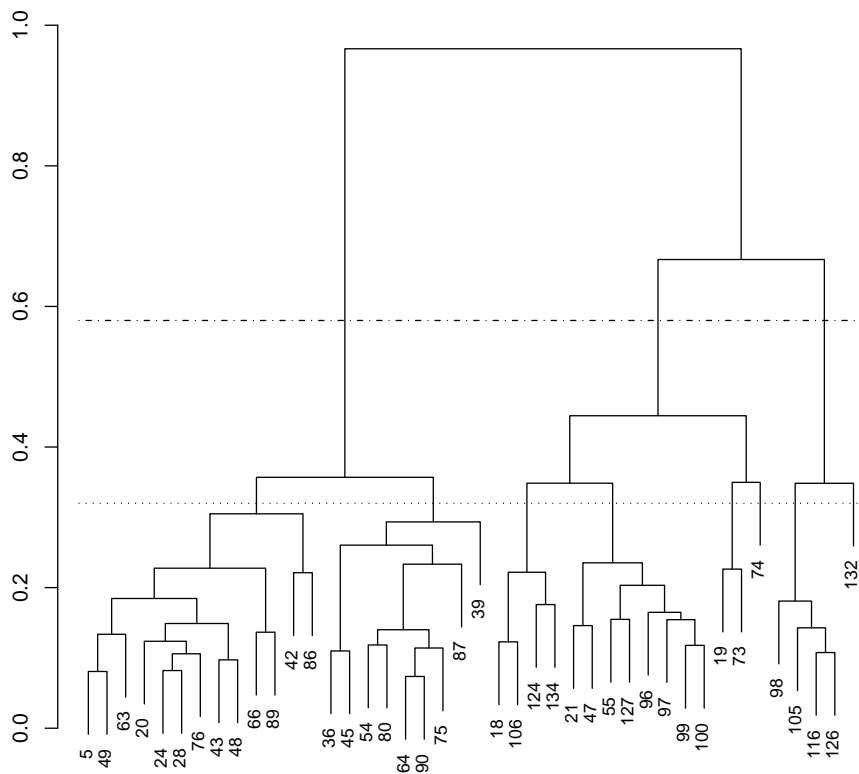


Figura 2.11: Dendrograma o diagrama de árbol resultante de la agrupación jerárquica de endotelios de pacientes entre 30 y 40 años

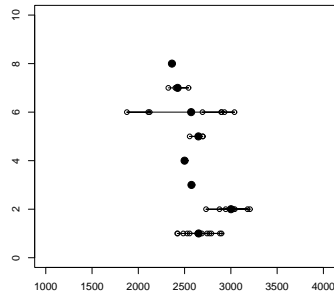
Grupo	Densidad	Hexagonalidad	CV
1 (13 eltos.)	2898.9	65.0	42.6
	2421.1	33.0	24.5
	2651.9	50.3	33.8
	161.6	10.9	5.4
2 (9 eltos.)	3206.0	71.0	45.9
	2732.0	38.0	21.6
	3001.3	56.5	35.3
	154.6	10.6	7.9
3 (2 eltos.)	2574.2	41.7	50.5
	2574.2	31.0	50.5
	2574.2	36.4	50.5
	0.0	7.6	0.0
4 (1 eltos.)	2499.7	34.0	58.8
	2499.7	34.0	58.8
	2499.7	34.0	58.8
	0.0	0.0	0.0
5 (4 eltos.)	2697.2	42.9	44.9
	2555.0	32.0	35.8
	2649.3	38.2	41.1
	81.6	4.6	4.7
6 (8 elto.)	3038.8	65.0	37.2
	1876.4	37.0	25.0
	2570.8	52.5	32.4
	459.1	9.1	4.4
7 (4 eltos.)	2542.3	68.0	42.2
	2324.6	40.0	36.1
	2424.6	50.3	39.5
	90.4	13.0	2.5
8 (1 eltos.)	2363.4	43.0	44.5
	2363.4	43.0	44.5
	2363.4	43.0	44.5
	0.0	0.0	0.0

Cuadro 2.9: Resumen de los parámetros clásicos en cada uno de los ocho clusters en que se ha dividido la muestra de pacientes entre 30 y 40 años. La primera fila de cada celda contiene el máximo, la segunda el mínimo, la tercera la media y la cuarta la desviación típica

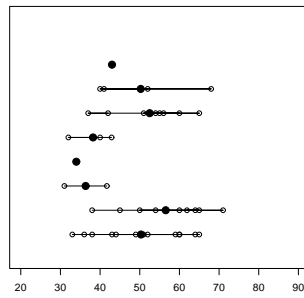


Grupo	Densidad	Hexagonalidad	CV
	3206.0	71.0	45.9
A	2421.1	33.0	21.6
(22 eltos.)	2785.0	52.9	34.4
	232.9	11.0	6.4
	3038.8	65.0	58.8
B	1876.4	31.0	25.0
(15 eltos.)	2583.7	45.3	37.8
	354.8	10.7	9.4
	2542.0	68.0	44.5
C	2324.6	40.0	36.1
(5 eltos.)	2412.3	48.8	40.5
	82.9	11.7	3.1

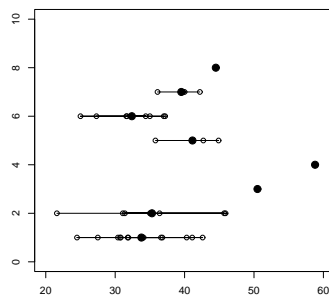
Cuadro 2.10: Resumen de los parámetros clásicos en cada uno de los tres clusters en que se ha dividido la muestra de pacientes entre 30 y 40 años. La primera fila de cada celda contiene el máximo, la segunda el mínimo, la tercera la media y la cuarta la desviación típica



(a)

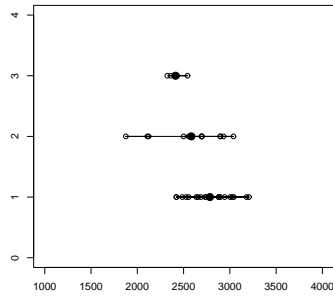


(b)

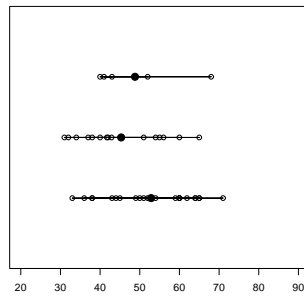


(c)

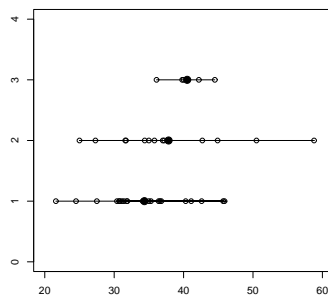
Figura 2.12: Representación de la densidad (a), hexagonalidad (b) y coeficiente de variación de las áreas (c) para cada uno de los ocho grupos obtenidos en la muestra de endotelios entre 20 y 30 años



(a)



(b)



(c)

Figura 2.13: Representación de la densidad (a), hexagonalidad (b) y coeficiente de variación de las áreas (c) para cada uno de los tres grupos obtenidos en la muestra de endotelios entre 20 y 30 años

## 2.4. Análisis local del endotelio corneal basado en distribuciones granulométricas y sus momentos

El contenido de esta sección se recoge en la referencia [53].

Hasta ahora se ha realizado un análisis global del endotelio. Sin embargo, podríamos considerar una metodología similar, salvo algunas diferencias lógicas, en la evaluación de las células endoteliales aisladamente. De entre los indicadores clásicos, uno se obtiene globalmente (la densidad) y los otros dos (hexagonalidad y coeficiente de variación de las áreas) se obtienen a partir de cantidades calculadas localmente: el número de células vecinas y el área de una célula dada. Veamos con un ejemplo cómo estas dos cantidades están obteniendo descripciones iguales de células bastante distintas. En la figura 2.15 aparecen dos parejas de células (cada fila contiene una) pertenecientes al endotelio representado en la figura 2.14 con valores aproximadamente iguales de los descriptores clásicos pero, sin embargo, con formas bastante diferentes. Esto pone en evidencia de nuevo, como ya adelantamos en la sección 2.2, la pobre descripción proporcionada por los descriptores clásicos en cuanto que no tienen en cuenta la morfología del objeto. La diferencia entre las células de cada pareja sí que se hace evidente cuando observamos su distribución granulométrica. En la figura 2.16 aparecen las funciones de distribución granulométrica calculadas utilizando un elemento estructurante circular de las células de la figura 2.15. Atendiendo a esta caracterización de las células es fácil distinguir las diferencias entre ellas.

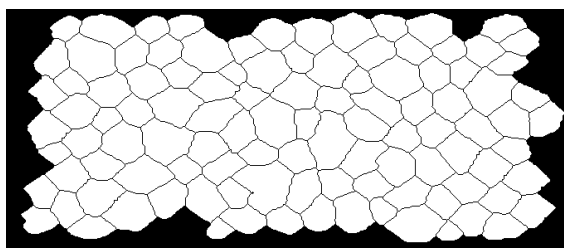


Figura 2.14: Endotelio corneal

El objetivo final de esta sección también es proporcionar una evaluación del endotelio pero en base a una evaluación previa de las células. Antes, veamos qué relación existe entre la distribución granulométrica del endotelio y la de las células que lo componen.

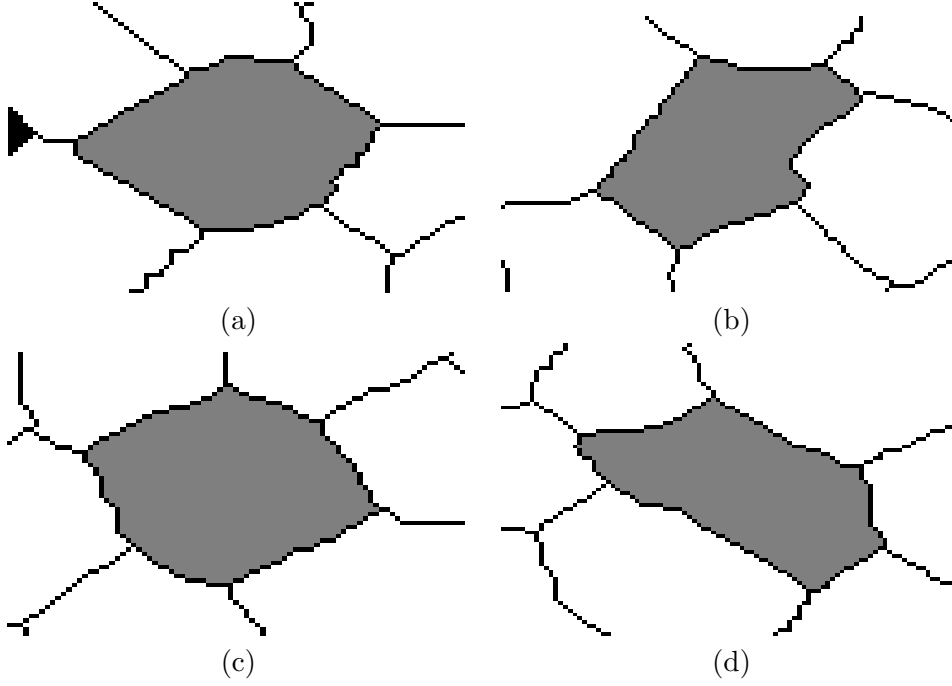


Figura 2.15: Parejas ((a-b) y (c-d)) de células hexagonales claramente distintas y con la misma área obtenidas del endotelio representado en la figura 2.14

Sea  $C$  un endotelio corneal y  $C_i$  con  $i = 1, \dots, N$  las células de las que está compuesto. Entonces  $C = \bigcup_{i=1}^N C_i$  donde las distintas células  $C_i$  son disjuntas. Sean  $F_C$  y  $F_{C_i}$  las funciones de distribución granulométrica del endotelio  $C$  y las células  $C_i$  respectivamente. Teniendo en cuenta que  $T$  es un conjunto convexo y compacto conteniendo al origen y que los  $C_i \circ T$  son conjuntos disjuntos tales que  $C \circ T = \bigcup C_i \circ T$  se sigue que  $\nu(C \circ T) = \sum_{i=1}^N \nu(C_i \circ T)$ . Además, en las condiciones anteriores,  $\lambda T$ , donde  $\lambda$  es un número positivo, es un conjunto convexo y compacto también, por tanto  $\nu(C \circ \lambda T) = \sum_{i=1}^N \nu(C_i \circ \lambda T)$ . Finalmente se obtiene

$$F_C(\lambda) = \sum_{i=1}^N \frac{\nu(C_i)}{\nu(C)} F_i(\lambda), \lambda \geq 0, \quad (2.15)$$

o sea, la distribución acumulada del endotelio corneal completo es una media ponderada de las distribuciones de tamaño acumuladas de las diferentes células. Los pesos vienen dados por los cocientes  $\frac{\nu(C_i)}{\nu(C)}$ . La derivada de la función  $F_C$ , la función de densidad de probabilidad de la correspondiente distribución

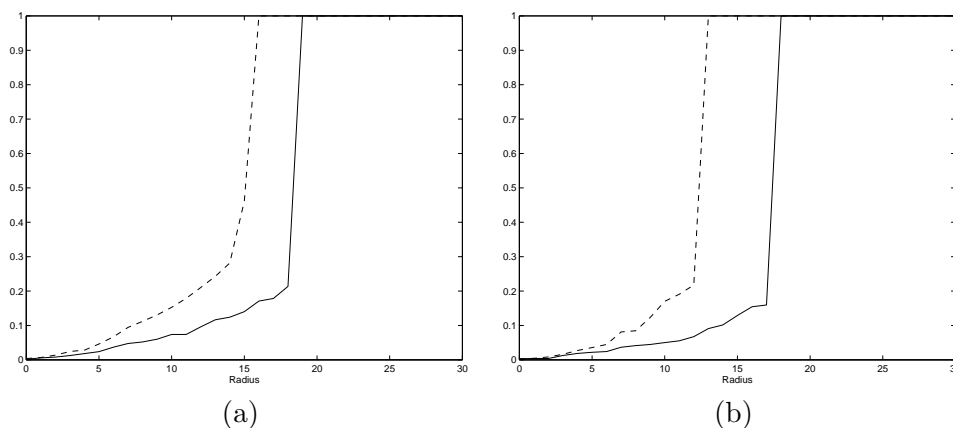


Figura 2.16: Funciones de distribución granulométrica circular de las células de la figura 2.15: (a) corresponde a la pareja (a-b) (línea continua y punteada respectivamente) y (b) corresponde a la pareja (c-d) (línea continua y punteada respectivamente)

de tamaño granulométrica, es igual a la media ponderada de las densidades de probabilidad de las distribuciones de tamaño granulométricas de las diferentes células. Formalmente, lo anterior se recoge en la siguiente ecuación.

$$f_C(\lambda) = \sum_{i=1}^N \frac{\nu(C_i)}{\nu(C)} f_i(\lambda), \lambda \geq 0, \quad (2.16)$$

donde  $f_C$  y  $f_{C_i}$  denotan las correspondientes densidades de probabilidad. Simplemente la distribución de tamaño granulométrico del endotelio completo es una mixtura de las distribuciones de tamaño granulométricas de las diferentes células donde la proporción en que interviene la  $i$ -ésima componente es el cociente entre el área de  $C_i$  y el de  $C$ .

En la figura 2.17 están representadas las funciones de distribución granulométrica circulares correspondientes a cinco células aleatoriamente elegidas del endotelio de la figura 2.14 junto con la función de distribución granulométrica del endotelio. La función de distribución del endotelio es una versión suavizada de la función de distribución de las diferentes células. También aparecen las envolventes inferior y superior de las diferentes funciones de distribución calculadas a partir de todas las células de todas nuestras imágenes.

El mismo procedimiento que hemos utilizado en la sección 2.3 para decidir si un endotelio era normal o no puede aplicarse al caso de las células. Recordemos que dicho procedimiento consistía en el cálculo de las envolventes

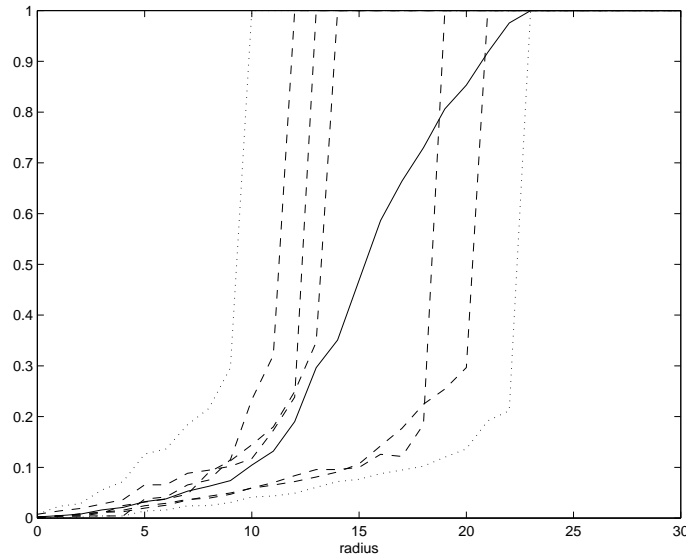


Figura 2.17: Funciones de distribución granulométricas de algunas células (líneas discontinuas), del endotelio corneal (línea continua) y envolturas superior e inferior (línea de puntos)

superior e inferior a partir de las funciones de distribución granulométricas de una cierta muestra de endotelios considerados normales (llamados 'endotelios control') y posteriormente en la comparación de la función de distribución del endotelio que queremos evaluar con las envolturas calculadas.

En este caso surge pues la necesidad de buscar células 'control'. La primera idea que aparece es la de considerar como células control todas las pertenecientes a endotelios control. Tomando dicho conjunto de células control y siguiendo con el procedimiento desarrollado en la sección 2.3 consistente en un test de aleatorización, podemos calcular para cada célula su  $p_g$ -valor asociado. Las funciones de distribución granulométrica que utilizamos son las circulares. Para asignar al endotelio un valor resumen a partir de los  $p_g$ -valores, es natural considerar la proporción de células del endotelio que tienen un  $p_g$ -valor más alto que una cierta cantidad umbral a la que denotamos  $\alpha$ . Esta proporción la denotamos mediante  $p_g(\alpha)$ . En resumen,  $p_g(\alpha)$  es la proporción de células normales usando las funciones de distribución granulométrica circulares y eligiendo como valor umbral  $\alpha$ . La tabla 2.11 contiene un breve resumen de los valores de  $p_g$  observados para las diferentes células y de los valores  $p_g(0,1)$  para los diferentes endotelios. Casi todas las proporciones  $p_g(0,1)$  son mayores que

0.8. Parece evidente que hay demasiadas células anormales que están siendo clasificadas como normales.

La tabla 2.3 muestra algunos estadísticos calculados sobre los parámetros clásicos en los pacientes que tienen valor de  $p_g(0,1)$  mayor que 0.8 por una parte, y en los pacientes cuyo valor de  $p_g(0,1)$  es menor que 0.8 por otra.

	Min.	Max.	Media	DT
$p_g$	0	0.96	0.40	0.32
$p_g(\alpha)$	0	1	0.84	0.16

Cuadro 2.11: Estudio descriptivo de los  $p_g$ -valores calculados sobre las células a partir de las funciones de distribución granulométrica circular y de los  $p_g(\alpha)$  con  $\alpha=0.1$

Los resultados obtenidos indican que hay que hacer una selección más elaborada de células control. Incluso los endotelios control contienen una cierta proporción de células anormales que, de ser utilizadas como control, conducirían a resultados poco o nada fiables. De esta forma resultarían clasificadas como normales prácticamente todas las células excepto los casos extremos. Además un número demasiado alto de células control llevaría a una pérdida de sensibilidad. En definitiva, es necesaria una selección más restrictiva ya que de ello depende en gran medida el buen funcionamiento del método. A continuación vamos a explicar el procedimiento de selección que finalmente hemos utilizado pero hay que recalcar de nuevo que en la aplicación clínica se requiere una muy cuidadosa selección de tales células control.

Se considerará una célula como control (o normal) si cumple tres condiciones:

1. Pertenecer a un paciente sin patología conocida cuyo endotelio corneal sea considerado sano por el oftalmólogo.
2. Tener seis vecinos (ser hexagonal), con lo cual la célula no puede pertenecer a la corona exterior de la imagen.
3. Su función de distribución granulométrica debe ser típica, o sea:
  - a) Para cada célula y para cada elemento estructurante  $T_i$ , se han estimado cuatro de los momentos de la función de distribución granulométrica asociada: media, varianza, asimetría y curtosis (ver sección 2.2).



- b) Sean  $m_1, \dots, m_n$  los valores observados correspondientes a un determinado momento granulométrico y sea  $\kappa_p$  el  $p$ -ésimo cuantil muestral, o sea,

$$\frac{\#\{m_i : m_i \leq \kappa_p\}}{n} \geq p \quad (2.17)$$

y

$$\frac{\#\{m_i : m_i \geq \kappa_p\}}{n} \geq 1 - p \quad (2.18)$$

donde  $\#$  es el cardinal del conjunto. Una célula será típica si cada uno de sus momentos se encuentra entre los cuantiles  $\kappa_{0,25}$  y  $\kappa_{0,75}$  estimados sobre las células que cumplen las condiciones (1) y (2).

Este procedimiento genera un conjunto de células control para cada elemento estructurante elegido  $T_i$ .

Para el análisis granulométrico de las células vamos a utilizar seis elementos estructurantes distintos:  $T_1$  (cuadrado unitario),  $T_2$  (disco unitario),  $T_3$ ,  $T_4$  y  $T_5$  (segmentos unitarios con un extremo en el origen y orientaciones 0, 45 y 90 grados respectivamente).

Dada una célula, cada uno de los elementos estructurantes proporciona un análisis distinto de la misma. Hemos utilizado más elementos estructurantes que en el análisis global debido a que podemos llegar de esa forma a un análisis más ajustado de la célula.

### 2.4.1. Relación de la información granulométrica con la hexagonalidad

La hexagonalidad es probablemente el parámetro de forma más usado en la descripción del endotelio y se define en función del número de vecinos que tiene cada célula (cantidad de células que están en contacto con ella). Queremos indagar sobre la relación existente entre el número de vecinos de una célula y su descripción granulométrica. Para ello proponemos dos experimentos.

#### Experimento 1

Como descripción granulométrica de las células vamos a utilizar algunos de los momentos granulométricos, los cuales han sido calculados como se explica en la sección 2.2. De la misma forma, utilizaremos la media,  $\mu_1$ ; la varianza,  $\mu_2 - \mu_1^2$ ; la asimetría,  $\mu_3/(\mu_2)^{3/2}$  y la curtosis,  $\mu_4/\mu_2^2$ .

En primer lugar, se han considerado por separado los momentos asociados a un elemento estructurante y se ha llevado a cabo un análisis discriminante

donde las variables independientes son los momentos y los diferentes grupos vienen definidos por el número de células vecinas. La cantidad de células que han sido clasificadas en el grupo adecuado, así como el porcentaje que dicha cantidad representa respecto a la cantidad total de células que integran el grupo, está recogida en la tabla 2.12. Cada elemento estructurante tiene asociadas dos filas: la primera está etiquetada con el nombre con que denotamos al elemento estructurante y contiene el número de células correctamente clasificadas usando los cuatro momentos correspondientes al elemento, mientras que la segunda fila contiene los porcentajes. Las dos últimas filas de la tabla contienen los resultados cuando usamos los momentos asociados a los cinco elementos estructurantes conjuntamente. La columna denominada 5-5 indica cuántas células que realmente tienen cinco vecinas han sido clasificadas como con cinco vecinas. La misma interpretación admite la notación 6-6 y 7-7. La última columna recoge cifras relativas a clasificación global. Para la realización de este análisis hemos tenido que renunciar a las células que están en el borde de la imagen ya que desconocemos su número de vecinas.

	5-5	6-6	7-7	Total
$T_1$	844	1371	514	2729
	67.1	40.9	51.9	48.7
$T_2$	874	1350	540	2764
	69.5	40.3	54.5	49.4
$T_3$	655	1439	561	2655
	52.1	43	56.6	47.4
$T_4$	661	1312	589	2562
	52.6	39.2	59.4	45.8
$T_5$	648	1513	529	2690
	51.6	45.2	53.4	48.1
Todos	703	1811	572	3086
	55.9	54.1	57.7	55.1

Cuadro 2.12: Clasificación correcta utilizando diferentes elementos estructurantes

Los mejores resultados globales los obtiene el disco con un 49.4% de células correctamente clasificadas, aunque los resultados obtenidos por el resto de elementos estructurantes no son muy distintos. Notar que la diferencia entre la mejor y la peor clasificación es sólo de un 3.6%. Sin embargo, las diferencias entre elementos son más acusadas observando las columnas 5-5, 6-6 y 7-7 separadamente. Finalmente, si se utilizan todos los elementos estructurantes

(última fila) no se obtienen resultados muy diferentes.

Momentos	5-5	6-6	7-7	Total
$\mu$	678	1720	561	2959
	53.9	51.3	56.6	52.9
$\mu, \sigma^2$	705	1812	561	3078
	56.1	54.1	56.6	55
$\mu, \sigma^2, sk$	703	1811	572	3086
	55.9	54.1	57.7	55.1
$\mu, \sigma^2, sk, ku$	703	1811	572	3086
	55.9	54.1	57.7	55.1

Cuadro 2.13: Clasificación correcta utilizando los diferentes momentos granulométricos ( $\mu$ , media;  $\sigma^2$ , varianza;  $sk$ , asimetría y  $ku$ , curtosis)

En la tabla 2.13 se muestran las cifras de clasificación correcta cuando se utilizan todos los elementos estructurantes pero se descartan algunos momentos. La primera fila muestra los resultados obtenidos cuando se utilizan sólo las medias asociadas a los diferentes elementos estructurantes, la segunda fila muestra los resultados obtenidos cuando se utilizan las medias y las varianzas, en la siguiente fila se añade al estudio la asimetría y en la última fila aparecen ya las cifras de clasificación correcta cuando utilizamos todos los momentos. Las últimas filas de las tablas 2.12 y 2.13 son coincidentes.

En este ejemplo hemos intentado averiguar la relación existente entre la distribución granulométrica de una célula y el número de células vecinas que tiene. Sin embargo no hemos tenido en cuenta el hecho de que la variable número de vecinas no sólo depende de la morfología de la célula que estudiamos sino también de la de las células que la rodean. Puede darse el caso de que una célula tenga forma regular pero esté en contacto con una célula deformada que acapare una parte proporcionalmente grande de su perímetro.

## Experimento 2

Una vez seleccionados los controles según el procedimiento explicado, se toma un elemento estructurante y se calcula la función de distribución granulométrica en todas las células para después clasificarlas como normales o anormales. La clasificación se realiza comparando la función de distribución con las envolventes obtenidas de las células control como ya hemos indicado en la sección 2.2. Cada columna en la tabla 2.14 corresponde a un elemento estructurante distinto y cada fila a un diferente número de vecinas. Las cifras que

se muestran son los porcentajes de células clasificadas como normales y que tienen un número de células vecinas determinado. Idealmente, el porcentaje de células clasificadas como normales pero con un número de vecinas distinto de seis debería ser casi nulo. En nuestros datos, se observa que las células con seis vecinas representan más del 60 %, mientras que las células con 5 ó 7 vecinas constituyen un porcentaje del total mucho más pequeño. Los porcentajes correspondientes a células con ocho vecinas clasificadas como normales es muy bajo y no hay células con cuatro o nueve vecinas clasificadas como normales.

Vecinos	$T_1$	$T_2$	$T_3$	$T_4$	$T_5$
Cuatro	0	0	0	0	0
Cinco	16	17	17	17	15
Seis	67	68	64	63	64
Siete	16	15	18	20	20
Ocho	1	0	1	1	1
Nueve	0	0	0	0	0

Cuadro 2.14: Distribución (porcentajes) de células clasificadas como normales según el número de células vecinas

### 2.4.2. Aplicación clínica: evaluación detallada de algunos casos

En esta sección una oftalmóloga ha examinado algunos casos con el fin de mostrar cuál puede ser la aplicación clínica de nuestro método.

En la figura 2.18 se muestran las imágenes especulares de ocho de nuestros casos. Los resultados del análisis se representan como sigue: en *gris oscuro* se encuentran las células con forma regular (hexagonal) (con la función de distribución granulométrica circular completamente contenida entre las envoltentes superior e inferior calculadas a partir de las funciones de distribución granulométricas circulares de las células normales), en *gris claro* se encuentran representadas las células consideradas 'casi regulares' con uno o dos puntos fuera de las envoltentes. Finalmente, las células *blancas* tienen más de dos puntos fuera de las envoltentes y son consideradas irregulares.

La figura 2.18 (a) corresponde a una mujer de 40 años que ha utilizado lentes de contacto blandas durante los últimos 20 años. En la figura 2.18 (b) aparece la imagen especular central de un hombre de 34 años que ha utilizado lentes de contacto rígidas también durante bastantes años. Las lentes de contacto reducen la permeabilidad corneal al oxígeno, limitando el suministro de oxígeno al endotelio. Aunque los índices proporcionados por el software co-

mercial eran aceptables en ambos casos (en el caso (a) la densidad es de 2356 células/ $mm^2$  y la hexagonalidad del 32 % y en el caso (b) estos dos índices toman el valor de 2555 células/ $mm^2$  y del 32 % respectivamente), el método propuesto muestra que el uso prolongado de lentes de contacto puede producir modificaciones sutiles en la morfología. A esta conclusión se llega observando el bajo número de células consideradas 'normales' o 'casi normales' por el método.

La figura 2.18 (c) muestra el endotelio corneal de una persona normal de 19 años. En este caso, la densidad y la hexagonalidad son respectivamente de 2971 células/ $mm^2$  y del 32 %. Se trata de valores altos. Observando la imagen, se aprecia que muchas de las células tienen forma regular (presentan algún tono de gris). En este endotelio normal hay coincidencia entre nuestro método y el software comercial. Pero incluso en un caso normal como éste hay un número de células que son catalogadas como anormales (las representadas en blanco), interpretamos la existencia de esas células como la variación normal que se produce en la forma de las células en cualquier endotelio.

La figura 2.18 (d) corresponde a un paciente miope de 22 años (la densidad es de 1316 células/ $mm^2$  y la hexagonalidad del 32 %). A este paciente le fue implantada una lente intraocular próxima a la córnea con el fin de corregirle la miopía. El trauma quirúrgico queda reflejado en el análisis ya que la mayoría de las células son catalogadas como irregulares (blancas). Notar que los índices clásicos podrían ser considerados por el oftalmólogo como 'no demasiado malos' y que una inspección visual cualitativa no sería capaz de separar este caso de otros considerados normales.

En la figura 2.18 (e) aparece un endotelio muy deteriorado. Esta imagen pertenece a un paciente de 67 años que ha sufrido una complicada extracción de cataratas la cual ha obligado a la implantación de una lente intraocular de cámara anterior. Este tipo de lente intraocular altera el endotelio con más frecuencia que las lentes intraoculares de cámara posterior. Además, al deterioro del endotelio contribuyó la prolongada duración de la intervención quirúrgica así como la adopción durante la misma de medidas imprevistas por el doctor. Tanto la densidad como la hexagonalidad son bajas en este caso (757 células/ $mm^2$  y 45 % respectivamente). Notar que en este caso también hay coincidencia con los indicadores clásicos, como en el caso normal considerado en (c). Esta coincidencia se da cuando estudiamos casos extremadamente patológicos o extremadamente normales. Sin embargo, consideramos que esos son precisamente los casos menos interesantes desde el punto de vista clínico ya que esos casos sólo son susceptibles de diagnóstico o pronóstico.

La figura 2.18 (f) pertenece a un paciente de 21 años al que se le ha

implantado una lente intraocular fijada en el iris hace 30 meses para corregir un error refractivo. Muchas de las células son irregulares en este caso. El endotelio corneal correspondiente al otro ojo de este paciente se encuentra en la figura 2.18(g). En este ojo también le fue implantada una lente intraocular del mismo tipo hace 18 meses. Este tipo de lentes se colocan cerca de la córnea y por tanto pueden dañar seriamente el endotelio. El estado del endotelio es un poco mejor en el segundo ojo, quizá debido a que el daño aparece de forma progresiva.

En la figura 2.18(h) se muestra el endotelio corneal de un paciente al que se le ha extraído una catarata hace 19 meses. Puede apreciarse que solamente dos células han sido consideradas 'normal', en un caso, y 'casi normal' en el otro.

Pensamos que este método puede ser muy útil en la detección de pequeños cambios en las células endoteliales pero sobre todo en el seguimiento de algunos casos en el tiempo. Por ejemplo casos de distrofia endotelial en los que las células sufren un deterioro progresivo. Las enfermedades que causan distrofia tienen un origen hereditario y comienzan a ser sintomáticas a partir de los 60 años, sin embargo mucho antes pueden empezar a producirse pequeños cambios.

## 2.5. Conclusiones

Últimamente han surgido nuevas técnicas de obtención de imágenes para el análisis de la córnea en vivo, principalmente la microscopía confocal, aunque la microscopía especular es todavía la técnica más utilizada en la práctica clínica. Sin embargo, el software comercial disponible para el análisis de este tipo de imágenes es demasiado simple y, frecuentemente, es incapaz de detectar pequeños cambios correspondientes, en algunos casos, a córneas clínicamente normales pero que están evolucionando a un estado patológico. En este capítulo se ha pretendido mostrar una nueva metodología para la descripción de endotelios corneales humanos por medio de sus imágenes especulares que tiene en cuenta el área de las células endoteliales y su forma conjuntamente. Posteriormente se indica la forma de evaluar el estado del endotelio a partir de la descripción realizada. En el caso global, la evaluación es mediante un test gráfico, fácilmente evaluable por un oftalmólogo, y los  $p$ -valores de determinados tests de aleatorización que comparamos con valores de referencia que dependen del error que estamos dispuestos a asumir. Aunque en general hay coherencia entre la evaluación mediante el procedimiento propuesto y el clásico, pensamos que el propuesto es más completo en base a la evaluación clínica

de algunos casos en los que esa coherencia no existe y a la agrupación realizada sobre la muestra mediante su descripción granulométrica, ambos análisis ponen de manifiesto lo incompleta que resulta la descripción clásica en algunas ocasiones.

En el caso local, la evaluación se produce por medio de un porcentaje, al igual que ocurre cuando tenemos en cuenta la hexagonalidad pero más fiable, no sólo debido a la descripción más completa de las células, sino que, además, este procedimiento tiene en cuenta también las células de la corona exterior.

Los comentarios realizados hasta ahora están basados en su mayoría en granulometrías circulares. Usando otros elementos estructurantes se han obtenido resultados similares. Sin embargo, sí que tiene gran influencia en el resultado la elección de las células control. De cara a una eventual aplicación del método, una buena elección de las células control es fundamental, mientras que la elección de elemento estructurante queda en segundo plano.

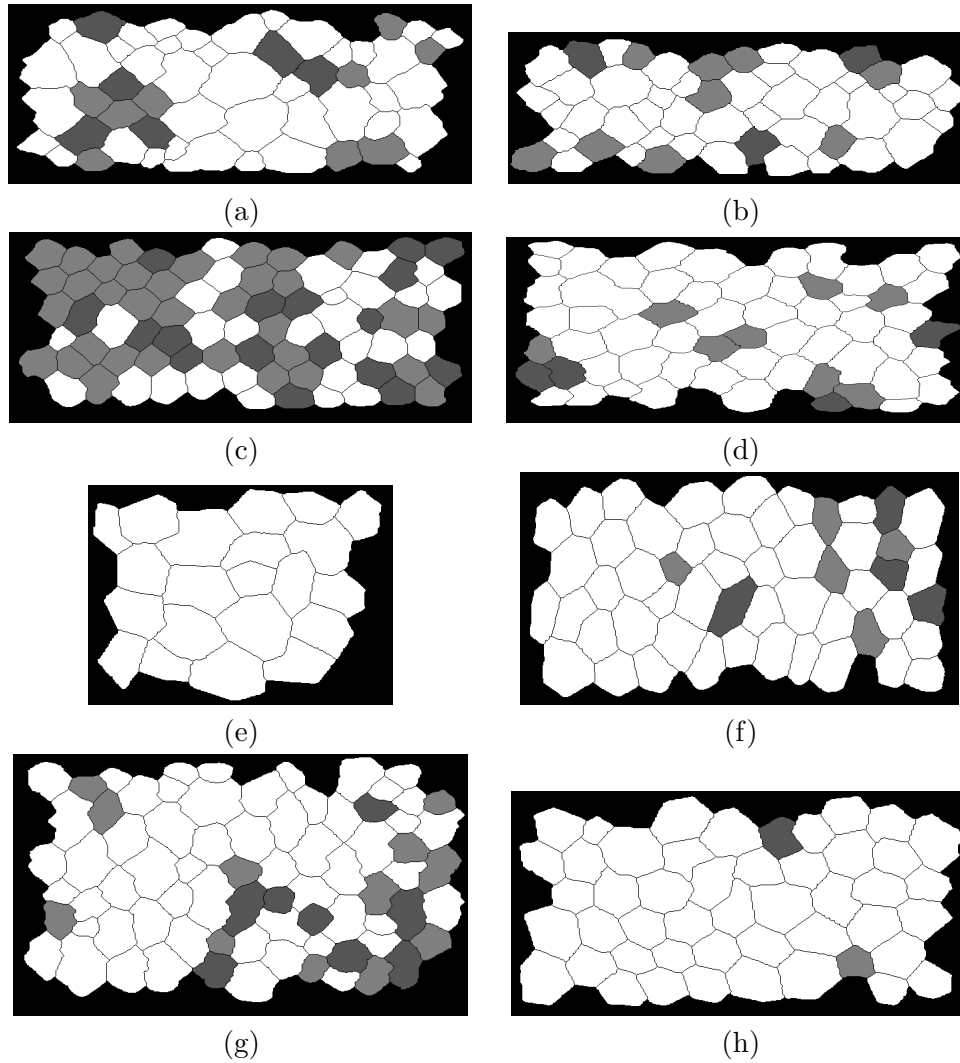


Figura 2.18: Células cuya función de distribución granulométrica circular tiene más de dos puntos fuera de las envolventes superior e inferior representadas en blanco, con uno o dos puntos fuera de las envolventes representadas en gris claro y completamente contenidas en las envolventes representadas en gris oscuro



## Capítulo 3

# Análisis del endotelio corneal mediante procesos puntuales

Este nuevo capítulo conserva el objetivo que nos planteamos en el capítulo anterior: la clasificación de endotelios corneales humanos. Sin embargo, partimos de una descripción de los endotelios y de un procedimiento de clasificación completamente distintos. Mientras que en el capítulo anterior nos hemos centrado en la morfología de la células que componen el endotelio corneal, en éste que nos ocupa vamos a prestar más atención a la localización de las mismas.

Genéricamente, en este capítulo vamos a utilizar una caracterización del endotelio basada en la localización de algunos de sus puntos relevantes. De esta manera, el resultado de dicha caracterización va a ser un patrón puntual. La clasificación de endotelios se traduce entonces en la clasificación o agrupamiento de un conjunto de patrones puntuales y es ahí adonde orientamos la metodología que se expondrá en lo que sigue.

Este tipo de imágenes ya fue descrito en [12] usando un patrón puntual bivariado compuesto por los centroides, o centros de masas, de las células y por los puntos triples (puntos en los que entran en contacto tres células). Posteriormente, para evaluar la normalidad de un endotelio, se comparaba su descripción en forma de patrón puntual con la descripción proveniente de una muestra de endotelios control de edad similar al endotelio analizado. La comparación con un grupo de endotelios control es adecuada siempre que este último se elija de forma consistente, cosa que no resulta fácil. Es muy probable que dos oftalmólogos distintos lleguen a dos elecciones también distintas de endotelios control. Otros trabajos relacionados son [14, 11].

Teniendo en cuenta lo anterior, hay que dejar claro que el enfoque que utilizamos en este capítulo es el que ya asumimos al final del anterior respecto

a la no elección de endotelios control. En la parte experimental aplicada a endotelios vamos a establecer grupos y valorar la normalidad o no de los casos que los componen cualitativamente.

Notar que no se pretende encontrar agrupaciones de puntos dentro de un patrón puntual, sino encontrar agrupaciones de patrones puntuales.

Vamos a simplificar la descripción del endotelio proporcionada en [12] ya que, en nuestro caso, consiste en un patrón univariado formado solamente por los centroides de las células. La figura 3.1 muestra el endotelio presentado en la figura 1.2(a) junto con la localización de los centroides de las células que lo componen.

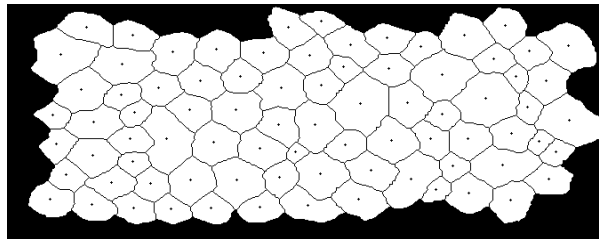


Figura 3.1: Localización de los centroides de las células del endotelio representado en 1.2(a)

El procedimiento que proponemos nos permite establecer grupos de endotelios parecidos basándonos en su caracterización mediante patrones puntuales. En primer lugar se establecen disimilaridades entre patrones puntuales y, posteriormente, éstos se agrupan mediante un procedimiento cluster. El establecimiento de las disimilaridades entre patrones puntuales va a estar en función de los dos tipos de distancias habitualmente utilizadas en la descripción de patrones puntuales: la distancia al vecino más próximo y la distancia de punto a suceso. En la sección 3.1 se indica la forma en que se van a describir los patrones puntuales y en la sección 3.2.1 se introducen las distintas disimilaridades así como los dos métodos de agrupamiento que utilizaremos. Baddeley y Gill [3] pusieron de manifiesto la analogía existente entre la estimación de determinadas características del proceso puntual a partir de las anteriores distancias y el análisis de supervivencia. Siguiendo esta idea básica, vamos a aprovechar los diferentes tests de comparación de dos muestras propuestos en la literatura sobre análisis de supervivencia para definir medidas de disimilaridad entre patrones puntuales observados en diferentes ventanas.

Para verificar la utilidad de las diferentes medidas de disimilaridad que proponemos, hemos realizado un estudio de simulación recogido en la sección

3.3.

Por último, en la sección 3.5 se presenta la aplicación del método a un conjunto de endotelios corneales humanos.

### 3.1. Descripción de patrones puntuales

En la sección 1.2.2 han sido introducidos los descriptores de patrones puntuales más usuales. En este capítulo vamos a utilizar descriptores basados en distancias.

Dado un conjunto de puntos característicos del endotelio, formado por los centroides de sus células, se van a calcular las distancias  $D$  (distancia al vecino más próximo) y  $E$  (distancia de punto a suceso) y, a partir de ellas, sus funciones de distribución empíricas,  $G$  (función de distribución de las distancias al vecino más próximo) y  $F$  (función de distribución de las distancias de punto a suceso o función de espacio vacío) respectivamente.

Es decir, en este contexto vamos a analizar el conjunto de puntos característicos del endotelio como un patrón puntual. Debido a que el patrón puntual se observa dentro de la ventana de muestreo, las diferentes distancias son datos censurados. Por ejemplo, si  $d_i$  denota la distancia del  $i$ -ésimo punto del patrón puntual,  $x_i$ , a su vecino más próximo dentro de la ventana de muestreo,  $W$ , y  $c_i$  es la distancia al complementario de  $W$ , puede ocurrir que la distancia real aunque desconocida al vecino más próximo  $d'_i$  sea más grande que  $c_i$  y menor que  $d_i$ , obviamente en el caso en que  $c_i < d_i$ . Sería el caso en que el verdadero vecino más próximo se encontrara fuera de la ventana de muestreo. El mismo comentario puede hacerse respecto a la distancia punto a suceso. Este problema se conoce en el contexto de la estadística espacial como el *efecto borde* o *efecto de la arista*. Cualquier estimador de las funciones  $F$  y  $G$  debe considerar esta distorsión y corregirla para lo cual hay propuestas varias *correcciones de borde o arista*.

Por tanto, las distancias observadas son realmente distancias censuradas. El valor de censura asociado con cada distancia depende de qué distancia estamos considerando (entre vecinos más próximos o entre punto y suceso) y de la localización del punto de muestreo. Por ejemplo, si se está examinando una distancia punto a suceso, entonces el punto de muestreo es un punto previamente elegido y el valor de censura es la distancia de este punto al marco de la ventana de muestreo. En el caso de la distancia al vecino más próximo, el punto de muestreo es un punto del patrón y el valor de censura corresponde de nuevo con la distancia al marco de la ventana de muestreo. Baddeley y Gill propusieron diferentes estimadores de  $F$  y  $G$  basándose en

esta idea. Aunque las distancias de censura son muy dependientes entre sí, los dos autores consiguieron demostrar que su técnica proporcionaba buenos estimadores. Consideraron censura por la derecha a pesar de ser conscientes de que lo natural sería considerar censura por intervalo.

En este trabajo también se tiene en cuenta, al igual que ya lo hicieron estos dos autores, la analogía entre el efecto borde y la censura aleatoria en el agrupamiento de patrones puntuales observados posiblemente sobre distintas ventanas de muestreo.

### 3.2. Diferentes formas de abordar el agrupamiento de procesos puntuales

Ésta es la parte metodológica del trabajo. En adelante, los  $m$  diferentes patrones puntuales serán denotados por  $s^{(i)}$  con  $i = 1, \dots, m$  donde el  $i$ -ésimo patrón puntual está formado por  $n_i$  puntos  $\{s_1^{(i)}, \dots, s_{n_i}^{(i)}\}$ . Los diferentes patrones puntuales pueden estar definidos en distintas regiones o ventanas de muestreo. Denotamos  $W^{(i)}$  a la ventana de muestreo donde observamos el patrón puntual  $s^{(i)}$ . Esta consideración es importante si pretendemos comparar patrones puntuales con distintas procedencias.

En el contexto del análisis estadístico de patrones puntuales, la ventana de muestreo es esencial ya que el efecto borde depende de ella. De hecho, no vamos a separar ambos objetos y los denotaremos conjuntamente,  $(s^{(i)}, W^{(i)})$ . En cierto sentido, vamos a agrupar pares de ese tipo.

Con el objeto de ilustrar el problema con que nos encontramos, veamos con un ejemplo hasta qué punto la ventana de observación puede influir en la descripción del patrón observado. Vamos a tomar un proceso puntual de Poisson con intensidad  $\lambda$ . La función de espacio vacío  $F$  teórica tiene, en este caso, la siguiente expresión

$$F(t) = 1 - e^{-\lambda t^2} \quad \text{con } t \geq 0. \quad (3.1)$$

Se ha generado una realización de este modelo en la ventana unidad  $[0, 1] \times [0, 1]$  con una intensidad de  $\lambda = 100$ . En la figura 3.2 se muestra la función teórica dada en la ecuación 3.1 junto con tres estimaciones de dicha función obtenidas a partir de la misma realización en distintas ventanas. La primera estimación usa todos los puntos generados en la ventana  $[0, 1] \times [0, 1]$ . La segunda y tercera estimaciones usan sólo los puntos en la ventana  $[0, 0,5] \times [0, 0,5]$  y  $[0, 0,3] \times [0, 0,3]$  respectivamente. Como se esperaba, las versiones estimadas son peores conforme las ventanas de muestreo (y por tanto la cantidad de distancias

disponibles) son más pequeñas. Notar que, en muchas ocasiones, observamos distintos patrones puntuales en distintas ventanas de muestreo. Para un  $t$  fijo, las diferentes estimaciones que comparamos tienen diferentes varianzas dependiendo del patrón puntual y de la ventana de muestreo asociada.

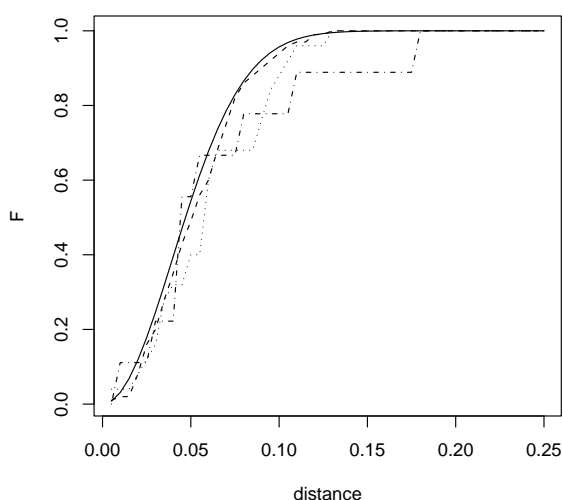


Figura 3.2: Función de distribución de las distancias de punto a suceso de un proceso puntual de Poisson: la función teórica (línea continua) y sus versiones estimadas a partir de una realización dada sobre las ventanas  $[0, 1] \times [0, 1]$  (línea discontinua),  $[0, 0,5] \times [0, 0,5]$  (línea de puntos),  $[0, 0,3] \times [0, 0,3]$  (línea de puntos y líneas)

Para agrupar procesos puntuales, va a ser necesario asignar una medida de disimilaridad a cualquier par de objetos de este tipo. La forma en que la ventana de muestreo puede afectar al cálculo de las disimilaridades propuestas es difícil de valorar. En este trabajo proponemos tres tipos de disimilaridades, el tercero de los cuales intenta evitar el condicionante de la ventana haciendo uso de métodos de análisis de supervivencia.

Como se ha comentado, la realización de una análisis cluster requiere el establecimiento de una medida de disimilaridad entre dos patrones puntuales, llamémosles  $s^{(i)}$  y  $s^{(j)}$ , y la elección de un procedimiento cluster. El apartado 3.2.1 considera diferentes medidas de disimilaridad entre patrones puntuales y el 3.2.2 presenta los procedimientos cluster que utilizaremos.

### 3.2.1. Disimilaridad entre patrones puntuales

Vamos a utilizar las *distancias al vecino más próximo*  $D$  y las *distancias de punto a suceso*  $E$  que aparecen en la sección 1.2.2 de la introducción para establecer la disimilaridad entre cualquier par de patrones puntuales. En unos casos emplearemos sus funciones de distribución (estimadas mediante estimadores con corrección de borde) y, en otros, emplearemos las distancias mismas.

En el primer caso, a cada patrón puntual le son asociadas dos funciones de distribución (la de vecino más próximo y la de punto a suceso). Sean  $\hat{F}_i$  y  $\hat{F}_j$  las correspondientes funciones de distribución de punto a suceso de  $(s^{(i)}, W^{(i)})$  y  $(s^{(j)}, W^{(j)})$  respectivamente. La disimilaridad entre ambos patrones puntuales puede ser definida como una distancia entre las correspondientes funciones de distribución, es decir,

$$D((s^{(i)}, W^{(i)}), (s^{(j)}, W^{(j)})) = d(\hat{F}_i, \hat{F}_j), \quad (3.2)$$

donde  $d$  es una métrica entre funciones. Usaremos la métrica  $L_2$ , es decir,

$$D_F((s^{(i)}, W^{(i)}), (s^{(j)}, W^{(j)})) = \int_0^{+\infty} (\hat{F}_i(t) - \hat{F}_j(t))^2 dt. \quad (3.3)$$

Obtenemos una segunda medida de disimilaridad sustituyendo en la anterior ecuación la función de distribución de punto a suceso por la función de distribución al vecino más próximo,

$$D_G((s^{(i)}, W^{(i)}), (s^{(j)}, W^{(j)})) = \int_0^{+\infty} (\hat{G}_i(t) - \hat{G}_j(t))^2 dt. \quad (3.4)$$

Sustituyendo la medida  $L_2$  por la  $L_\infty$  obtendríamos la correspondiente medida de disimilaridad. Sin embargo, como este valor se usa en el test de comparación de dos muestras de Kolmogorov-Smirnov, propondremos una disimilaridad basada en el  $p$ -valor resultante de este test y no en la distancia original.

En el segundo caso comparamos dos patrones puntuales directamente a partir de las distancias observadas de punto a suceso o al vecino más próximo. En este caso, la manera más natural de definir una medida de disimilaridad es comparando ambos conjuntos de distancias mediante un test de hipótesis de comparación de dos muestras. Sea  $p$  el correspondiente  $p$ -valor asociado a este test. Valores altos del  $p$ -valor indican alta similaridad, por tanto,  $1-p$  puede considerarse una medida posible de disimilaridad.

Hemos recurrido a dos tests no paramétricos clásicos: el test de Kolmogorov-Smirnov (KS) y el test  $\chi^2$ . Aunque se trata de dos procedimientos muy conocidos, vamos a recordarlos brevemente. El estadístico del test de Kolmogorov-Smirnov es simplemente la métrica  $L_\infty$  entre las funciones de distribución acumuladas estimadas.

Para aplicar el test  $\chi^2$ , las distancias observadas (con  $d_i$  y  $e_i$  denotamos las distancias de punto a suceso y al vecino más próximo respectivamente) se agrupan en  $k$  intervalos. Sean  $N_{ij}$  (con  $i = 1, 2, j = 1, \dots, k$ ) el número de distancias observadas en el patrón puntual  $i$ -ésimo contenidas en el  $j$ -ésimo intervalo, entonces el estadístico  $\chi^2$  toma la forma:

$$Q = \sum_{i=1,2} \sum_{j=1}^k \frac{(N_{ij} - N_{i \cdot} N_{\cdot j} / N)^2}{N_{i \cdot} N_{\cdot j} / N}, \quad (3.5)$$

donde  $N_{i \cdot} = \sum_{j=1}^k N_{ij}$ ,  $N_{\cdot j} = \sum_{i=1,2} N_{ij}$  y  $N = \sum_{i=1,2} \sum_{j=1}^k N_{ij}$ .

El estadístico  $Q$  sigue una distribución  $\chi^2$  con  $k-1$  grados de libertad bajo la hipótesis nula,  $Q \sim \chi^2(k-1)$ .

Es muy importante notar que los test de Kolmogorov-Smirnov y  $\chi^2$  procesan las distancias sin tener en cuenta en ningún momento el efecto borde ni tampoco la naturaleza censurada de este tipo de datos. Sin embargo, como se ha comentado anteriormente, un patrón puntual y su ventana de muestreo asociada pueden ser descritos por medio de un conjunto de distancias censuradas con sus correspondientes valores de censura. Al final, la comparación entre dos patrones puntuales puede ser aproximada mediante la comparación de dos conjuntos de distancias censuradas.

Existen varios tests de hipótesis para la comparación de dos muestras de datos censurados. En la introducción (ver la sección 1.2.3) figuran los más usuales: el test log-rank [29] y un estadístico de tipo Kolmogorov-Smirnov [16]. Notar que la variable tiempo o duración propia del contexto del análisis de supervivencia, se reemplaza aquí por la variable distancia observada. La diferencia más importante entre ambos tipos de variable es que las distintas distancias no son independientes entre sí, es decir, no son una muestra aleatoria. Esta situación se hace más evidente cuanto mayor es la intensidad de puntos del patrón puntual. Vamos a obviar este hecho considerando que los  $p$ -valores obtenidos de los tests tomados del análisis de supervivencia no son sino aproximaciones de los valores correctos que desconocemos. Dentro del análisis de supervivencia se denomina *tiempo de ocurrencia (observado)* al instante en que es observado un suceso. En este contexto, vamos a reemplazar la anterior variable por otra integrada por *distancias de ocurrencia*, es decir, distancias

entre un punto del patrón y su vecino más próximo o entre un punto externo y el punto más próximo del patrón puntual, siempre que sean menores que las distancias al borde de la ventana.

A partir de cada tipo de distancia,  $D$  o  $E$ , y cada opción propuesta se generará una disimilaridad distinta. Las disimilaridades derivadas de  $D$  y de  $E$  con un mismo método pueden ser combinadas. Hemos considerado dos posibles combinaciones: el máximo y la suma. Combinar disimilaridades obtenidas por distintos procedimientos es más cuestionable. Notar que, en el caso de las disimilaridades generadas por  $p$ -valores, es bien sabido que, bajo la hipótesis nula de que dos patrones puntuales son realizaciones de un mismo proceso puntual, cada  $p$ -valor estaría uniformemente distribuido en el intervalo unitario  $[0, 1]$ . Sin embargo, en el caso de utilizar conjuntamente dos tests distintos para definir la disimilaridad, los  $p$ -valores obtenidos serían dependientes.

En resumen, hemos contemplado cinco disimilaridades derivadas de las distancias  $D$  y  $E$ : la distancia  $L_2$  entre las funciones de distribución correspondientes y  $1-p$ , siendo  $p$  el  $p$ -valor resultante de aplicar el test  $\chi^2$ , el test de Kolmogorov-Smirnov, el test log-rank y el test propuesto en [16], tomando el  $\alpha$  de la definición igual a 1. Además también hemos considerado para cada opción la combinación (máximo y suma) de las disimilaridades obtenidas a partir de  $D$  y  $E$ . Es decir, contamos con veinte disimilaridades distintas en total.

### 3.2.2. Procedimientos cluster

En la sección anterior se han propuesto varias medidas de disimilaridad. En este punto hay que elegir el procedimiento cluster que utilizaremos para completar el trabajo, es decir, para agrupar patrones puntuales espaciales. Vamos a usar un método de partición y un método jerárquico. La explicación detallada de ambos métodos puede encontrarse en [25].

El método de partición se denomina PAM, Partitioning Around Medoids, y se basa en la búsqueda de  $k$  objetos típicos donde  $k$  es el número de grupos que queremos obtener. A los objetos típicos mencionados se les llama *medoides*. En torno a esos  $k$  objetos se agrupa el resto de forma que la suma de disimilaridades de cada objeto con respecto a su medioide más próximo sea mínima. El principal inconveniente de este método es que tiene que conocerse el número de clusters con antelación.

El método PAM consiste en dos fases. La primera fase, llamada BUILD, consiste en la obtención de un cluster inicial mediante la selección sucesiva de objetos de la muestra hasta obtener  $k$  objetos. El primero de esos objetos se elige de forma que la suma de disimilaridades de él con el resto de objetos



de la muestra es lo más pequeña posible. Se trata, por tanto del objeto más centrado en la muestra, si las disimilaridades se obtienen como distancias entre objetos. En nuestro caso no tiene por qué ocurrir. Para la elección de los objetos representativos siguientes se sigue el criterio de buscar entre aquéllos que no han sido seleccionados previamente, uno que resulte avalado por el resto de objetos no seleccionados en el sentido de que la diferencia entre la disimilaridad que se viene produciendo entre los objetos representativos y los que no lo son se vea reducida. En [25] se encuentra el procedimiento detallado a seguir.

La segunda fase, llamada SWAP, se intenta mejorar el conjunto de objetos representativos y por tanto, el agrupamiento conseguido por este conjunto. Esto se consigue considerando todos los pares de objetos  $(i, h)$  en los cuales el objeto  $i$  ha sido seleccionado y el objeto  $h$  no. Entonces se determina el efecto que se produce en la agrupación cuando deja de considerarse  $i$  como objeto representativo y pasa a considerarse a  $h$  como tal. Para evaluar la agrupación se observa la suma de disimilaridades entre cada objeto y el objeto representativo más similar. Cada una de las sustituciones que se produce en la fase SWAP se realiza con el objetivo de minimizar esa suma.

Los agrupamientos obtenidos mediante el procedimiento PAM admiten una representación gráfica basada en *siluetas*. En ellas se muestra cuáles son los objetos que están bien ubicados en el cluster y aquéllos que ocupan posiciones intermedias. La silueta se obtiene mediante la representación de las cantidades

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, \quad (3.6)$$

donde  $i$  es el objeto perteneciente al cluster  $A$  el cual se quiere evaluar,  $a(i)$  es la disimilaridad media del objeto  $i$  con el resto de objetos de  $A$  y  $b(i)$  es el mínimo de las disimilaridades entre  $i$  y el resto de clusters distintos de  $A$ . La media de los valores  $s(i)$  se denomina la *anchura media de la silueta*. Calculando este valor para distintas elecciones de  $k$ , podemos obtener el *coeficiente de la silueta* como el máximo obtenido sobre  $k$  al que se denota  $SC$  y que mide hasta qué punto el algoritmo ha captado la estructura cluster existente en los datos.

El clustering jerárquico ha sido obtenido utilizando el método AGNES, AGlomerative NESTing. Este método actúa realizando fusiones sucesivas de los objetos. Como es típico en cualquier método aglomerativo jerárquico, en un primer momento hay tantos clusters como objetos. En el según paso fusionamos los dos objetos más similares o más próximos para formar un cluster de dos objetos. Para ello buscamos el valor más pequeño de la matriz de disimilaridades. Para continuar es crucial tener claro la forma en que va a ser

calculada la disimilaridad entre dos clusters ya que dependiendo de ella van a obtenerse resultados distintos. Los autores de [25] se deciden por el método 'UPGMA' (unweighted pair-group average method) cuya definición es como sigue: dados dos clusters  $R$  y  $Q$ , denotando  $|R|$  y  $|Q|$  su número de objetos, la disimilaridad  $d(R, Q)$  entre los dos cluster se define como la media de todas las disimilaridades  $d(i, j)$  donde  $i$  es un objeto de  $R$  y  $j$  es un objeto de  $Q$ .

Aunque el diagrama habitual para representar los resultados de un análisis aglomerativo es el diagrama de árbol, éste se vuelve bastante complicado conforme el número de objetos aumenta. En [25] se propone una representación distinta: el *banner*. En el banner aparecen los objetos, o los clusters, que se unen en cada paso y además la disimilaridad que había entre ellos (ver la referencia anterior). A partir de la anterior representación puede obtenerse un coeficiente de la fiabilidad del cluster obtenido, el *coeficiente aglomerativo*, calculado como la media de las disimilaridades que se dan entre dos clusters previa a su fusión normalizadas a 1. Puede interpretarse como la anchura media del banner.

### 3.3. Estudio de simulación

Hemos realizado un estudio de simulación para evaluar la potencialidad de las distintas disimilaridades propuestas. Se han simulado varios tipos de patrones puntuales con diferentes valores de los parámetros. Para cada simulación, se ha realizado un análisis cluster usando las disimilaridades definidas en cada uno de los procedimientos cluster que consideramos.

#### 3.3.1. Validez del cluster obtenido

Para valorar el resultado del estudio de simulación, hemos comparado la agrupación resultante de la aplicación de los dos métodos cluster con las distintas disimilaridades con la partición real mediante cuatro índices: el estadístico de Rand ajustado, el coeficiente de Jaccard, el índice de Fowlkes y Mallows y el estadístico  $\hat{\Gamma}$  de Hubert normalizado. Veamos brevemente cuál es su definición. En [46] se encuentra un estudio más profundo del problema y, en particular, una presentación más detallada de los índices.

Sea  $C = (C_1, \dots, C_m)$  la agrupación resultante de la aplicación de un algoritmo cluster específico y  $P = (P_1, \dots, P_s)$  la partición real. Sea  $N$  el número de objetos que queremos agrupar (en nuestro caso, patrones puntuales). Además, sea  $a$  el número de pares de objetos tales que ambos objetos pertenecen al mismo grupo en  $C$  y al mismo grupo en  $P$ ,  $b$  el número de pares

de objetos tales que ambos objetos pertenecen al mismo grupo en  $C$  pero a distintos grupos en  $P$ ,  $c$  el número de pares de objetos con sus elementos perteneciendo a distintos grupos en  $C$  pero al mismo grupo en  $P$  y  $d$  el número de pares de objetos tales que sus elementos pertenecen a distintos grupos en  $C$  y a distintos grupos en  $P$ . Entonces,  $a + b + c + d = M$ , donde  $M$  es el número de posibles pares. Sea  $m_1 = a + b$  el número de pares cuyos elementos pertenecen a un mismo grupo en  $C$  y  $m_2 = a + c$  el número de pares cuyos elementos pertenecen al mismo grupo en  $P$ .

El *estadístico de Rand ajustado* se define como

$$R = \frac{2(ad - cb)}{2ad + (a + d)(b + c) + b^2 + c^2}. \quad (3.7)$$

El *coeficiente de Jaccard* es

$$J = \frac{a}{a + b + c}, \quad (3.8)$$

donde  $0 \leq J \leq 1$ ; el *índice de Fowlkes y Mallows* se obtiene como

$$FM = \frac{a}{\sqrt{m_1 m_2}}, \quad (3.9)$$

donde  $0 \leq FM \leq 1$  y, por último, el *estadístico de Hubert normalizado*  $\hat{\Gamma}$  es

$$\hat{\Gamma} = \frac{Ma - m_1 m_2}{\sqrt{m_1 m_2 (M - m_1)(M - m_2)}}, \quad (3.10)$$

donde  $-1 \leq \hat{\Gamma} \leq 1$ .

Las cuatro cantidades que acabamos de introducir toman valores más altos cuanto mejor es la calidad de la agrupación resultante, es decir, cuanto más se parece a la estructura real.

La función de densidad de probabilidad de estos estadísticos bajo la hipótesis nula de que no hay estructura cluster es desconocida y se ha tenido que recurrir a técnicas de Monte Carlo. Teniendo en cuenta que una agrupación determinada  $C$  puede interpretarse como una aplicación  $g$  del conjunto de objetos en  $1, \dots, m$ , donde  $m$  es el número de grupos, entonces, puede obtenerse una muestra de los anteriores estadísticos bajo la hipótesis nula  $H_0$  produciendo  $k$  aplicaciones aleatorias  $g_i$ ,  $i = 1, \dots, k$  correspondientes a otras tantas agrupaciones  $C_i$  con  $i = 1, \dots, k$ , las cuales se comparan con la partición real  $P$ . De esta forma pueden obtenerse los  $p$ -valores de Monte Carlo para cada estadístico de los anteriores.

### 3.3.2. Experimentos

El objetivo de este estudio experimental es la evaluación de la capacidad que tienen las distintas disimilaridades expuestas para distinguir entre patrones puntuales generados mediante modelos distintos y para agrupar aquéllos que son realizaciones de un mismo proceso puntual.

Hemos elegido tres tipos de procesos puntuales: el proceso puntual de Poisson, es decir, la aleatoriedad espacial completa; el proceso Cluster de Poisson correspondiente a un modelo agregado y, finalmente, el proceso de Strauss que corresponde a un modelo regular.

El *proceso de Poisson* es una referencia constante en la teoría de procesos puntuales. Representa la aleatoriedad espacial completa y constituye el mecanismo estocástico más simple posible para la generación de patrones puntuales espaciales. En la práctica, este modelo es un estándar con el que se compara cualquier patrón antes de abordar un análisis más completo ya que, en primer lugar, debe descartarse la aleatoriedad completa en los datos. Un proceso puntual de Poisson homogéneo de intensidad  $\lambda$  es un proceso puntual que cumple los siguientes requerimientos:

1. El número de puntos en un conjunto acotado  $B$  sigue una distribución de Poisson con media  $\lambda\nu(B)$  donde  $\nu(B)$  es el área de  $B$ .
2. Dado  $\Phi(B) = n$ , las  $n$  observaciones en  $B$  forman una muestra independiente de la distribución uniforme en  $B$ .

El proceso de Poisson se utiliza en la definición de modelos más complicados. Es el caso del *proceso Cluster de Poisson* que es generado como se explica a continuación: primero se genera un proceso puntual de Poisson de intensidad  $\rho$ . A este primer proceso se le denomina *proceso de padres*. A cada *padre* se le asocia un entero aleatorio no negativo, el número de hijos. Los hijos de un padre dado se colocan en torno del padre independientemente y con una distribución de probabilidad dada. El proceso puntual final está compuesto sólo por los hijos, es decir, el proceso formado por los padres no es considerado. En nuestros experimentos el número de hijos será un número fijo  $N$  y las posiciones relativas respecto a los padres seguirán una distribución normal radialmente simétrica con media cero y varianza  $\sigma^2$ .

Un proceso de Strauss es un proceso puntual de Markov. Los procesos puntuales de Markov sirven como marco para la modelización de procesos de inhibición. Se dice que un proceso puntual es de Markov de rango  $\rho$  si, dada una realización del proceso en  $A - \{s\}$ , ésta depende sólo de los sucesos en  $B(s, \rho) - \{s\}$ , donde  $B(s, \rho)$  es la bola cerrada de radio  $\rho$  centrada en  $s$ . En el proceso

de Strauss, la intensidad condicional de una ocurrencia en un punto  $x$  dada una realización del proceso en un conjunto  $B$  es  $\alpha c^{s_x}$  donde  $s_x$  es el número de ocurrencias a una distancia  $r$  o menos de  $x$ . A estas ocurrencias se les denomina vecinos de  $x$ . El parámetro  $c$ , con  $0 \leq c \leq 1$ , describe la interacción entre vecinos mientras que el parámetro  $\alpha$  controla la intensidad del proceso. En nuestros experimentos el proceso Strauss simulado estará condicionado al número total  $N$  de puntos en la ventana.

Hemos considerado seis experimentos. Con ello pretendemos evaluar el éxito del método en distintas situaciones. Notar que en todo momento consideramos patrones puntuales con la misma intensidad ya que patrones con distinta intensidad son fácilmente discriminables.

Los seis experimentos que hemos diseñado están organizados de igual manera: en cada uno de ellos, se consideran tres grupos de patrones puntuales y se aplican los procedimientos cluster PAM y AGNES a la unión de todos los patrones considerando los distintos tipos de disimilaridades que hemos descrito en 3.2.1. Cada uno de los tres grupos de patrones puntuales está formado por 100 realizaciones de un mismo proceso puntual generadas en la ventana unidad  $W = [0, 1] \times [0, 1]$ . Como conocemos la verdadera partición, podemos calcular los estadísticos de Rand, Jaccard, Fowkles y Mallows y  $\hat{\Gamma}$  de Hubert.

El primer experimento considera los siguientes tres grupos: 100 realizaciones de un proceso puntual de Poisson con intensidad  $\lambda=120$ ; 100 realizaciones de un modelo Strauss con  $r=0.04$ ,  $c=0$  y  $N=120$  y por último, el tercer grupo contiene 100 realizaciones de un proceso Cluster de Poisson con  $\rho=20$ ,  $N=6$  y  $\sigma^2=0.001$ . En este experimento se pretende observar qué ocurre cuando tratamos con patrones puntuales procedentes de modelos distintos. Puede probarse fácilmente que todos los procesos puntuales utilizados tienen la misma intensidad (igual a 120). En la figura 3.3 se muestra un patrón puntual de cada tipo.

En el segundo experimento se han considerado tres diferentes procesos puntuales cluster de Poisson. Hemos mantenido fijo el número de hijos,  $N=6$ , y la intensidad en el proceso de padres,  $\rho=20$ , pero hemos variado el valor de  $\sigma^2$ . En particular, hemos tomado  $\sigma^2=0.0005$ ,  $0.001$  y  $0.002$ . Este parámetro es el que controla el grado de agregación. El objetivo es comprobar la respuesta del método ante agregaciones distintas. La figura 3.4 muestra una realización de cada proceso.

El tercer experimento considera 3 procesos de Strauss diferentes, es decir, diferentes grados de regularidad, aunque en todos los casos  $N=120$ . Como valor fijo de  $c$  hemos tomado  $c=0$  y como valor de  $r$  en cada grupo hemos elegido  $r=0.01$ ,  $0.04$  y  $0.07$ . La figura 3.5 muestra un ejemplo de cada tipo de

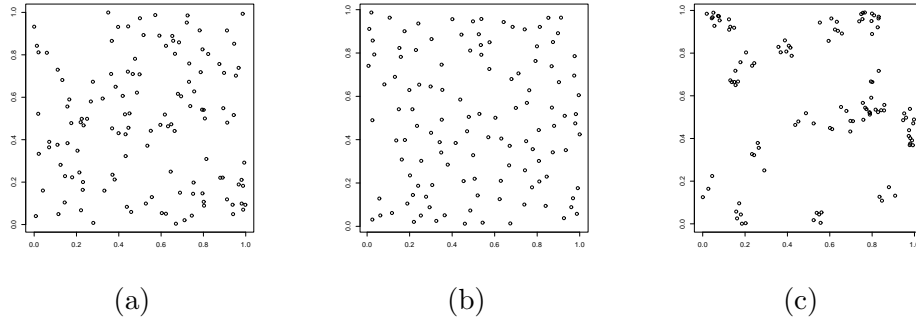


Figura 3.3: Primer experimento: (a) patrón puntual de Poisson con  $\lambda=120$ , (b) patrón puntual de Strauss con  $r=0.05$ ,  $c=0$  y  $N=120$ , y (c) patrón puntual cluster de Poisson con  $\rho=20$ ,  $N=6$  y  $\sigma^2=0.001$

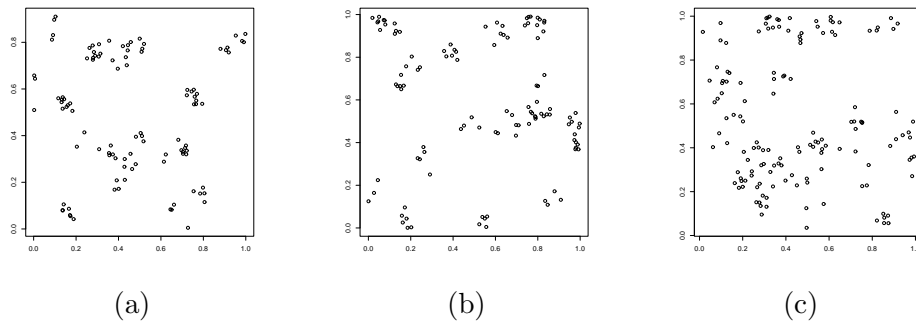


Figura 3.4: Segundo experimento: patrones puntuales cluster de Poisson con  $\rho=20$ ,  $N=6$  y (a)  $\sigma=0.0005$ , (b)  $\sigma=0.001$  y (c)  $\sigma=0.002$

proceso puntual.

En el cuarto experimento también consideramos procesos puntuales de Strauss. En esta ocasión tomamos  $N=120$ ,  $r=0.04$  y los distintos grupos vienen definidos por los tres valores siguientes de  $c$ : 0, 0.05 y 0.07. La figura 3.6 muestra diferentes realizaciones de los anteriores modelos.

El quinto experimento tiene en cuenta tres procesos puntuales cluster de Poisson cada uno con un número distinto de clusters. De nuevo construimos tres grupos de 100 patrones puntuales cada uno, con  $\sigma^2=0.0005$  en común y distintos valores de  $N$ : 2, 4 y 6. Los procesos de padres han sido generados respectivamente con  $\rho=60$ ,  $\rho=30$  y  $\rho=20$ . La figura 3.7 muestra un ejemplo de cada uno de estos grupos.

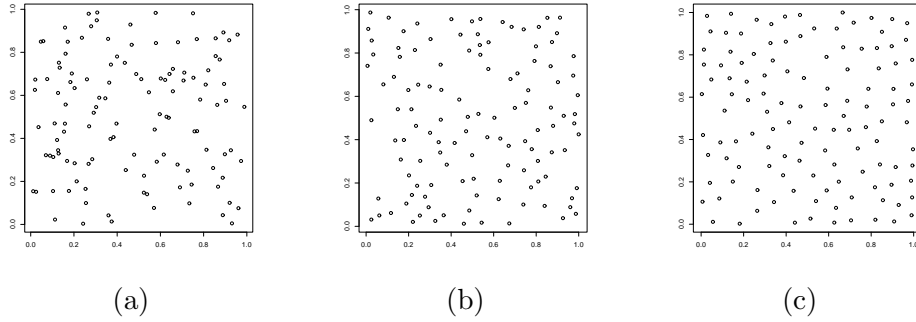


Figura 3.5: Tercer experimento: diferentes patrones puntuales Strauss con  $c=0$ ,  $N=120$  y (a)  $r=0.01$ , (b)  $r=0.04$  y (c)  $r=0.07$

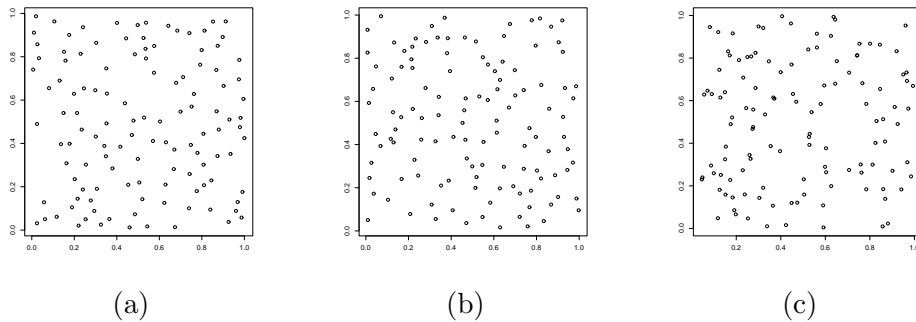


Figura 3.6: Cuarto experimentos: diferentes patrones puntuales Strauss con  $N=120$ ,  $r=0.04$  y (a)  $c=0$ , (b)  $c=0.05$  y (c)  $c=0.07$

Todos los procesos puntuales considerados en los anteriores experimentos tienen en común el valor de la intensidad. Eso quiere decir que la proporción de distancias censuradas en todos ellos es aproximadamente igual. En el sexto experimento queremos investigar la influencia que tiene la intensidad en el comportamiento del método. Los procesos puntuales que hemos considerado en este experimento son los mismos que hemos considerado en el segundo experimento con distinto valor de  $\rho$ , en este caso hemos tomado  $\rho=12$  y de esta forma hemos reducido la intensidad en un 40%. Se trata de ver qué ocurre cuando aumenta la proporción de distancias censuradas en el experimento.

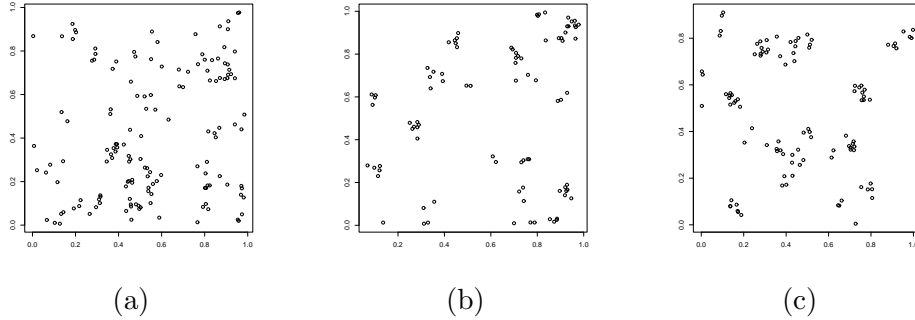


Figura 3.7: Quinto experimento: diferentes patrones puntuales cluster de Poisson con  $\sigma=0.0005$  y (a)  $\rho=60$ ,  $N=2$  (b)  $\rho=30$ ,  $N=4$  y (c)  $\rho=20$ ,  $N=6$

### 3.4. Resultados

Las tablas 3.1, 3.2, 3.3, 3.4, 3.5 y 3.6 contienen los estadísticos de Rand ajustado (R), Jaccard (J), Fowkles y Mallows (FM) y el  $\hat{\Gamma}$  de Hubert junto con su correspondiente  $p$ -valor de Monte Carlo siempre que sea mayor que 0,05.

Todas las tablas tienen la misma estructura. La primera fila indica el procedimiento cluster utilizado: de partición o jerárquico. La columna encabezada por  $L^2$  usa como disimilaridad está métrica entre las funciones de distribución obtenidas para las distancias de punto a suceso (fila etiquetada  $E$ ), y para las distancias al vecino más próximo (fila etiquetada  $D$ ) o ambas consideradas conjuntamente (fila etiquetada  $sum$ ). La columna encabezada  $KS$  muestra los resultados alcanzados usando como medida de disimilaridad  $1-p$  donde  $p$  es el  $p$ -valor obtenido del test de Kolmogorov-Smirnov aplicado a las distancias  $E$  o a las distancias  $D$ . En los casos en que la disimilaridad se obtiene en forma del complementario de algún  $p$ -valor, las filas etiquetadas  $sum$  y  $max$  indican la suma o el máximo de las anteriores disimilaridades. La columna etiquetada  $Chi2$  contiene los resultados cuando utilizamos el test  $\chi^2$  en el cálculo de la disimilaridad, la etiquetada con  $LR$  se refiere al test log-rank y con  $SF$  estamos refiriéndonos al test de Smirnov-Fleming.

Se obtienen buenos resultados en general ya que en casi todos los casos se rechaza la hipótesis de etiquetado aleatorio con un  $p$ -valor de 0. Hay algunos pocos casos en que, basándonos en los estadísticos de Rand ajustado y el  $\hat{\Gamma}$  de Hubert, no puede rechazarse tal hipótesis (la mayoría de ellos surgen cuando se utiliza el procedimiento cluster jerárquico en el experimento 6).

Además, parece que las distancias al vecino más próximo dan mejores resultados que las distancias de punto a suceso. Esta diferencia es menos evi-



dente en el experimento 5. Sólo en este caso se produce mejora considerando el máximo y la suma de disimilaridades.

Respecto a los algoritmos cluster utilizados sólo puede decirse que el que mejores resultados ha producido de los dos es el método de partición. Cada método intenta explicar los datos de una forma distinta y la discusión de lo apropiado o no de cada uno queda fuera del ámbito de este trabajo.

De los resultados obtenidos en el primer experimento (ver tabla 3.1) concluimos que todas las disimilaridades son capaces de discriminar perfectamente entre patrones aleatorios, regulares y agrupados.

Los resultados del segundo experimento no son tan buenos. Sin embargo, debe tenerse en cuenta la dificultad del problema dada por el gran parecido entre los dos primeros procesos de los tres que intentamos discriminar (ver figura 3.4). Lo cierto es que, a nivel visual, resulta bastante difícil distinguirlos. Los mejores resultados son los obtenidos con las disimilaridades resultantes de los tests que consideran censura.

En el tercer experimento, se alcanzan resultados muy buenos, excepto para las disimilaridades basadas en los tests  $\chi^2$  y Kolmogorov-Smirnov donde no se considera ningún tipo de corrección de borde.

Respecto al cuarto experimento, los resultados no son muy buenos, pero de nuevo debemos hacer notar la gran similitud existente entre los patrones integrantes del primer y segundo grupo (ver figura 3.6).

En el quinto experimento, se alcanzan los mejores resultados combinando la información proveniente de los dos tipos de distancias (la distancia al vecino más próximo y de punto a suceso). Resulta bastante razonable si tenemos en cuenta que, en este caso, las distancias  $D$  y  $E$  no están dando información redundante ya que, cuando el número de hijos (por padre) se incrementa y el número de padres se reduce, las distancias  $E$  deberían hacerse más grandes y las  $D$  más pequeñas. La disimilaridad que mejor resultado da en este experimento es la basada en el test log-rank.

Los resultados del sexto experimento son peores en comparación con los del segundo. Sin embargo, si comparamos ambos experimentos respecto a las distancias de punto a suceso (ver tablas 3.2 y 3.6) podemos apreciar la importancia de la censura por la derecha. En el segundo experimento, considerando las distancias  $E$ , se tiene que las disimilaridades generadas sin considerar censura son similares (incluso en algunos casos mejores) a las otras disimilaridades. Si consideramos las distancias  $E$  generadas en el sexto experimento junto con el método cluster de partición, los tests log-rank y Smirnov-Fleming dan los mejores resultados. De hecho, se rechaza la hipótesis de etiquetado aleatorio, cosa que no ocurre con el resto de disimilaridades cuando consideramos los

estadísticos Rand ajustado y  $\hat{\Gamma}$  de Hubert. El resultado parece obvio puesto que cuanto más pequeña es la intensidad, más distancias  $E$  censuradas nos encontramos. No ocurre lo mismo con las distancias  $D$ : cuando la intensidad decrece, a su vez decrece tanto el número de distancias censuradas como el número de las que no lo son.

En esta situación, no es posible señalar una de las disimilaridades como la mejor. Sí podemos decir que, en casi todos los casos, es una de las siguientes disimilaridades la que mejores resultados da:  $L^2$ ,  $LR$  y  $SF$ , con lo cual se confirma la importancia de considerar el efecto borde en el problema.

Sobre si mejoran los resultados al considerar la censura por la derecha, no hay una respuesta clara. Parece que los resultados mejoran pero no se puede ser categórico en esta cuestión.

### **3.5. Agrupamiento de endotelios corneales humanos**

En este capítulo hemos abordado el agrupamiento de patrones puntuales en base a cinco tipos de disimilaridades. Hay muchos tipos de imágenes que pueden ser descritas mediante un patrón puntual y las imágenes de endotelio corneal constituyen uno de ellos. La forma en que transformamos una imagen de endotelio corneal en un patrón puntual es considerando los centroides de las células que lo componen.

Vamos a analizar ahora la cuestión del agrupamiento de endotelios basándonos en las disimilaridades que hemos presentado.

En el capítulo 2 han sido explicadas las deficiencias que presenta la descripción usual que realizan los paquetes comerciales de las imágenes de endotelio corneal. Recordar que dicha descripción consiste en tres parámetros, a los que se ha denominado parámetros clásicos: (i) densidad celular, (ii) hexagonalidad y (iii) coeficiente de variación de las áreas celulares. Ya fue puesto de manifiesto entonces que los anteriores tres parámetros sólo son capaces de diferenciar endotelios muy distintos. En este capítulo se pretende, al igual que en el capítulo 2, mejorar la descripción aportada por los parámetros clásicos.

Un endotelio corneal puede ser descrito mediante un patrón puntual consistente en los centroides de sus células. Dado un conjunto de endotelios y sus representaciones en forma de patrón puntual asociada podemos calcular la disimilaridad entre cualquier par. En este caso, hemos limitado el problema al considerar sólo las distancias al centroide más próximo y  $1-p$  como la disimilaridad entre patrones puntuales (endotelios) donde  $p$  es el  $p$ -valor resultante del test log-rank.

Cuadro 3.1: Estadísticos de validación del cluster (p-valor) en el experimento 1

		Partición						Jerárquico					
		$L^2$	KS	Chi2	LR	SF	$L^2$	KS	Chi2	LR	SF		
E	R	0.617	0.594	0.433	0.678	0.628	0.555	0.546	0.555	0.508	0.555		
	J	0.597	0.574	0.450	0.648	0.605	0.586	0.578	0.586	0.553	0.586		
	FM	0.748	0.729	0.621	0.786	0.754	0.762	0.755	0.762	0.737	0.762		
	$\hat{\Gamma}$	0.618	0.594	0.433	0.678	0.628	0.612	0.599	0.612	0.564	0.612		
D	R	1	1	1	0.887	0.990	1	0.569	1	0.904	1		
	J	1	1	1	0.859	0.987	1	0.596	1	0.880	1		
	FM	1	1	1	0.924	0.993	1	0.771	1	0.936	1		
	$\hat{\Gamma}$	1	1	1	0.887	0.990	1	0.627	1	0.904	1		
sum	R	1	0.725	0.709	0.860	0.797	0.555	0.555	0.569	0.941	0.869		
	J	1	0.689	0.676	0.829	0.763	0.586	0.586	0.596	0.925	0.840		
	FM	1	0.816	0.806	0.907	0.865	0.762	0.762	0.771	0.961	0.913		
	$\hat{\Gamma}$	1	0.725	0.709	0.860	0.797	0.612	0.612	0.627	0.941	0.869		
max	R	1	0.585	1	0.941	0.990	0.556	0.565	1	0.537	1		
	J	1	0.570	1	0.925	0.987	0.586	0.594	1	0.571	1		
	FM	1	0.726	1	0.961	0.993	0.762	0.769	1	0.749	1		
	$\hat{\Gamma}$	1	0.586	1	0.941	0.990	0.612	0.625	1	0.588	1		

Cuadro 3.2: Estadísticos de validación del cluster (p-valor) en el experimento 2

		Partición						Jerárquico					
		$L^2$	KS	Chi2	LR	SF	$L^2$	KS	Chi2	LR	SF		
E	R	0.082	0.099	0.061	0.080	0.084	0.098	0 (0.4)	0 (0.4)	0.039	0.106		
	J	0.241	0.261	0.232	0.251	0.243	0.283	0.328	0.327	0.283	0.288		
	FM	0.389	0.415	0.377	0.402	0.391	0.446	0.568	0.566	0.457	0.452		
	$\hat{\Gamma}$	0.082	0.099	0.061	0.080	0.084	0.101	0 (0.4)	0 (0.4)	0.044	0.109		
D	R	0.612	0.448	0.454	0.661	0.628	0.415	0 (0.4)	0.481	0.593	0.602		
	J	0.589	0.469	0.466	0.632	0.602	0.494	0.330	0.529	0.576	0.584		
	FM	0.741	0.639	0.636	0.774	0.752	0.686	0.572	0.710	0.731	0.737		
	$\hat{\Gamma}$	0.612	0.449	0.454	0.661	0.628	0.469	0 (0.4)	0.521	0.593	0.602		
sum	R	0.236	0.099	0.371	0.386	0.451	0.002 (0.27)	0.120	0.451	0.075	0.079		
	J	0.331	0.262	0.409	0.423	0.464	0.302	0.305	0.508	0.281	0.276		
	FM	0.498	0.378	0.581	0.594	0.634	0.507	0.476	0.692	0.447	0.438		
	$\hat{\Gamma}$	0.237	0.038	0.371	0.387	0.451	0.002 (0.27)	0.127	0.488	0.079	0.082		
max	R	0.167	0.394	0.443	0.393	0.392	0.011	0 (0.4)	0.476	0 (0.4)	0 (0.4)		
	J	0.295	0.438	0.459	0.433	0.424	0.308	0.330	0.525	0.312	0.312		
	FM	0.456	0.611	0.629	0.605	0.596	0.515	0.572	0.706	0.529	0.529		
	$\hat{\Gamma}$	0.167	0.396	0.443	0.394	0.392	0.016	0 (0.4)	0.513	0 (0.4)	0 (0.4)		

Cuadro 3.3: Estadísticos de validación del cluster (p-valor) en el experimento 3

		Partición						Jerárquico					
		$L^2$	KS	Chi2	LR	SF	$L^2$	KS	Chi2	LR	SF		
E	R	0.530	0.176	0.156	0.514	0.452	0.447	0 (0.4)	0 (0.4)	0.378	0.414		
	J	0.523	0.322	0.298	0.513	0.470	0.485	0.330	0.330	0.445	0.480		
	FM	0.687	0.492	0.461	0.678	0.640	0.659	0.572	0.572	0.623	0.663		
	$\hat{\Gamma}$	0.530	0.181	0.158	0.514	0.453	0.456	0 (0.4)	0 (0.4)	0.390	0.441		
D	R	1	0.503	0.411	0.932	1	1	0 (0.4)	0 (0.4)	0.560	1		
	J	1	0.544	0.444	0.913	1	1	0.330	0.330	0.590	1		
	FM	1	0.724	0.616	0.954	1	1	0.572	0.572	0.766	1		
	$\hat{\Gamma}$	1	0.545	0.412	0.932	1	1	0 (0.4)	0 (0.4)	0.618	1		
sum	R	1	0.176	0.212	0.876	0.922	1	0 (0.4)	0 (0.4)	0.941	1		
	J	1	0.322	0.335	0.847	0.900	1	0.330	0.330	0.924	1		
	FM	1	0.492	0.505	0.917	0.948	1	0.572	0.572	0.961	1		
	$\hat{\Gamma}$	1	0.181	0.216	0.876	0.922	1	0 (0.4)	0 (0.4)	0.941	1		
max	R	1	0.440	0.442	0.922	1	1	0 (0.4)	0 (0.4)	0.961	1		
	J	1	0.496	0.497	0.902	1	1	0.329	0.330	0.949	1		
	FM	1	0.677	0.679	0.948	1	1	0.570	0.572	0.974	1		
	$\hat{\Gamma}$	1	0.468	0.470	0.923	1	1	0.001 (0.23)	0 (0.4)	0.961	1		

Cuadro 3.4: Estadísticos de validación del cluster (p-valor) en el experimento 4

		Partición						Jerárquico					
	$L^2$	KS	Chi2	LR	SF	$L^2$	KS	Chi2	LR	SF			
E	R	0.179	0.044	0.024	0.266	0.243	0.292	0 (0.4)	0 (0.4)	0.246	0.235		
	J	0.302	0.275	0.236	0.342	0.330	0.413	0.330	0.330	0.378	0.372		
	FM	0.464	0.443	0.384	0.510	0.497	0.603	0.572	0.572	0.562	0.555		
D	$\hat{I}$	0.180	0.047	0.025	0.267	0.243	0.322	0 (0.4)	0 (0.4)	0.263	0.252		
	R	0.512	0.086	0.149	0.407	0.518	0.555	0.006 (0.11)	0 (0.4)	0.451	0.560		
	J	0.510	0.305	0.354	0.435	0.527	0.586	0.325	0.330	0.515	0.587		
sum	FM	0.675	0.484	0.554	0.606	0.692	0.762	0.556	0.572	0.704	0.761		
	$\hat{I}$	0.512	0.096	0.184	0.407	0.522	0.612	0.014	0 (0.4)	0.504	0.611		
	R	0.524	0.051	0.034	0.398	0.439	0.546	0 (0.4)	0 (0.4)	0.502	0.551		
max	J	0.522	0.277	0.251	0.426	0.456	0.579	0.330	0.330	0.545	0.582		
	FM	0.687	0.444	0.406	0.598	0.626	0.756	0.572	0.572	0.726	0.759		
	$\hat{I}$	0.525	0.055	0.035	0.398	0.439	0.600	0 (0.4)	0 (0.4)	0.547	0.605		
max	R	0.505	0.232	0.075	0.417	0.508	0.551	0.009	0 (0.4)	0.427	0.547		
	J	0.508	0.384	0.278	0.446	0.508	0.582	0.324	0.330	0.502	0.581		
	FM	0.674	0.576	0.443	0.616	0.674	0.759	0.553	0.572	0.694	0.759		
$\hat{I}$	0.506	0.262	0.078	0.417	0.508	0.606	0.021	0 (0.4)	0.483	0.606			

Cuadro 3.5: Estadísticos de validación del cluster (p-valor) en el experimento 5

		Partición					Jerárquico				
		$L^2$	KS	Chi2	LR	SF	$L^2$	KS	Chi2	LR	SF
E	R	0.470	0.320	0.373	0.458	0.459	0.422	0.393	0.331	0.369	0.378
	J	0.477	0.397	0.415	0.468	0.469	0.453	0.445	0.414	0.455	0.445
	FM	0.646	0.572	0.587	0.638	0.639	0.624	0.620	0.591	0.642	0.624
	$\hat{\Gamma}$	0.470	0.325	0.374	0.458	0.459	0.424	0.399	0.341	0.399	0.390
D	R	0.572	0.450	0.339	0.579	0.454	0.486	0 (0.4)	0.457	0.484	0.467
	J	0.556	0.466	0.391	0.566	0.473	0.530	0.330	0.511	0.529	0.516
	FM	0.715	0.636	0.563	0.723	0.643	0.710	0.572	0.693	0.709	0.696
	$\hat{\Gamma}$	0.572	0.450	0.339	0.580	0.455	0.521	0 (0.4)	0.491	0.519	0.498
sum	R	0.521	0.433	0.522	0.523	0.530	0.492	0.394	0.517	0.667	0.477
	J	0.518	0.461	0.520	0.518	0.522	0.528	0.445	0.557	0.636	0.524
	FM	0.683	0.632	0.684	0.683	0.686	0.703	0.620	0.737	0.778	0.705
	$\hat{\Gamma}$	0.521	0.435	0.523	0.523	0.530	0.516	0.400	0.566	0.667	0.512
max	R	0.531	0.600	0.581	0.661	0.633	0.527	0 (0.4)	0.541	0.498	0.487
	J	0.523	0.581	0.566	0.631	0.606	0.555	0.330	0.574	0.540	0.532
	FM	0.686	0.735	0.723	0.773	0.755	0.727	0.572	0.752	0.720	0.712
	$\hat{\Gamma}$	0.531	0.600	0.581	0.661	0.633	0.555	0 (0.4)	0.593	0.538	0.524

Cuadro 3.6: Estadísticos de validación del cluster (p-valor) en el experimento 6

	Partición					Jerárquico					
	$L^2$	KS	Chi2	LR	SF	$L^2$	KS	Chi2	LR	SF	
E	R	0.006 (0.11)	0.003 (0.21)	0.007 (0.08)	0.010	0.016	0.019	0 (0.4)	0 (0.4)	0.002 (0.27)	0.004 (0.17)
	J	0.209	0.211	0.214	0.237	0.214	0.244	0.327	0.311	0.311	0.295
	FM	0.346	0.349	0.353	0.387	0.353	0.396	0.564	0.572	0.525	0.490
D	$\hat{F}$	0.006 (0.11)	0.003 (0.21)	0.007 (0.08)	0.010	0.016	0.020	-0.001 (0.49)	0 (0.4)	0.004 (0.17)	0.005 (0.13)
	R	0.348	0.170	0.156	0.236	0.232	0 (0.4)	0 (0.4)	0 (0.4)	0.011	0 (0.4)
	J	0.393	0.296	0.287	0.339	0.328	0.330	0.330	0.320	0.320	0.566
sum	FM	0.565	0.457	0.447	0.506	0.494	0.572	0.572	0.541	0.541	0.566
	$\hat{F}$	0.348	0.170	0.156	0.256	0.232	0 (0.4)	0 (0.4)	0.022	0 (0.4)	0.003 (0.21)
	R	0.018	0.019	0.055	0.122	0.156	0.006 (0.11)	0 (0.4)	0 (0.4)	0.199	0.003 (0.21)
max	J	0.218	0.212	0.236	0.299	0.303	0.323	0.330	0.330	0.204	0.531
	FM	0.358	0.350	0.382	0.467	0.468	0.553	0.572	0.572	0.505	0.531
	$\hat{F}$	0.018	0.019	0.055	0.126	0.158	0.015	0 (0.4)	0 (0.4)	0.204	0.006 (0.11)
max	R	0.063	0.182	0.086	0.129	0.205	0 (0.4)	0 (0.4)	0.005 (0.13)	0.005 (0.13)	0.003 (0.21)
	J	0.238	0.295	0.254	0.313	0.341	0.330	0.330	0.288	0.316	0.316
	FM	0.384	0.456	0.406	0.490	0.515	0.572	0.572	0.475	0.536	0.536
max	$\hat{F}$	0.063	0.182	0.086	0.138	0.212	0 (0.4)	0 (0.4)	0.007 (0.08)	0.005 (0.13)	0.005 (0.13)



En lo que sigue se comparará el agrupamiento que se produce utilizando los parámetros clásicos con el producido por nuestro procedimiento. El resultado final es analizado por un médico. Los endotelios analizados pertenecen a 23 individuos con edades entre 17 y 28 años.

Se ha utilizado como algoritmo cluster el Partitioning Around Medoids (PAM).

El algoritmo PAM requiere que le sea suministrado a priori el número de clusters en que ha de ser descompuesta la muestra. En un caso práctico como el que estamos analizando, esa cantidad suele ser desconocida. Sin embargo, la elección del número de clusters puede hacerse mediante la *anchura media de la silueta* que ya ha sido introducida en la sección 3.2.2 tomando como número de clusters aquél que mayor valor de esa cantidad proporcione.

La agrupación óptima en nuestro caso se ha obtenido con 7 grupos. Para este número de grupos, la anchura media de la silueta es 0.7, lo cual indica que hay una clara estructura cluster detrás de los datos. Sin embargo, si usamos los parámetros clásicos como descriptores del endotelio y realizamos el agrupamiento mediante el mismo algoritmo cluster (PAM) obtenemos una anchura media de la silueta de 0.4 (con 6 grupos), lo cual indica una estructura cluster muy débil. Los descriptores han sido estandarizados, ya que sus rangos de variación son muy distintos, y se ha utilizado la distancia euclídea como medida de la disimilaridad entre endotelios. La figura 3.8 donde está representada la densidad celular versus la hexagonalidad para los casos que forman parte del análisis, viene a confirmar la práctica inexistencia de estructura en la muestra.

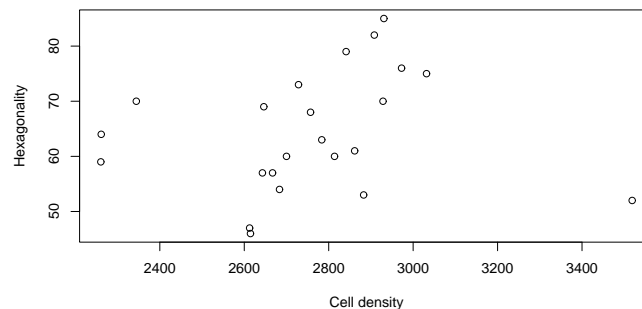


Figura 3.8: Densidad celular versus hexagonalidad

Vamos a analizar los parámetros clásicos dentro de cada uno de los siete grupos obtenidos con el procedimiento basado en la descripción mediante pa-

trones puntuales. La tabla 3.7 muestra el mínimo, la media y el máximo de la densidad celular y la hexagonalidad dentro de cada grupo.

Cuadro 3.7: Descripción clásica de los grupos obtenidos mediante nuestro procedimiento

	Cardinalidad	Densidad celular (min-mean-max)	Hexagonalidad (min-mean-max)
1	2	2260-2261-2261	59-61.5-64
2	1	3519-3519-3519	52-52-52
3	3	2613-2661-2757	46-53.7-68
4	3	2784-2827-2883	53-58.7-63
5	9	2344-2702-2973	54-65.1-76
6	4	2841-2885-2931	61-76.8-85
7	1	3032-3032-3032	75-75-75

De la observación de la tabla 3.7, se extrae la conclusión de que los descriptores clásicos no llegan a detectar diferencias salvo en los casos extremos. En la figura 3.9, está representado el medioides de cada grupo. De esta forma puede apreciarse gráficamente las diferencias entre los distintos grupos. Notar que los descriptores clásicos no pueden discriminar esas clases.

Además, la figura 3.9 revela algunos hechos interesantes. El grupo 1 es fácilmente detectable mediante los parámetros clásicos debido a que los dos casos que lo componen tienen una baja densidad. Además con la nueva descripción, también podemos decir que se caracterizan por una distribución irregular de los centroides. El grupo 2 constituye una observación anómala respecto a los parámetros clásicos debido a su muy alta densidad. Si, por otra parte, nos remitimos a la figura 3.9 (b) vemos que también se caracteriza por una distribución de los centroides irregular. Se trata de un caso muy distinto al integrante del grupo 7 que también está caracterizado por una densidad muy alta. He aquí un ejemplo claro de cómo los parámetros clásicos, en determinados casos, no pueden distinguir endotelios claramente distintos. Los grupos 4 y 6 constituyen otro ejemplo de este tipo. Ambos grupos tienen densidades altas y hexagonalidades intermedias o altas, sin embargo la distribución del patrón puntual mediante el cual son representados son muy distintas. En el caso del grupo 6 es bastante regular cosa que no ocurre en el grupo 4. Un comentario similar puede hacerse para los grupos 3 y 5. Ambos grupos presentan densidades y hexagonalidades intermedias, pero la disposición espacial de los centroides es muy distinta en uno y otro. Es más regular en los patrones del grupo 5.

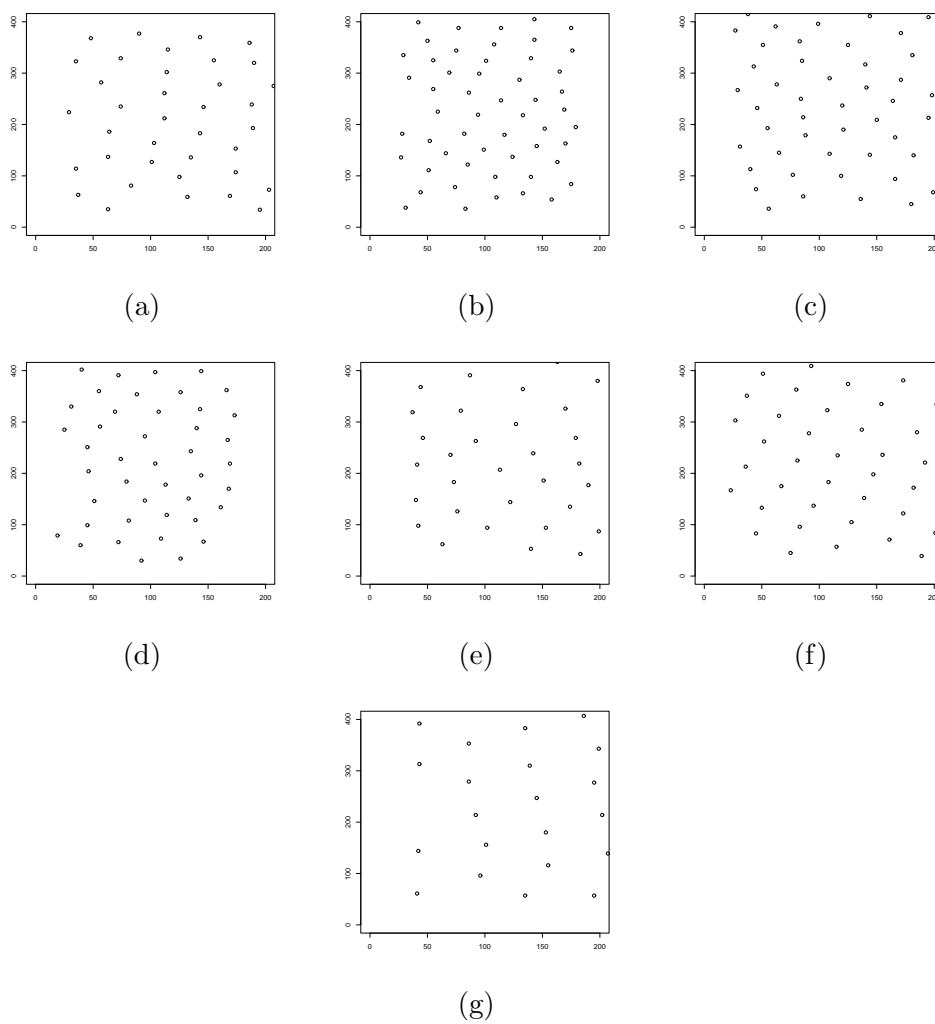


Figura 3.9: Medioides para los grupos: (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6 , (g) 7

Los resultados que hemos expuesto han sido examinados por una oftalmóloga la cual opinó que los grupos 1, 2, 3 y 4 estaban formados por endotelios enfermos. La experta confió más en la caracterización producida por nuestro procedimiento que en la producida por los parámetros clásicos a la hora de tomar decisiones ya que según ella refleja mejor las diferencias entre endotelios sanos y enfermos.

### 3.6. Conclusiones y comentarios

Este capítulo trata el problema del agrupamiento de patrones puntuales. Cada patrón puntual considerado es descrito mediante dos tipos de distancias: la distancia punto a suceso y la distancia al vecino más próximo. Este tipo de descripción de patrones puntuales es ampliamente usada. Posteriormente, se han considerado diferentes disimilaridades entre patrones puntuales. En primer lugar, se han considerado métricas entre las funciones de distribución acumuladas estimadas para ambos tipos de distancias (caso de la disimilaridad  $L^2$ ). En segundo lugar, se han calculado disimilaridades comparando directamente las distancias observadas (las disimilaridades obtenidas de esta forma han sido la  $KS$ ,  $Chi2$ ,  $LR$  y  $SF$ ). Ambos tipos de distancias consideradas son en realidad observaciones censuradas como fue apuntado por Baddeley y Gill ([3]). Aunque en realidad son observaciones censuradas por intervalo, las hemos considerado como censuradas a la derecha de la misma manera que hacen en su trabajo los autores que hemos mencionado. En las disimilaridades dadas por  $LR$  y  $SF$  hemos tenido en cuenta la censura. En ese caso, las disimilaridades son  $1-p$ , donde  $p$  es el  $p$ -valor de dos tipos distintos de test propuestos en la literatura de análisis de supervivencia para la comparación de dos muestras de valores censurados por la derecha.

Hemos realizado un estudio de simulación con distintos tipos de procesos puntuales: aleatorios, regulares y agregados; y la conclusión es que el procedimiento puede distinguir perfectamente entre los tres tipos de proceso. Sin embargo, los resultados hacen que no podamos decantarnos sobre qué tipo de similaridades son mejores: las que tienen en cuenta la censura o las que no lo hacen. Sí queda claro que se obtienen peores resultados si no se tiene en cuenta el efecto borde.

Las distancias al vecino más próximo son las que mejor resultado dan quizá porque contienen información de tercer grado. Sólo en el quinto experimento ha influido de forma determinante en los resultados la información proporcionada por la distancia de punto a suceso.

Por último, hemos proporcionado una aplicación clínica basada en una

muestra de imágenes de endotelios corneales humanos. El principal objetivo con este tipo de imágenes era determinar el grado de normalidad o anormalidad de los casos. Para ello asignamos a cada imagen el patrón puntual dado por los centroides de las células. Las imágenes consideradas anormales son las que presentan una disposición espacial anormal de los centroides de las células.

El análisis cluster que hemos realizado sobre la muestra de endotelios se ha basado en las disimilaridades de tipo  $LR$  (una de las que mejor resultado han dado en los ejemplos simulados) y se han logrado detectar de forma satisfactoria los casos anormales.



## Capítulo 4

# Segmentación del sistema vascular retiniano humano a partir de funciones de pertenencia

En este capítulo se va a mostrar una aplicación de la teoría de conjuntos difusos al proceso de imagen, en concreto a la segmentación. La segmentación de una imagen es la división de ésta en zonas con significado, es decir en las zonas correspondientes a los objetos representados. El método de segmentación suele diseñarse en función del objeto que se quiere segmentar y muchos de ellos dependen de ciertos parámetros que no quedan completamente determinados sino que deben ser elegidos por el usuario. La segmentación final es una imagen binaria que depende de los valores de los parámetros y se trata, por tanto, de una posibilidad entre un conjunto de opciones posibles.

En nuestro caso, nos vamos a ocupar del problema de la segmentación de los vasos sanguíneos en imágenes de fondo de ojo. La importancia de disponer de la segmentación de esta estructura retiniana a nivel clínico ya ha sido indicada en la sección 1.1.2. Su importancia también queda reflejada en la cantidad de métodos de segmentación de vasos sanguíneos en la retina que han sido propuestos en la literatura. A lo largo de este capítulo, el interés se centra en tres de esos métodos, sin embargo el objetivo no es evaluarlos ni compararlos, simplemente han sido utilizados para generar funciones de pertenencia a vaso sanguíneo. Por eso su elección no resulta especialmente relevante en el desarrollo de la metodología. En la sección 4.2 se indica cuáles

son los métodos elegidos y en qué consisten. Además también se explica cómo, a partir de cada uno de ellos, ha sido generada una función de pertenencia a vaso distinta. La estructura vascular es considerada a partir de ese momento como un conjunto difuso. Sin embargo, como ya hemos avanzado, el objetivo es conseguir segmentaciones y, para ello, van a aplicarse dos tipos distintos de promedios de conjuntos difusos calculados sobre los  $\alpha$ -cortes. El conjunto promedio del difuso resultante es un conjunto nítido (crisp) o binario y por tanto una segmentación. En la sección 4.3 se dan las dos definiciones de promedio de un conjunto difuso a aplicar: el promedio de Vorob'ev y el de Baddeley-Molchanov. En la sección 4.4 se muestra la aplicación de la metodología a las funciones de pertenencia a vaso obtenidas a partir de tres imágenes de fondo de ojo distintas.

## 4.1. Material

Las imágenes que utilizaremos pertenecen a la base construida por Adam Hoover dentro del proyecto denominado 'Threshold probing of the blood vessel network'. Se puede acceder a dichas imágenes en la dirección <http://www-parl.clemson.edu/stare/probing>. El proyecto anterior a su vez forma parte del proyecto 'STARE' (STRUCTURED Analysis of the Retina) concebido e iniciado en 1975 por Michael Goldbaum con el objetivo general de obtener diagnósticos automáticos de las enfermedades del ojo humano [19],[24].

Se trata de imágenes de fondo de ojo a color y por tanto tienen tres bandas (RGB). A partir de ellas se ha construido una única imagen a niveles de gris ponderando cada banda por las cantidades [0,2989, 0,5870, 0,1140]. Todas las imágenes de esa base vienen acompañadas de un par de segmentaciones manuales realizadas por A. Hoover y V. Kouznetsova. Las segmentaciones manuales nos servirán para poder evaluar el procedimiento que presentamos.

## 4.2. Construcción de la función de pertenencia a vaso a partir de algunos métodos de segmentación

En la literatura se han propuesto procedimientos de segmentación automática del árbol vascular retiniano de muy distinta naturaleza: basados en filtros de acoplamiento, morfológicos, de seguimiento y basados en redes neuronales. El interés en lo que sigue va a estar centrado en tres métodos pertenecientes a los dos primeros tipos. El método basado en filtros de acoplamiento utilizado



es el diseñado por Chaudhuri et al. [7] y los métodos basados en morfología son los propuestos por F. Zana y J.C. Klein [52, 51] y por C. Heneghan et al. [23, 17]. Otros métodos de segmentación de vasos no considerados aquí pero que pueden ser incorporados del mismo modo son [47, 54, 41].

El procedimiento diseñado por Chaudhuri et al. consiste en la aplicación de un filtro de acoplamiento con distintas orientaciones. En cada píxel nos quedamos con la respuesta máxima del filtro, que se producirá cuando la función kernel de dicho filtro esté alineada con la dirección del vaso, en el caso de que el píxel forme parte de alguno. En el diseño de la función kernel se consideran dos importantes características de los vasos sanguíneos: su estructura lineal a trozos y la forma gaussiana de sus perfiles transversales. Teniendo en cuenta tales características, construyen una función kernel en dos dimensiones cuya expresión viene dada por

$$K(x, y) = -\exp(-x^2/2\sigma^2) \text{ para } |y| \leq L/2, \quad (4.1)$$

donde  $L$  y  $\sigma$  son cantidades fijas. La aplicación de este procedimiento da como resultado una imagen a niveles de gris sobre la cual, finalmente, se aplica el umbral automático diseñado por Otsu [35].

El umbral de Otsu es un umbral automático que consiste en la obtención del nivel de gris que divide al conjunto total de píxeles en dos clases de forma que la varianza entre clases sea máxima respecto a la varianza intra-clase o bien respecto a la varianza total.

Los procedimientos de Zana y Klein [52] y Heneghan et al. [23] son de tipo morfológico. La Morfología Matemática es una potente herramienta cuando la forma del objeto de interés a segmentar es conocida a priori. Esta situación se da en el caso del sistema vascular ya que su estructura lineal a trozos es una de sus características más importantes. Ambos procedimientos constan de una primera fase en que resultan enfatizadas todas las partes lineales de la imagen, y, en particular, los vasos, mediante aperturas (ver [42], página 89) con elementos estructurantes lineales de tamaño  $L$  previamente establecido por el usuario. Una posterior reconstrucción (ver [42], página 159) de la imagen transformada en la original, produce una nueva imagen donde desaparecen algunos objetos con anatomía distinta a la de los vasos. Sin embargo, esta primera fase no es capaz de eliminar otros objetos que comparten la linealidad con los vasos. En la siguiente fase de ambos métodos se asume que los perfiles transversales a los vasos pueden, en general, aproximarse mediante una forma de campana de Gauss o gaussiana y por tanto se asume, de la misma manera, que tienen ciertas propiedades diferenciales de segundo orden. En este sentido, se diseñan sendos filtros de forma que los objetos que no compartan dichas

propiedades con los vasos puedan ser descartados. Como aproximación a la segunda derivada, Zana y Klein aplican el Laplaciano (ver [18], página 58) a la imagen previamente suavizada con un filtro gaussiano (ver [18], página 53). Heneghan depura un poco más el procedimiento al calcular la segunda derivada en todas las direcciones posibles después de haber aplicado un filtro gaussiano de suavización en la dirección perpendicular. Por último, Zana y Klein aplican un filtro alternante (ver [42], página 215) a la imagen resultante y se quedan como píxeles de vaso con aquéllos cuyo valor es mayor o igual que 1. Heneghan modifica la imagen segunda derivada dando valor cero a los píxeles con segunda derivada negativa y aplica un filtro alternante similar al anterior pero no establece ningún valor de referencia sino que aplica un umbral por histéresis cuyos parámetros han de ser determinados por el usuario y avisa de que esta elección tiene un gran impacto en la segmentación final.

En todos estos métodos, el usuario se ve obligado a tomar decisiones sobre algunos de los parámetros. En el método de Chaudhuri, debe elegirse la longitud  $L$  y la desviación típica  $\sigma$  para el diseño de la función kernel. Posteriormente se aplica un umbral automático. Se va a mostrar cómo en algunos casos este umbral no proporciona la mejor de las segmentaciones posibles. El método propuesto por Heneghan obliga al establecimiento de parámetros, los más importantes de los cuales son la longitud del elemento estructurante, pero, sobre todo, los parámetros del umbral por histéresis final. En el método de Zana-Klein es la longitud del elemento estructurante lineal empleado el único parámetro importante que queda libre.

Veamos cómo eludir la toma de decisiones sobre los parámetros nos conduce a una nueva imagen a niveles de gris donde han sido enfatizados los vasos sanguíneos retinianos y que puede interpretarse como una función de pertenencia a los mismos.

En el caso del método de Chaudhuri y el de Heneghan, consideramos que la función de pertenencia a vaso viene dada por la imagen resultante en el instante previo a la aplicación del umbral con el que culminan ambos procedimientos. La motivación en el caso de Chaudhuri es la comprobación de que, en determinados casos, la imagen binaria proporcionada por el umbral automático no mejoraba el mejor de los  $\alpha$ -cortes obtenidos a partir de la imagen a niveles de gris posterior al filtro (función de pertenencia). En el caso del procedimiento de Heneghan, la elección del umbral final es competencia exclusiva del usuario por lo que el procedimiento se vuelve subjetivo y queda a expensas de una decisión que no tiene por qué ser siempre acertada.

Como resultado de aplicar el procedimiento de Zana y Klein se obtiene siempre una imagen binaria ya que el establecimiento del umbral final no in-

volucra al usuario sino que ha sido fijado por los autores y forma parte del método. En este caso, la obtención de la imagen a niveles de gris se realiza de forma completamente distinta. El método de Zana y Klein, como ya se expuesto, implica una serie de transformaciones morfológicas basadas en un elemento estructurante lineal cuya longitud,  $L$ , debe ser determinada al principio del procedimiento. Una vez tomada esta decisión, el procedimiento da como resultado una imagen binaria (una segmentación). La pregunta que surgiría a cualquier usuario que decide aplicar el procedimiento es cuál es el valor adecuado de  $L$ . Los autores dan alguna indicación, pero lo cierto es que no queda determinado. La imagen a niveles de gris que se considera en este caso es la suma normalizada de las imágenes binarias obtenidas variando el tamaño del elemento estructurante lineal  $L$ , donde  $L \in \{5, 6, \dots, 25\}$ . De esta forma se aprovecharía conjuntamente la información proveniente de segmentaciones con distintos elementos estructurantes lineales a la cual, según nuestro criterio, no hay que renunciar.

Notar que, tanto en el método de Chaudhuri como en el de Heneghan, también hay una elección previa del valor de algunos parámetros. Hemos elegido los valores sugeridos por los autores. Hay que recordar que la imagen de trabajo resultante de ambos métodos (considerada como una función de pertenencia a vaso en este enfoque) es una imagen a niveles de gris y por tanto resulta más robusta ante cambios en estos parámetros. La consecuencia de cambiar algún parámetro hará que aumenten o disminuyan algunos grados de pertenencia asociados a algunos píxeles determinados pero dichos píxeles siguen teniendo influencia en el análisis posterior.

En la figura 4.1 aparecen la función de pertenencia a vaso generada con el método de Chaudhuri, (b); con el método de Zana y Klein, (c) y con el método de Heneghan, (d), todas ellas correspondientes a la imagen representada en (a). Todas las funciones de pertenencia que consideramos están en el rango 0-1, es decir, su nivel de gris más alto es 1 y el más bajo es 0.

En definitiva, por cada método de segmentación utilizado, obtenemos una función de pertenencia a vaso. El objetivo final es el de segmentar el árbol vascular retiniano a partir de esas funciones de pertenencia.

### 4.3. Promedio de un conjunto difuso

En la sección 4.2 se ha puesto de manifiesto que algunos métodos de segmentación del árbol vascular pueden proporcionar como resultado una función de pertenencia al objeto de interés. El mismo razonamiento es aplicable a la mayoría de los métodos de segmentación, independientemente del objeto a

que se apliquen. El problema que surge entonces es el de asociar un conjunto nítido (binario) al conjunto difuso caracterizado por la función de pertenencia obtenida.

Algunos autores ya han estudiado este problema pero fundamentalmente aplicado a números difusos, es decir, difusos definidos en  $\mathbb{R}$ , ver por ejemplo [15].

Este problema va a ser abordado usando resultados existentes en la teoría de conjuntos compactos aleatorios. Algunas buenas referencias sobre el tema son [9, 21, 43]. La forma en que conectaremos el problema con la teoría es pasando a considerar los  $\alpha$ -cortes de la función de pertenencia al difuso como realizaciones de conjunto compacto aleatorio dado tomando un valor  $\alpha$  en  $[0, 1]$  al azar según una distribución de probabilidad dada. El resumen del conjunto difuso por medio de un conjunto nítido no será sino el resumen del conjunto aleatorio asociado. En cualquier modelo estocástico, suele utilizarse la media como cantidad representativa. En el caso que ahora se presenta en que se dispone de un conjunto de realizaciones de un conjunto compacto aleatorio se va a utilizar el concepto de conjunto medio. Hay diferentes definiciones de dicho concepto en la literatura pero tres de las más importantes son la media de Aumann [44], la media de Vorob'ev [44] y la distancia media de Baddeley-Molchanov [4] que, en adelante, pasaremos a denotar media de Baddeley-Molchanov.

Ralescu [36] definió el promedio de un conjunto difuso basándose en la integral de conjuntos definida por Aumann [2] cuya definición recordamos a continuación. Sea  $(\Omega, \mathcal{A}, \mu)$  un espacio de medida no atómico finito. Consideremos entonces  $F : \Omega \rightarrow \mathcal{P}(\mathbb{R}^n)$  una función de conjuntos tal que  $F(\omega) \neq \emptyset$  para todo  $\omega \in \Omega$ . Se asume que  $\{(\omega, x) : x \in F(\omega)\}$  pertenece a la correspondiente  $\sigma$ -álgebra producto. La familia de selectores de  $F$  se define como  $S(F) = \{f : f(\omega) \in F(\omega) \text{ casi por todas partes, con } f \text{ integrable}\}$ . Finalmente, la integral de Aumann viene dada por

$$\int_{\omega} F d\mu = \left\{ \int_{\Omega} f d\mu : f \in S(F) \right\}. \quad (4.2)$$

A partir de la integral de conjuntos de Aumann, Ralescu definió el nivel promedio de un conjunto difuso  $u$  como

$$A_{\alpha} u = \int_{[0,1]} Lu(\alpha) d\alpha = \int_{[0,1]} L_{\alpha}(u) d\alpha \quad (4.3)$$

donde la integral debe ser interpretada como la integral de Aumann del mapa de nivel  $Lu$  respecto a la medida de Lebesgue definida sobre el intervalo uni-

tario  $[0, 1]$ . La notación utilizada se corresponde con la referencia [36] donde pueden encontrarse más detalles al respecto.

A continuación se expone una formulación equivalente de este concepto no contemplada en [36]. Dado un conjunto fijo  $X$ , su *función soporte* se define como

$$h(X, v) = \sup\{v * x : x \in X\} \tag{4.4}$$

donde  $v * x$  denota el producto interno de  $v$  con  $x$  y  $v$  varía sobre la circunferencia unidad centrada en el origen. El promedio de Aumann del mapa de nivel, denotado por  $A_\alpha u$  sería entonces el conjunto convexo cuya función soporte viene dada por

$$h(A_\alpha u, v) = \int_0^1 r_{L_\alpha(u)}(v) d\alpha. \tag{4.5}$$

La prueba se encuentra en [1].

Según lo anterior, se obtiene en primer lugar la función soporte de la media del mapa de nivel y, por tanto, la media del conjunto a continuación.

Si sustituimos en la anterior definición el conjunto fijo  $X$  por un conjunto aleatorio  $\Phi$  siendo  $h(\Phi, \cdot)$  su función soporte (por tanto, una función aleatoria) tal y como se definió en la sección 1.2.4, entonces la media de Aumann de  $\Phi$ ,  $E_\alpha \Phi$ , se define como el conjunto convexo con función soporte dada por la media o el valor esperado de la variable aleatoria  $h(\Phi, v)$ , es decir

$$h(E_\alpha \Phi, v) = E h(\Phi, v). \tag{4.6}$$

Esta definición fue dada por Aumann en [2].

La cuestión clave es que, de la misma forma que hemos utilizado la definición de Aumann del nivel promedio de un conjunto difuso en el marco de la teoría de conjuntos compactos aleatorios, podemos utilizar otras definiciones de media existentes en el contexto de esa teoría.

Los  $\alpha$ -cortes de un conjunto difuso pueden verse como realizaciones de un conjunto compacto aleatorio según quedó establecido en [36]. Sea  $\Phi$  un conjunto compacto aleatorio definido como sigue: se genera un valor uniforme  $\alpha$  en el intervalo unidad  $[0, 1]$  y, dado  $\alpha$ ,  $\Phi = L_\alpha(u)$ . Obviamente, tenemos que asumir que  $L_\alpha(u)$  es cerrado para todo  $\alpha$ , es decir,  $u$  es una función semi continua superiormente. Si  $\Phi$  se define de esta forma, entonces la media de Aumann de este conjunto aleatorio (ver [4]) coincide con el nivel promedio propuesto por Ralescu.

**Definición 1 (Promedio de Ralescu-Aumann de un conjunto difuso)**  
*Sea  $u$  un conjunto difuso y sea  $\Phi$  un conjunto compacto aleatorio definido como:  $\Phi = L_\alpha(u)$  donde  $\alpha$  es un valor aleatorio en  $[0, 1]$  con distribución de*

probabilidad  $\mathcal{P}_I$ . Entonces el promedio de Ralescu-Aumann de un conjunto difuso,  $A_{ra}(u)$ , es el conjunto nítido convexo con función soporte dada por

$$h(A_{ra}(u), v) = E_{\Phi} h(\Phi, v) = \int_0^1 h(L_{\alpha}(u), v) \mathcal{P}_I(d\alpha). \quad (4.7)$$

donde  $v \in \partial B(0, 1)$ , siendo  $\partial B(0, 1)$  la circunferencia unitaria.

De esta manera, hemos utilizado una media tomada de la teoría de conjuntos compactos aleatorios para definir el promedio asociado a un conjunto difuso dado. Pero hay otras definiciones de media de un conjunto aleatorio que pueden resultar útiles.

#### 4.3.1. Promedio de Vorob'ev de un conjunto difuso

Notar que la media de Aumann y su aplicación a conjuntos difusos, el promedio de Ralescu-Aumann, tiene un serio inconveniente. Aun no siendo conjuntos convexos los diferentes  $\alpha$ -cortes  $L_{\alpha}(u)$ , la media resultante sí lo es. En el caso de la segmentación de vasos retinianos, los objetos a segmentar son claramente no convexos.

Hay definiciones alternativas en la literatura que no comparten la misma limitación. Dos de ellas son particularmente importantes: la media de Vorob'ev y la media de Baddeley-Molchanov.

En la sección 1.2.4 ha sido introducida la media de Vorob'ev de un conjunto compacto aleatorio. La definición se aplica a un conjunto difuso como sigue.

**Definición 2 (Promedio de Vorob'ev de un conjunto difuso)** *Sea  $u$  un conjunto difuso y sea  $\Phi$  un conjunto compacto aleatorio definido como:  $\Phi = L_{\alpha}(u)$ , donde  $\alpha$  es un valor aleatorio en  $[0, 1]$  con distribución de probabilidad  $\mathcal{P}_I$ . El promedio de Vorob'ev del conjunto difuso  $u$ ,  $A_v(u)$ , se define como*

$$A_v(u) = L_{\alpha^*}(u), \quad (4.8)$$

donde el  $\alpha$ -corte  $L_{\alpha^*}(u)$  verifica la ecuación

$$\nu(L_{\alpha^*}(u)) = \int_{[0,1]} \nu(L_{\alpha}(u)) \mathcal{P}_I(d\alpha). \quad (4.9)$$

Si hay diferentes  $\alpha^*$ -cortes verificando la ecuación anterior entonces se toma el valor menor de  $\alpha^*$ .

A continuación se presentan algunos ejemplos sobre el promedio de Vorob'ev de algunos modelos de conjunto difuso comunes.

Un número difuso (conjunto difuso sobre  $\mathbb{R}$ )  $\tilde{M}$  se dice que es  $L - R$  y se denota por  $\tilde{M} = (m_L, m_R, c_1, c_2)_{L,R}$ , si su función de pertenencia tiene la siguiente forma:

$$\mu_{\tilde{M}}(r) = \begin{cases} L\left(\frac{m_L - r}{c_1}\right) & \text{si } r \leq m_L \\ 1 & \text{si } m_L \leq r \leq m_R \\ R\left(\frac{r - m_R}{c_2}\right) & \text{si } r \geq m_R \end{cases}$$

donde  $L$  y  $R$  son las funciones de referencia, es decir,  $L, R : [0, +\infty[ \rightarrow [0, 1]$  son funciones estrictamente decrecientes en el soporte de  $\tilde{M}$  semi continuas superiormente tales que  $L(1) = R(1) = 0$ . Por ejemplo, cuando  $L(x) = R(x) = \max\{0, 1 - x\}$ ,  $\tilde{M}$  es un número difuso trapezoidal. Si, además,  $m_L = m_R$  entonces  $\tilde{M}$  es un número difuso triangular. En la figura 4.3 aparecen algunos ejemplos de este tipo de números difusos.

Consideremos un número difuso  $L - R$ ,  $u = (m_L, m_R, c_1, c_2)_{L,R}$  con funciones de referencia

$$L(x) = \max\{0, 1 - x^p\}, \text{ y } R(x) = \max\{0, 1 - x^q\}, \quad (4.10)$$

con  $p, q \geq 1$ . Para este conjunto difuso, se cumple que

$$\nu(L_\alpha(u)) = m_R - m_L + c_2 R^{-1}(\alpha) + c_1 L^{-1}(\alpha), \quad (4.11)$$

donde  $L^{-1}(\alpha) = (1 - \alpha)^{1/p}$  y  $R^{-1}(\alpha) = (1 - \alpha)^{1/q}$ . Finalmente,

$$\int_0^1 \nu(L_\alpha(u)) d\alpha = m_R - m_L + \frac{c_2 q}{1 + q} + \frac{c_1 p}{1 + p}. \quad (4.12)$$

En este caso, hay que encontrar  $\alpha^*$  cumpliendo que

$$m_R - m_L + c_2 R^{-1}(\alpha^*) + c_1 L^{-1}(\alpha^*) = m_R - m_L + \frac{c_2 q}{1 + q} + \frac{c_1 p}{1 + p}. \quad (4.13)$$

El siguiente resultado se obtiene de forma inmediata.

**Proposición 1** *Sea  $u = (m_L, m_R, c_1, c_2)_{L-R}$  un número difuso con funciones de referencia*

$$L(x) = R(x) = \max\{0, 1 - x^p\}, \quad (4.14)$$

entonces

$$A_v(u) = L_{\alpha^*}(u) = [m_L - c_1 \frac{p}{1+p}, m_R + c_2 \frac{p}{1+p}] \quad (4.15)$$

donde

$$\alpha^* = 1 - \left(\frac{p}{1+p}\right)^p. \quad (4.16)$$

La figura 4.2 muestra un ejemplo particular de la proposición previa. Los parámetros elegidos han sido  $m_L = -5$ ,  $m_R = 5$ ,  $c_1 = 3$ ,  $c_2 = 4$ ,  $p = 2$ . El segmento horizontal central que aparece en el fondo de la representación es el número difuso esperado dado en la proposición 1.

Notar que  $\alpha^*$  crece con  $p$ . Si  $p = q = 1$ , tenemos un número difuso triangular,  $\alpha^* = 0,5$  y

$$A_v(u) = [m_L - \frac{c_1}{2}, m_R + \frac{c_2}{2}]. \quad (4.17)$$

No se puede obtener una solución cerrada para la media de Vorob'ev del conjunto difuso si  $p$  y  $q$  son distintos pero sí puede utilizarse una solución numérica. Por ejemplo, si  $p = 3$ ,  $q = 2$ ,  $c_1 = c_2 = 1$  entonces tenemos  $\alpha^* = 0,565$ .

El siguiente ejemplo consiste en un número difuso L-R con funciones de referencia exponenciales. La figura 4.3 muestra diferentes ejemplos.

**Proposición 2** Sea  $u = (m_L, m_R, c_1, c_2)_{L-R}$  un número difuso L-R con funciones de referencia dadas por

$$L(x) = R(x) = \max\{0, a_p(1 - \exp[-p(1-x)])\} \quad (4.18)$$

para  $p \neq 0$ , donde  $a_p = \frac{1}{1-e^{-p}}$ . En ese caso, la media de Vorob'ev viene dada por

$$A_v u = [m_L - c_1(1 - \frac{1}{p} + \frac{e^p}{1-e^{-p}}), m_R + c_2(1 - \frac{1}{p} + \frac{e^p}{1-e^{-p}})]. \quad (4.19)$$

**Demostración 1** Tenemos que

$$\begin{aligned} & \int_0^1 \nu(L_\alpha(u)) d\alpha = \\ & \int_0^1 (m_R - m_L + c_1(1 + \frac{\ln((1-\alpha)(1-e^{-p}))}{p}) + c_2(1 + \frac{\ln(1-(1-e^{-p})\alpha)}{p})) d\alpha = \\ & m_R - m_L + c_1 + c_2 + \frac{c_1}{p}(-1-p + \frac{p}{1-e^{-p}}) + \frac{c_2}{p}(-1-p + \frac{p}{1-e^{-p}}). \end{aligned} \quad (4.20)$$



Debemos encontrar el valor de  $\alpha$ ,  $\alpha^*$ , tal que la expresión contenida en la ecuación 4.20 sea igual a esta otra

$$m_R - m_L + c_1 + c_2 + \frac{c_1}{p} \ln(1 - \alpha^*(1 - e^{-p})) + \frac{c_2}{p} \ln(1 - \alpha^*(1 - e^{-p})). \quad (4.21)$$

Podemos emplear la siguiente simplificación:  $\frac{(e^{-p}-1)(1+p)+p}{1-e^{-p}} = \ln(1 - \alpha^*(1 - e^{-p}))$ , y por tanto

$$\alpha^* = (1 - \exp(\frac{pe^{-p}}{1 - e^{-p}} - 1))(\frac{1}{1 - e^{-p}}) \quad (4.22)$$

La figura 4.4 muestra un ejemplo del anterior resultado cuando  $m_L = -5$ ,  $m_R = 5$ ,  $c_1 = 3$ ,  $c_2 = 4$  y  $p = -1$ . El segmento horizontal que aparece en el fondo de la figura es el número difuso esperado obtenido en la anterior proposición.

### 4.3.2. Promedio de Baddeley-Molchanov de un conjunto difuso

Finalmente, Baddeley y Molchanov propusieron un concepto de media de conjuntos basada en su representación por medio de la función distancia. La presentación detallada se encuentra en la referencia original [4]. En la sección 1.2.4 hemos recordado la definición. De nuevo, los  $\alpha$ -cortes van a ser considerados como realizaciones de conjunto compacto aleatorio para definir la media de Baddeley-Molchanov de un conjunto difuso.

#### Definición 3 (Promedio de Baddeley-Molchanov de un conjunto difuso)

Sea  $u$  un conjunto difuso y sea  $\Phi$  el conjunto compacto aleatorio definido como:  $\Phi = L_\alpha(u)$  donde  $\alpha$  es un valor aleatorio en  $[0, 1]$  con distribución de probabilidad  $\mathcal{P}_I$ . Sea  $d$  una función distancia generalizada (ver sección 1.2.4) verificando los requisitos que hemos mencionado anteriormente. Si  $\bar{d}_u(x) = \int_0^1 d(x, L_\alpha(u)) \mathcal{P}_I(d\alpha)$  y  $N_\epsilon = \{x : \bar{d}_u(x) \leq \epsilon\}$ . El promedio de Baddeley-Molchanov del conjunto difuso  $u$ ,  $A_{bm}u$ , se define como  $N_{\epsilon^*}$  donde

$$\epsilon^* = \operatorname{argmin} D(1_{N_\epsilon}, \bar{d}_u) \quad (4.23)$$

donde  $1_A$  es la función indicatriz de  $A$  y  $D$  es una métrica entre funciones.

Consideremos un número difuso triangular simétrico,  $u = (m, c)$ . En primer lugar hemos calculado  $\bar{d}_u(x)$  para  $x \in [m - c, m + c]$  como  $\bar{d}_u(x) =$

$\int_0^1 d(x, L_\alpha(u))d\alpha$ , obteniendo que:  $\bar{d}_u(x) = \frac{1}{2} \frac{(m-x)^2}{c}$  para  $x \in [m-c, m+c]$ . Entonces determinamos  $N_\epsilon = [m - \sqrt{2c\epsilon}, m + \sqrt{2c\epsilon}]$  para  $\epsilon \leq \frac{c}{2}$  y

$$d(x, N_\epsilon) = \begin{cases} m - \sqrt{2c\epsilon} - x & \text{si } x \leq m - \sqrt{2c\epsilon} \\ 0 & \text{si } x \in N_\epsilon \\ x - m - \sqrt{2c\epsilon} & \text{si } x \geq m + \sqrt{2c\epsilon} \end{cases}$$

Y, posteriormente, calculamos la distancia entre las funciones  $\bar{d}_u(\cdot)$  y  $d(\cdot, N_\epsilon)$  como  $D(\bar{d}_u(\cdot), d(\cdot, N_\epsilon)) = \int (\bar{d}_u(x) - d(x, N_\epsilon))^2 dx$ . Sin pérdida de generalidad, podemos asumir que  $m = 0$  y, entonces, tenemos que:

$$\begin{aligned} D(\bar{d}_u(\cdot), d(\cdot, N_\epsilon)) = & \int_{-c}^{-\sqrt{2c\epsilon}} \left( \frac{x^2}{2c} + \sqrt{2c\epsilon} + x \right)^2 dx + \int_{-\sqrt{2c\epsilon}}^{\sqrt{2c\epsilon}} \frac{x^2}{2c} dx + \int_{\sqrt{2c\epsilon}}^c \left( \frac{x^2}{2c} - x + \sqrt{2c\epsilon} \right)^2 dx \\ & - \frac{2}{3} c\epsilon^2 - \frac{2}{3} \sqrt{2c\epsilon}^3 - \frac{4}{3} c^2 \sqrt{2c\epsilon} + 4c^2\epsilon + \frac{4c^3}{15}. \end{aligned} \quad (4.24)$$

Tenemos que encontrar el  $\epsilon^*$  que minimice esta función de  $\epsilon$ , lo cual es equivalente a minimizar  $h_c(\epsilon) := -\epsilon^2 - 2(\epsilon + c)\sqrt{2c\epsilon} + 6c\epsilon$  para  $0 \leq \epsilon \leq \frac{c}{2}$ .

Se ha obtenido como mínimo global de  $h_1(\epsilon)$  el valor  $\epsilon_1 = 0,101021$  y, se ha probado que  $c\epsilon_1$  es el mínimo global de  $h_c(\epsilon)$ . Primeramente, se calcula la primera derivada de  $h_c(\epsilon)$  para  $\epsilon > 0$ :  $h'_c(\epsilon) = \frac{2}{\sqrt{2c\epsilon}}(-\epsilon\sqrt{2c\epsilon} - 3c\epsilon - c^2 + 3c\sqrt{2c\epsilon})$ . En particular, para  $c = 1$ , se tiene que  $h'_1(\epsilon) = \frac{2}{\sqrt{2\epsilon}}(-\epsilon\sqrt{2\epsilon} - 3\epsilon - 1 + 3\sqrt{2\epsilon})$ . Por último, se comprueba que si  $\epsilon^*$  es un cero de  $h'_1(\epsilon)$ , entonces  $c\epsilon^*$  es un cero de  $h'_c(\epsilon)$ :  $h'_c(c\epsilon^*) = \frac{2c}{\sqrt{2c\epsilon^*}}(-\epsilon^*\sqrt{2c\epsilon^*} - 3c\epsilon^* - 1 + 3\sqrt{2c\epsilon^*}) = \frac{2c}{\sqrt{2c\epsilon^*}}h'_1(\epsilon^*)$ . Por ejemplo, si  $c = 1$  entonces  $\epsilon^* = 0,101021$  y  $u_{\epsilon^*} = [m - 0,4495, m + 0,4495]$ .

En [28] se propone una modificación de la media de Baddeley-Molchanov. La motivación es la dificultad que supone a veces encontrar el  $\epsilon^*$  de la definición original. La modificación mencionada puede aplicarse en el contexto de los conjuntos difusos como sigue.

**Definición 4 (Promedio de Baddeley-Molchanov modificado)** *Sea  $u$  un conjunto difuso y sea  $\Phi$  el conjunto difuso aleatorio definido como:  $\Phi = L_\alpha(u)$  donde  $\alpha$  es un valor aleatorio en  $[0, 1]$  con distribución de probabilidad  $\mathcal{P}_I$ . Sea  $d$  una función de distancia generalizada verificando los requerimientos que ya hemos mencionado. Si  $\bar{d}_u(x) = \int_0^1 d(x, L_\alpha(u))\mathcal{P}_I(d\alpha)$  y  $N_\epsilon = \{x : \bar{d}_u(x) \leq \epsilon\}$ . El promedio de Baddeley-Molchanov modificado de un conjunto difuso  $u$ ,  $A_{mbm}u$ , se define como  $N_{\epsilon^*}$  donde*

$$\nu(N_{\epsilon^*}) = \int_{[0,1]} \nu(L_\alpha(u))\mathcal{P}_I(d\alpha). \quad (4.25)$$

En la sección 4.4 cuando hablamos de media de Baddeley-Molchanov nos estamos refiriendo a la dada en la definición 4.

#### 4.4. Aplicación de la metodología

En esta sección van a presentarse las funciones de pertenencia obtenidas a partir de los tres métodos de segmentación mencionados en la sección 4.2 aplicados a tres imágenes concretas de fondo de ojo y, posteriormente, la segmentación a la que dan lugar aplicando los promedios de conjuntos difusos propuestos en la sección 4.3.

En la figura 4.1 aparece una imagen de fondo de ojo (a), acompañada de las funciones de pertenencia que se obtienen al aplicar los métodos de segmentación de Chaudhuri et al., (b); de Zana y Klein, (c), y de Heneghan et al., (d). En los métodos de Chaudhuri y de Heneghan hay que asignar valores a ciertos parámetros para los cuales hemos utilizado los sugeridos por los autores. En el primer caso se ha tomado  $\sigma$  igual a 2 y  $L$  igual 9 donde  $\sigma$  es la desviación típica de la curva gaussiana que constituye la sección transversal de la función kernel y  $L$  es su longitud. Para aplicar el procedimiento de Heneghan, hemos tomado la longitud de los elementos estructurantes lineales igual a 17 ( $L_1$  y  $L_2$ ) y la desviación típica que aparece en los operadores gaussiano y laplaciano del gaussiano,  $\rho$ , igual a 1,75. La función de pertenencia obtenida mediante el método de Zana y Klein es la media aritmética de las segmentaciones obtenidas a partir de 21 elementos estructurantes lineales con longitudes distintas. Las longitudes elegidas son las del conjunto  $\{5, 6, \dots, 25\}$ . Todas las funciones de pertenencia toman valores entre 0 y 1.

Para aplicar la metodología expuesta en la sección 4.3 hay que pasar a considerar los  $\alpha$ -cortes de las funciones de pertenencia a vaso. Se ha tomado  $\alpha \in \{0,05, \dots, 0,95\}$  y, por lo tanto, han resultado 19  $\alpha$ -cortes o conjuntos a promediar.

Van a ser aplicadas dos definiciones distintas de promedio de conjuntos difusos: el promedio de Vorob'ev y el de Baddeley-Molchanov. El promedio resultante con ambas definiciones va a depender de la distribución de probabilidad que asignemos a  $\alpha$ . Como modelo general para la distribución de probabilidad se ha elegido la distribución beta.

La densidad de la distribución de probabilidad beta con parámetros  $a$  y  $b$  ( $a \geq 0$  y  $b \geq 0$ ) viene dada por

$$f(x | a, b) = \begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} & \text{para } 0 \leq x \leq 1, \\ 0 & \text{en otro caso} \end{cases}$$

Se han considerado distintas distribuciones beta sobre  $\alpha$  variando los parámetros  $a$  y  $b$  en el conjunto  $\{1, 2, \dots, 30\}$ . En los ejemplos aplicados es imposible disponer de todos los  $\alpha$ -cortes con  $\alpha \in [0, 1]$ . Ya se ha mencionado que vamos a considerar  $\alpha$  variando en el conjunto  $\{0,05, 0,1, \dots, 0,95\}$ . Para asignar una densidad a esos valores, dados  $a$  y  $b$ , se ha calculado la  $f$  sobre ellos y después se ha dividido por la suma. Es una discretización de la beta continua que se conoce como la distribución beta discreta.

Con la distribución de probabilidad asignada a  $\alpha$  en la definición de los promedios, estamos dando más importancia a unos  $\alpha$ -cortes que a otros pero, aún así, en la media participan todos ellos, aunque al final nos quedemos con uno solo, como ocurre en el caso de la definición de Vorob'ev.

Para encontrar en cada imagen de las tres consideradas y en cada uno de los tres métodos los parámetros de la distribución beta que conduce al mejor de los conjuntos medios, hemos tenido que comparar la imagen promedio (binaria) resultante con las segmentaciones manuales realizadas por Hoover y Kouznetsova sobre las imágenes que estamos utilizando. Se ha utilizado una medida de error basada en los falsos positivos, es decir, los píxeles que han sido segmentados como vaso por el método pero no lo han sido por la segmentación manual, y en los falsos negativos, es decir, los píxeles que no han sido segmentados como vaso por el método pero sí forman parte de un vaso según la segmentación manual. La expresión para el error que ha sido utilizada ha sido la siguiente

$$\frac{1}{2} \frac{eI_h + eII_h}{A_h} + \frac{1}{2} \frac{eI_k + eII_k}{A_k} \quad (4.26)$$

donde  $eI_h$  ( $eII_h$ ) es el error de tipo I (de tipo II) de la segmentación resultante respecto a la segmentación manual realizada por Hoover y  $eI_k$  ( $eII_k$ ) es el error de tipo I (de tipo II) de la segmentación resultante respecto a la segmentación manual realizada por Kouznetsova.  $A_h$  y  $A_k$  son las áreas de la segmentación manual de Hoover y Kouznetsova respectivamente. La característica que hemos tenido en cuenta al elegir esta medida de error es su sencillez.

En la tabla 4.1 aparece el error asociado a los distintos  $\alpha$ -cortes de las funciones de pertenencia mostradas en la figura 4.1. La imagen de fondo de ojo mostrada en 4.1(a) es la imagen 0077 de la base de datos de imágenes de Hoover. Respecto a la función de pertenencia obtenida según Chaudhuri, el  $\alpha$ -corte que menor error produce (es decir, el que mejor segmenta el árbol vascular) es el correspondiente a  $\alpha = 0,5$ . Respecto a Zana y Klein el mejor resultado se obtiene con  $\alpha = 0,6$ . Y, por último, el mejor resultado en el caso de Heneghan se alcanza con  $\alpha = 0,25$ .

Cuadro 4.1: Error cometido al segmentar mediante  $\alpha$ -cortes sobre las funciones de pertenencia 4.1(b) (Chaudhuri), 4.1(c) (Zana-Klein) y 4.1 (d) (Heneghan)

$\alpha$	Método de Segmentación		
	Chaudhuri	Zana and Klein	Heneghan
0.05	6.3347	0.9419	4.4440
0.10	6.3336	0.8040	1.2765
0.15	6.3318	0.6902	0.6799
0.20	6.3260	0.6192	0.5085
0.25	6.2741	0.5746	0.4676
0.30	6.2156	0.5411	0.4834
0.35	5.8696	0.5135	0.5320
0.40	4.0691	0.4928	0.5999
0.45	1.1915	0.4781	0.6575
0.50	0.5871	0.4683	0.7157
0.55	0.6597	0.4632	0.7714
0.60	0.7820	0.4605	0.8259
0.65	0.8741	0.4606	0.8721
0.70	0.9302	0.4695	0.9143
0.75	0.9650	0.4766	0.9548
0.80	0.9841	0.4883	0.9825
0.85	0.9929	0.5240	0.9905
0.90	0.9977	0.5598	0.9979
0.95	0.9995	0.5873	0.9982

Sobre la imagen a niveles de gris generada por el método de Chaudhuri a partir de la imagen 0077, hemos obtenido el umbral de Otsu [35], es decir, hemos acabado de aplicar el procedimiento de Chaudhuri tal y como fue diseñado originalmente. El resultado ha sido un nivel de gris umbral de 0,4941 y un error asociado de 0,5911, es decir más alto que el proporcionado por el  $\alpha$ -corte con  $\alpha = 0,5$  (ver tabla 4.1).

De la misma manera hemos utilizado el umbral por histéresis con la imagen a niveles de gris obtenida mediante el método de Heneghan. En [23], los parámetros del umbral por histéresis son denominados  $t_{low}$  y  $t_{high}$ . A partir de la imagen 0077 y habiendo normalizado los niveles de gris resultantes a valores en el intervalo  $[0, 1]$ , se obtiene que la segmentación con error mínimo es la asociada a  $(t_{low}, t_{high}) = (0,2510, 0,2549)$ , o bien,  $(t_{low}, t_{high}) = (0,2510, 0,2588)$ , con error asociado igual a 0,4673. Notar que el error correspondiente al  $\alpha$ -corte con  $\alpha = 0,24$  es de 0,4671.

Asociado a cada función de pertenencia, hemos calculado un total de 900 conjuntos medios de Vorob'ev, uno para cada elección distinta de parámetros  $a$  y  $b$  de la distribución beta, con  $a, b \in \{1, \dots, 30\}$ . Para cada conjunto medio de los anteriores se ha valorado su error asociado comparando con las segmentaciones manuales y ha resultado que varias distribuciones beta producían el conjunto promedio con error mínimo. Notar que el promedio de Vorob'ev siempre coincide con uno de los  $\alpha$ -cortes debido a la forma en que ha sido definido. Por tanto, en el ejemplo considerado, el promedio con error mínimo debe coincidir con el  $\alpha$ -corte correspondiente a  $\alpha = 0,5$  en el caso del Chaudhuri,  $\alpha = 0,6$  en el caso de Zana y Klein y  $\alpha = 0,25$  en el caso de Heneghan, como así ocurre (ver figura 4.5).

En la figura 4.5 (a), (b) y (c) aparecen respectivamente las medias de Vorob'ev obtenidas sobre las distintas funciones de pertenencia de la figura 4.1 junto con su error asociado. Observar que los errores coinciden con el error más bajo de los observados en cada columna.

Cuando se aborda el cálculo del promedio de Baddeley-Molchanov utilizando la misma batería de distribuciones beta, obtenemos resultados muy malos sobre las funciones de pertenencia obtenidas mediante el método de Chaudhuri y Heneghan pero buenos resultados con la obtenida mediante el método de Zana y Klein. Los malos resultados son debidos, en parte a la configuración especial de la colección de  $\alpha$ -cortes y, en parte, a la limitación impuesta a los parámetros de la distribución beta ya que no pueden ser mayores que 30. Entre los  $\alpha$ -cortes correspondientes a los dos primeros métodos mencionados se encuentran muy pocos que sean realmente informativos, es decir, hay muchos  $\alpha$ -cortes que, o bien segmentan toda la ventana como vaso (los asociados a va-

lores de  $\alpha$  pequeños), o bien no segmentan prácticamente ninguna zona como vaso (los correspondientes a valores de  $\alpha$  altos). En ese caso la distribución de probabilidad de  $\alpha$  debería estar muy concentrada en torno a los valores de interés y, los valores de los parámetros  $a$  y  $b$  de la distribución beta considerados no lo están permitiendo. Cuando no se limita el rango de variación de dichos parámetros, se llega en ambos casos a una distribución beta prácticamente concentrada en un valor de  $\alpha$ , aquél asociado al  $\alpha$ -corte con menor error.

En el caso en que se considera la función de pertenencia obtenida mediante el método de Zana y Klein, los resultados son distintos y bastante más satisfactorios. En la figura 4.6 aparece el resultado de aplicar el promedio de conjunto difusos de Baddeley-Molchanov a la función de pertenencia 4.1(c) cuando  $\alpha$  sigue la distribución beta representada en 4.6(b). El error cometido al segmentar el árbol vascular con el promedio obtenido es de 0.4452. Como puede observarse es menor que el error de cualquiera de los  $\alpha$ -cortes (ver tabla 4.1).

A partir de las segmentaciones obtenidas con distintos tamaños de elemento estructurante lineal hemos obtenido una nueva segmentación que tiene en cuenta la información de todas la anteriores y que las mejora.

En resumen, el conjunto promedio de Baddeley-Molchanov es ventajoso sobre el de Vorob'ev ya que, en el peor de los casos, se limita a elegir un  $\alpha$ -corte como promedio pero en otras ocasiones combina la información de la secuencia de  $\alpha$ -cortes para mejorarla.

Hemos aplicado la misma metodología a otras imágenes ejemplo de la base de Hoover, en concreto a las imágenes 0162 y 0163. En la figura 4.7 está representada la imagen de fondo de ojo 0162 de la base de Hoover, (a), junto con las funciones de pertenencia a vaso generadas mediante el método de Chaudhuri, (b), Zana y Klein, (c), y Heneghan, (d). De nuevo son funciones tomando valores en el intervalo unitario,  $[0, 1]$ . Hemos aplicado el mismo procedimiento de cálculo del promedio del conjunto difuso y hemos obtenido los conjuntos promedio (segmentaciones) representados en la figura 4.8.

De la misma manera, en la figura 4.9 aparecen la imagen original 0163 y las funciones de pertenencia a vaso generadas con los tres métodos de segmentación utilizados.

Y en la figura 4.10 aparecen los distintos conjuntos promedio obtenidos.

En estos dos ejemplos el promedio de Baddeley-Molchanov aplicado a la función de pertenencia obtenida mediante el método de Zana y Klein también mejora la segmentación producida por el  $\alpha$ -corte mejor. En el caso del promedio de Vorob'ev, el método simplemente elige el mejor de los  $\alpha$ -cortes, el que menor error produce o el que más se parece a las segmentaciones manuales.

## 4.5. Conclusiones

En este capítulo se ha planteado el problema de la segmentación de vasos sanguíneos en imagen de fondo de ojo a partir de funciones de pertenencia. Para cada imagen de fondo de ojo se han obtenido tres funciones de pertenencia a vaso sanguíneo a partir de tres métodos distintos de segmentación como se ha indicado en la sección 4.2. Entonces se han aplicado distintas definiciones de conjunto promedio de un difuso.

Aunque la metodología no se ha planteado con el objetivo de establecer diferencias entre los métodos de segmentación utilizados, conviene notar que es el método de Zana y Klein el que mejores resultados proporciona según la experiencia recogida hasta este punto.

Respecto a los dos promedios de conjunto difuso utilizados, notar que la media de Vorob'ev ha funcionado bien sobre todas las funciones de pertenencia, sin embargo, la media de Baddeley-Molchanov sólo ha funcionado bien sobre la función de pertenencia generada mediante el procedimiento de Zana y Klein. Lo cierto es que las funciones de pertenencia son generadas de forma muy distinta en el método de Zana y Klein respecto a los dos restantes ya que, en ese caso, los  $\alpha$ -cortes constituyen auténticas segmentaciones.

La aplicación del método propuesto implica la utilización de segmentaciones manuales o expertas de las que no se dispone en general. Eso convierte al método en impracticable. Sin embargo, en base a las tres imágenes de fondo de ojo analizadas, se observa que las modas de las distribuciones beta asociadas a cada método están concentradas en zonas distintas. No hay coincidencia, pero sí se aprecia algún tipo de pauta. Es decir, la experiencia apunta a tipologías de distribuciones beta distintas en cada método.

Por otra parte, esta metodología aporta un punto de vista más amplio sobre el problema de la segmentación ya que lo hace más independiente del usuario.



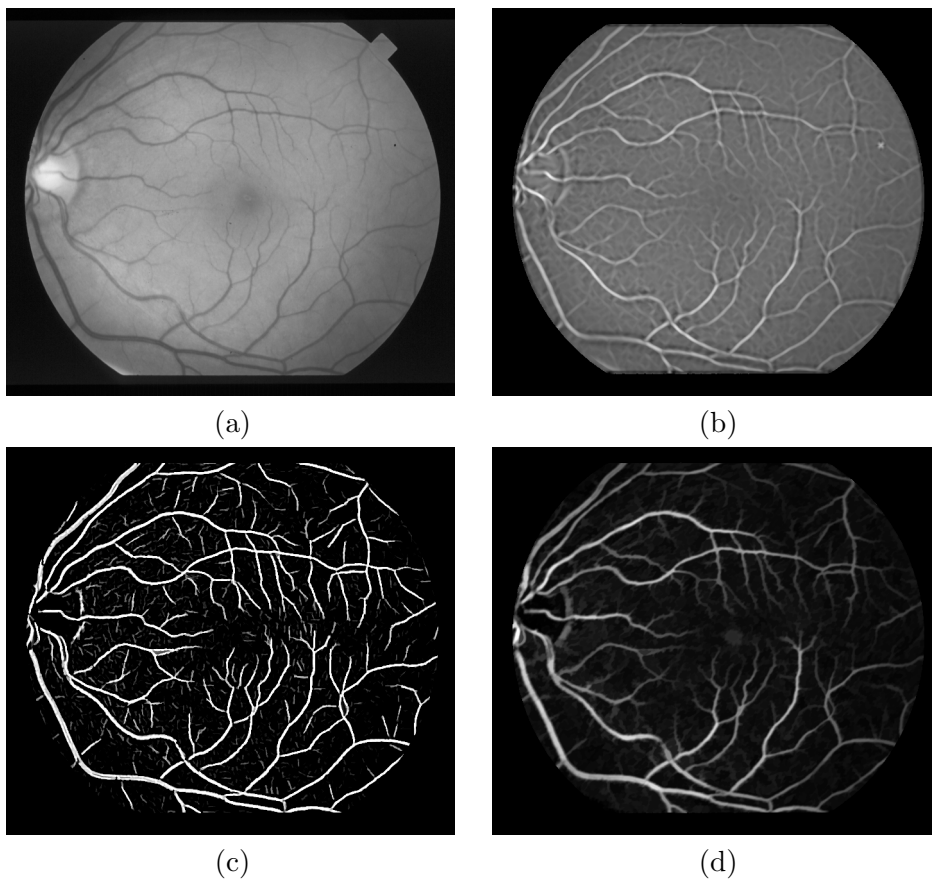


Figura 4.1: Imagen original (a), función de pertenencia a vaso obtenida a partir de (a) mediante: el método de Chaudhuri (b), el método de Zana y Klein (c) y el método de Heneghan (d)

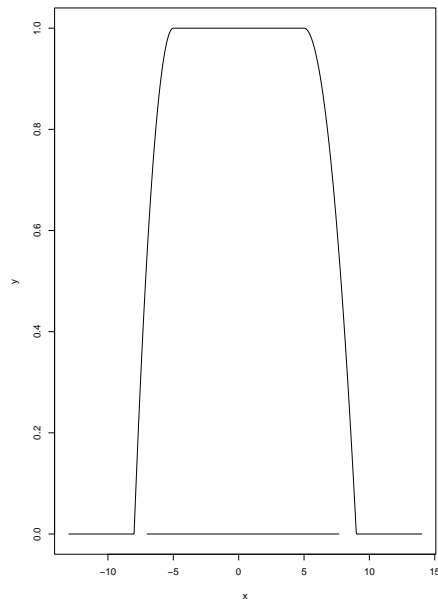


Figura 4.2: Número difuso L-R con funciones de referencia  $L(x) = R(x) = \max\{0, 1 - x^p\}$  para  $p = 2$  y el número difuso esperado representado en la parte inferior de la figura

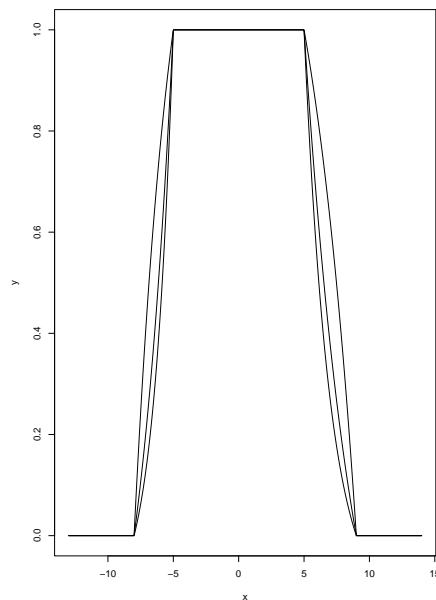


Figura 4.3: Diferentes números difusos L-R con funciones de referencia  $L(x) = R(x) = \max\{0, a_p(1 - \exp[-p(1 - x)])\}$  para  $p = -2, -1, 1$

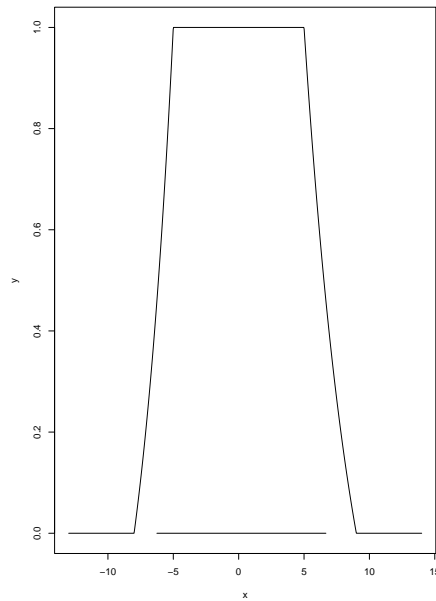


Figura 4.4: Número difuso L-R con funciones de referencia  $L(x) = R(x) = \max\{0, 1 - x^p\}$  para  $p = 2$  y el número difuso esperado en la parte inferior de la figura

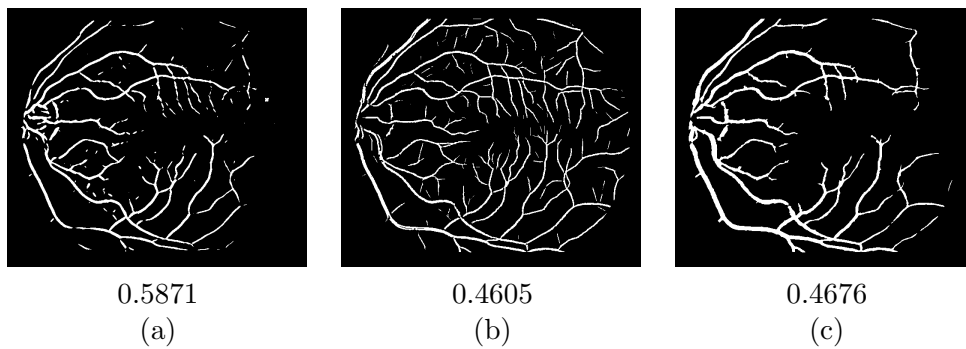


Figura 4.5: Media de Vorob'ev calculada sobre las distintas funciones de pertenencia a vaso obtenidas a partir de la imagen 4.1(a) mediante el método de Chaudhuri, (a); Zana y Klein, (b), y Heneghan, (c). Debajo de cada conjunto medio aparece el error asociado

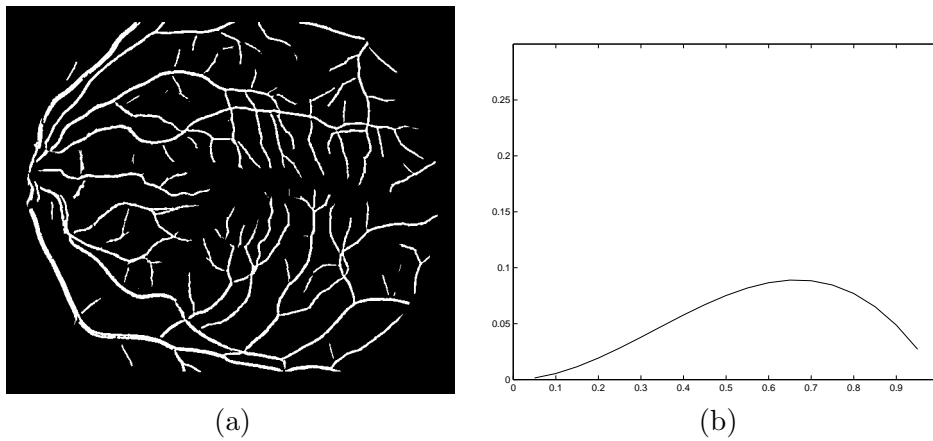


Figura 4.6: Promedio de Baddeley-Molchanov calculado sobre la función de pertenencia a vaso obtenida a partir de la imagen 4.1(a) mediante el método de Zana y Klein y función de densidad de la distribución beta correspondiente

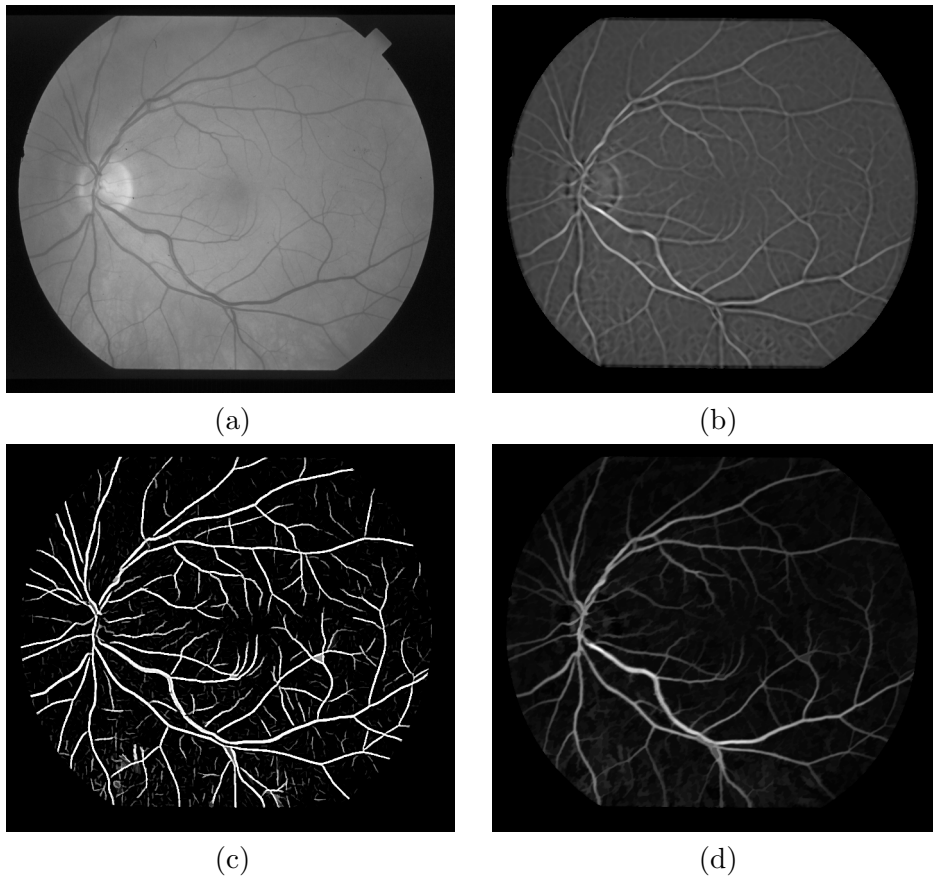


Figura 4.7: Imagen 0162 original (a), función de pertenencia a vaso obtenida a partir de (a) mediante: el método de Chaudhuri (b), el método de Zana y Klein (c) y el método de Heneghan (d)

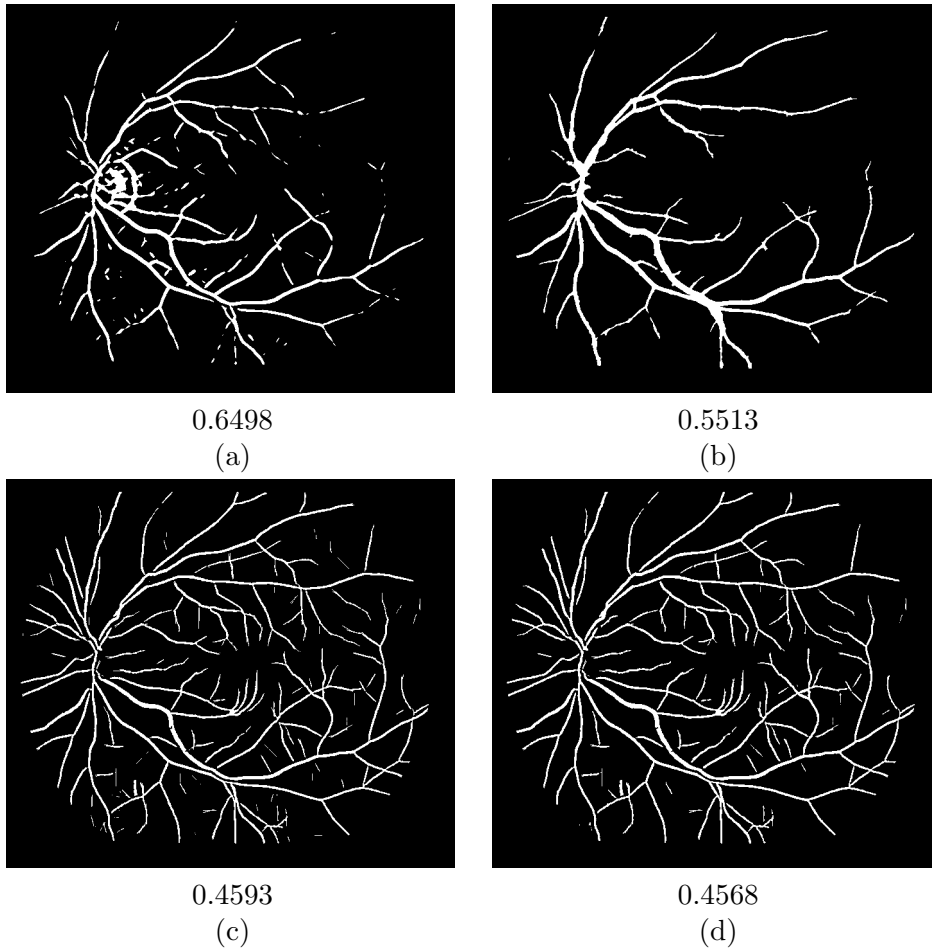


Figura 4.8: Media de Vorob'ev correspondiente a la función de pertenencia a vaso generada mediante el método de Chaudhuri sobre la imagen 4.7(a), (a), media de Vorob'ev correspondiente a la función de pertenencia generada mediante el método de Heneghan, (b), y, por último, medias de Vorob'ev, (c), y de Baddeley-Molchanov, (d), correspondientes a la función de pertenencia generada mediante el método de Zana y Klein. Debajo de cada imagen aparece el error asociado al conjunto promedio correspondiente

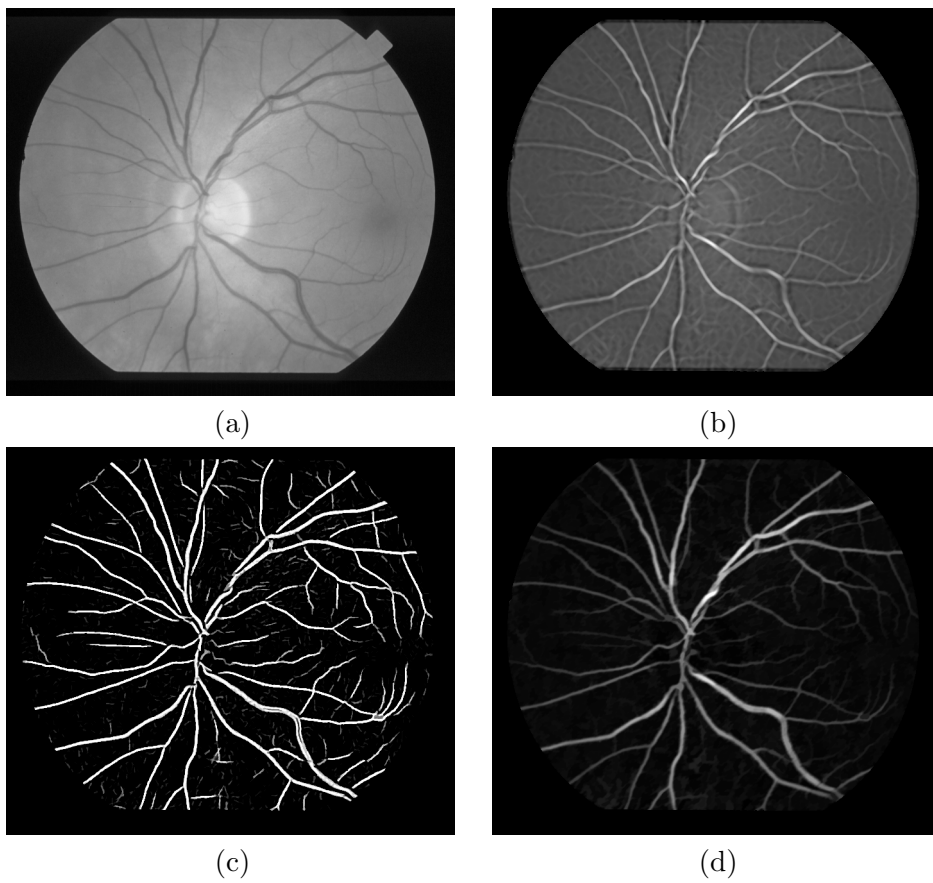


Figura 4.9: Imagen 0163 original (a), función de pertenencia a vaso obtenida a partir de (a) mediante: el método de Chaudhuri (b), el método de Zana y Klein (c) y el método de Heneghan (d)



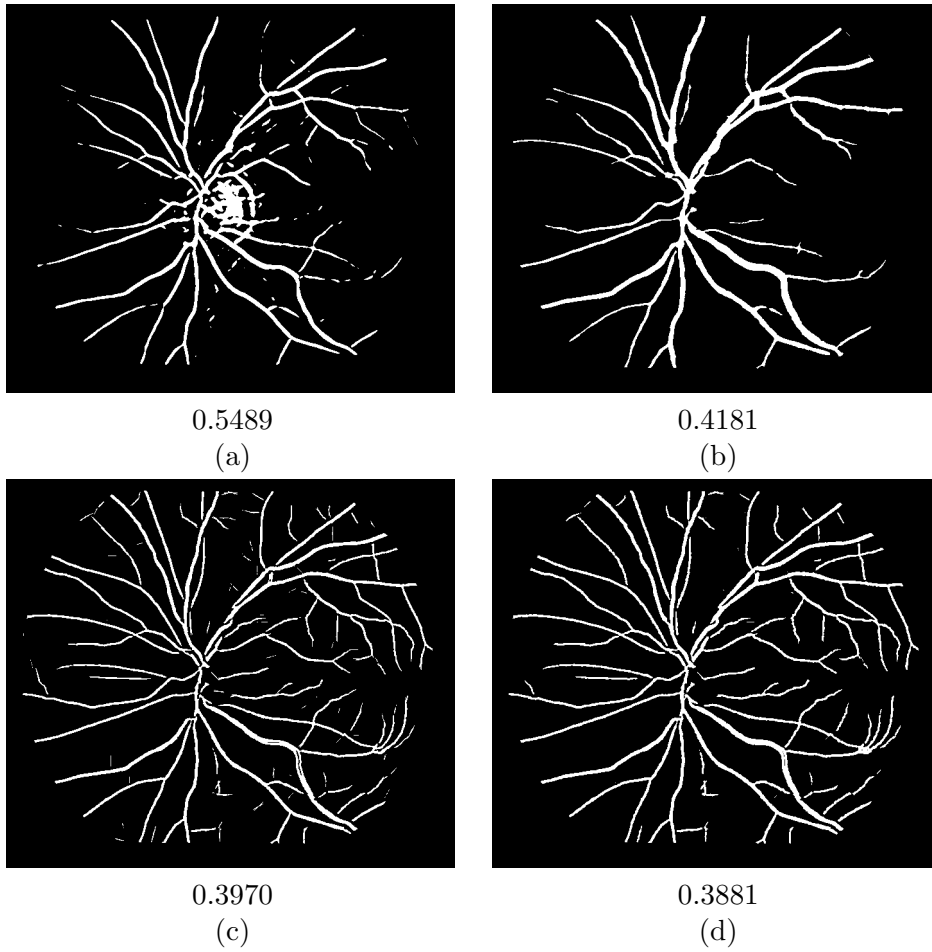


Figura 4.10: Media de Vorob'ev correspondiente a la función de pertenencia a vaso generada mediante el método de Chaudhuri sobre la imagen 4.9(a), (a), media de Vorob'ev correspondiente a la función de pertenencia generada mediante el método de Heneghan, (b), y, por último, medias de Vorob'ev, (c), y de Baddeley-Molchanov, (d), correspondientes a la función de pertenencia generada mediante el método de Zana y Klein. Debajo de cada imagen aparece el error asociado al conjunto promedio correspondiente

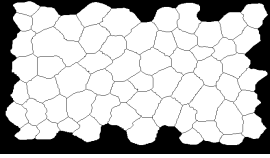
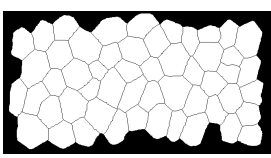
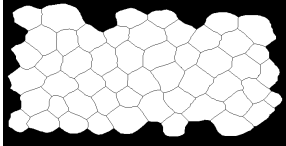


## Apéndice A

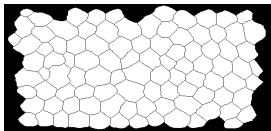
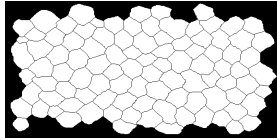
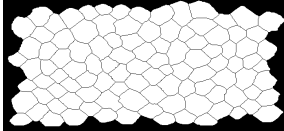
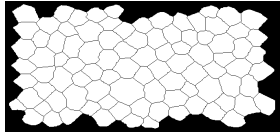
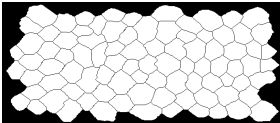
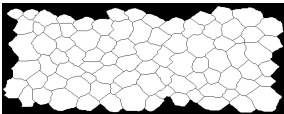
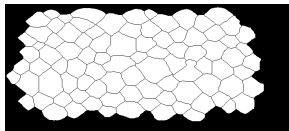
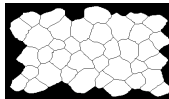
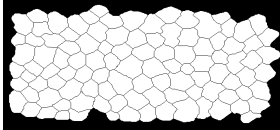
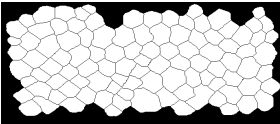
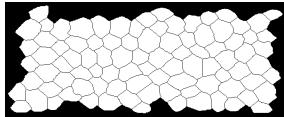
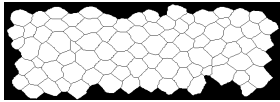
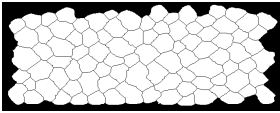
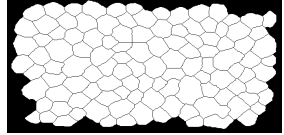
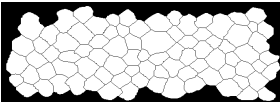
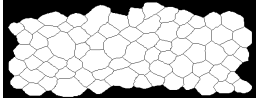
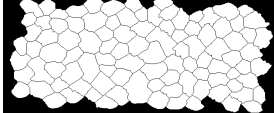
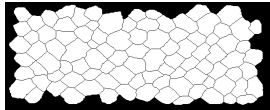
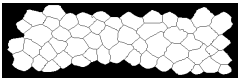
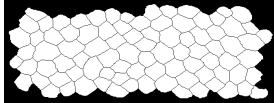
# Agrupaciones de endotelios corneales humanos obtenidos a partir de su descripción granulométrica

### A.1. Grupos de endotelios corneales humanos de individuos entre 20 y 29 años

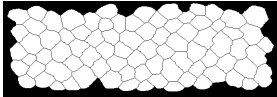
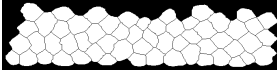
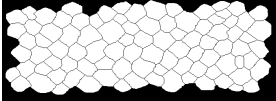
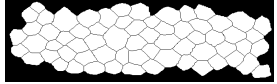

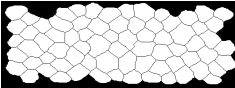
Cuadro A.1: Grupo A

Grupo 5	
	
e25 (1316.1, 61, 35.3)	e77 (1185, 44, 28)
Grupo 6	
	
e136 (3031.7, 75, 25.8)	

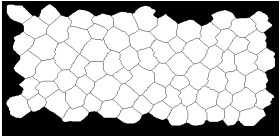
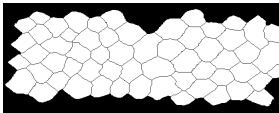
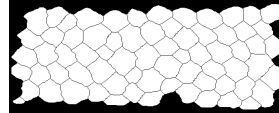
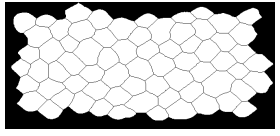
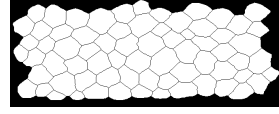
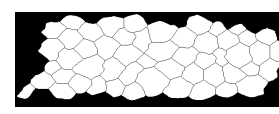
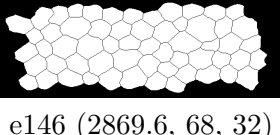


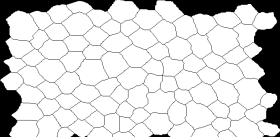
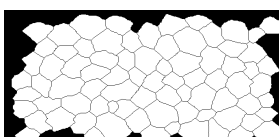
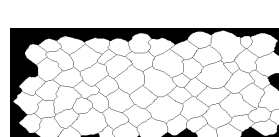

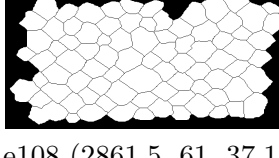

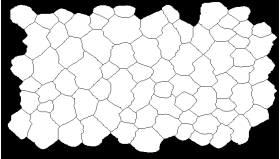
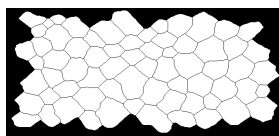
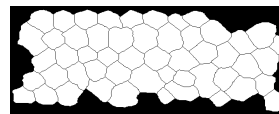
Cuadro A.2: Grupo B

Grupo 2		
		
e7 (NA, 80, NA)	e26 (2384, 73, 26.3)	e92 (2385.6, 59, 29)
		
e8 (2260.3, 59, 30.9)	e91 (2351.6, 58, 34.6)	e125 (3476.2, 52, 39.8)
		
e133 (3864.3, 60, 35.4)	e23 (2261.2, 64, 33.9)	
Grupo 3		
		
e2 (2598.8, 56, 30.6)	e11 (2699.1, 70, 31.9)	e22 (2604.3, 70, 29.1)
		
e52 (2802.8, 67, 31.2)	e93 (2614.7, 46, 32.3)	e10 (2643.6, 45, 34.6)
		
e12 (2741.4, 56, 36.5)	e71 (2612.7, 47, 38.7)	e13 (NA, 40, NA)
		
e40 (2951.7, 49, 33.7)	e69 (2907.9, 53, 36.7)	e16 (2911.8, 60, 26.5)

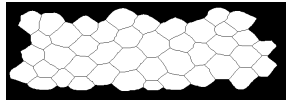
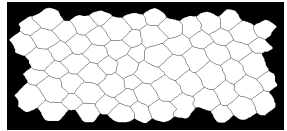
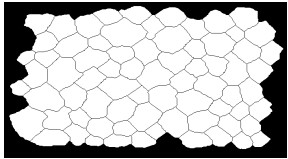
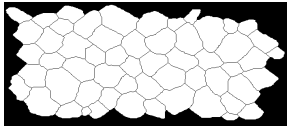
A.1 Grupos de endotelios corneales humanos de individuos entre 20 y 29 años 129

 e41 (2757, 68, 27.9)	 e46 (2783.6, 63, 22.7)	 e72 (2882.8, 53, 31.7)
 e50 (2813.8, 60, 27.8)	Grupo 4	
 e15 (3518.8, 52, 31.6)	 e51 (3018.7, 60, 23.7)	

Cuadro A.3: Grupo C

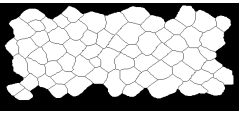
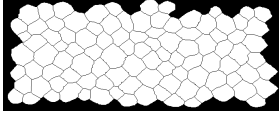
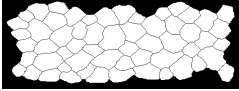
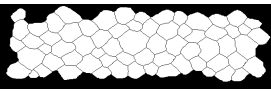
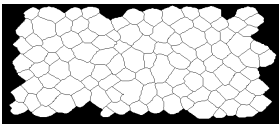
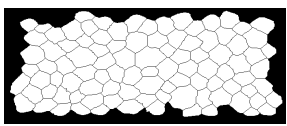
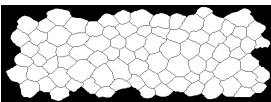
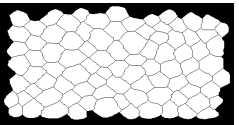



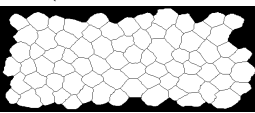
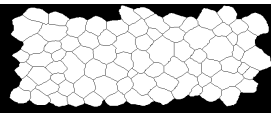
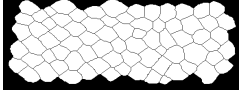


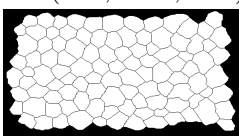
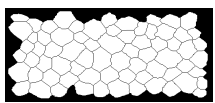
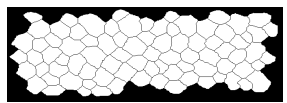
Grupo 1		
		
e4 (1912, 61, 24.5)	e118 (2761.8, 75, 29.6)	e120 (2705.3, 69, 28.6)
		
e141 (2646.3, 69, 22.9)	e137 (3262.6, 65, 31.6)	e147 (2928.6, 70, 36.5)
		
e146 (2869.6, 68, 32)	e140 (2873.8, 68, 24.6)	e148 (2908, 82, 26.2)
		
e107 (2728.5, 73, 34.5)	e109 (2623.5, 58, 34.2)	e114 (2700, 60, 37.3)
		
e121 (2683.5, 54, 34.1)	e108 (2861.5, 61, 37.1)	e122 (2667, 57, 42.4)
Grupo 7		
		
e78 (1501.4, 32, 41.2)	e131 (2658.5, 62, 38.7)	e110 (2337, 69, 37.8)

A.1 Grupos de endotelios corneales humanos de individuos entre 20 y 29 años 131

	
e149 (2694.2, 70, 27.6)	e130 (2734, 77, 23.1)
Grupo 8	
	
e103 (2182.2, 51, 39.5)	e104 (2166.6, 64, 41.1)

## A.2. Grupos de endotelios corneales humanos de individuos entre 30 y 39 años

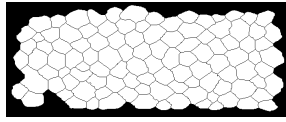
Cuadro A.4: Grupo A

Grupo 1		
		
e5 (2554.9, 38, 31.8)	e49 (2767.9, 60, 30.4)	e63 (2527.9, 44, 36.8)
		
e20 (2898.9, 65, 36.6)	e24 (2789.3, 49, 34.1)	e28 (2742.2, 60, 30.8)
		
e76 (2881.7, 52, 31.9)	e43 (2687.3, 64, 27.5)	e48 (2636.5, 51, 30.7)
		
e66 (2657.1, 43, 42.6)	e89 (2421.1, 33, 40.3)	e42 (2424.2, 59, 24.5)
		
e86 (2485.5, 36, 41.1)		
Grupo 2		
		
e36 (NA, 61.9, NA)	e45 (3026, 50, 45.7)	e54 (2732, 54, 31.4)
		
e80 (2944, 60, 31.1)	e64 (3179.8, 65, 35.3)	e90 (3042.4, 38, 36.4)

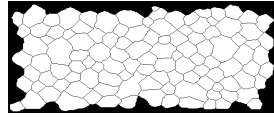


*A.2 Grupos de endotelios corneales humanos de individuos entre 30 y 39 años* 133

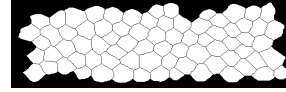
---



e75 (3002.8, 64, 35)

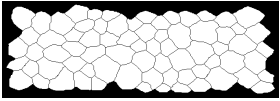

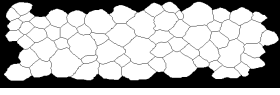
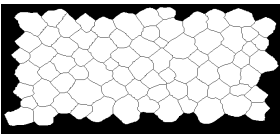
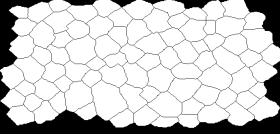
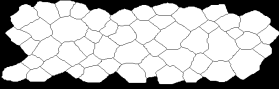
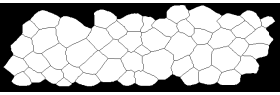
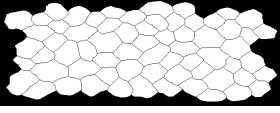
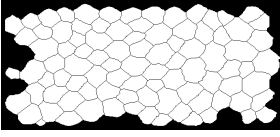


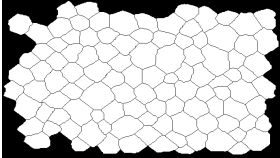
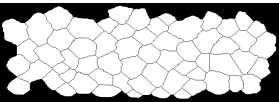
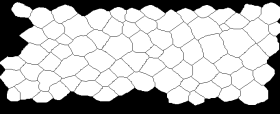



e87 (3206, 45, 45.9)

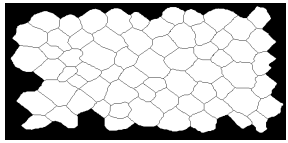
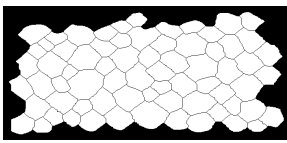
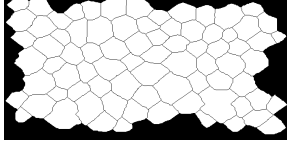
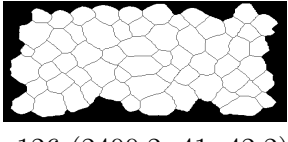
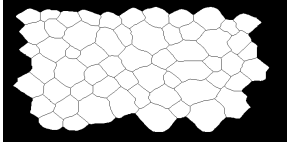


e39 (2877.3, 71, 21.6)

Cuadro A.5: Grupo B

Grupo 3		
		
e19 (NA, 41.7, NA)	e73 (2574.2, 31, 50.5)	
Grupo 4		
		
e74 (2499.7, 34, 58.8)		
Grupo 5		
		
e18 (NA, 42.9, NA)	e106 (2697.2, 40, 35.8)	e124 (2555, 32, 44.9)
		
e134 (2695.6, 38, 42.7)		
Grupo 6		
		
e21 (2107.6, 42, 31.7)	e47 (2122.5, 37, 34.4)	e55 (1876.4, 55, 25)
		
e127 (2693.8, 60, 27.3)	e96 (2931.1, 56, 35)	e97 (2899.4, 65, 37.2)
		
e99 (3038.8, 51, 37)	e100 (2896.4, 54, 31.6)	

Cuadro A.6: Grupo C

Grupo 7		
		
e98 (2324.6, 52, 36.1)	e105 (2542.3, 40, 39.8)	e116 (2431.2, 68, 40)
		
e126 (2400.2, 41, 42.2)		
Grupo 8		
		
e132 (2363.4, 43, 44.5)		



# Bibliografía

- [1] Z. Arstein and R. A. Vitale. A strong law of large numbers for random compact sets. *Annals of Probability*, 3:879–882, 1975.
- [2] R. J. Aumann. Integrals of set-valued functions. *J. Math. Anal. Appl.*, 12:1–12, 1965.
- [3] A. Baddeley and R.D. Gill. Kaplan-meier estimators of distance distributions for spatial point processes. *The Annals of Statistics*, 25(1):263–292, 1997.
- [4] A. Baddeley and I. Molchanov. Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision*, 8:79–92, 1998.
- [5] H.L. Blatt, G.N. Rao, and J.V. Aquavella. Endothelial cell density in relation to morphology. *Invest Ophthalmol Vis Sci*, 18(8):856–9, Aug 1979.
- [6] W.M. Bourne, Nelson L.R., and Hodge D.O. Central corneal endothelial cell changes over a ten-year period. *Invest Ophthalmol Vis Sci*, 38(3):779–82, Mar 1997.
- [7] S. Chaudhuri, S. Chatterjee, N. Katz, M. Nelson, and M. Goldbaum. Detection of blood vessels in retinal images using two-dimensional matched filters. *IEEE Transactions on medical imaging*, 8(3):263–269, September 1989.
- [8] J. Cotinat. Microscopie spéculaire de l' endothelium cornéen. *J Fr Ophthalmol*, 22(2):255–261, 1999.
- [9] Noel A.C. Cressie. *Statistics for Spatial Data. Revised Edition*. John Wiley and Sons, New York, 1993.

- [10] A.C. Davison and D.V. Hinkley. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.
- [11] M.E. Díaz and G. Ayala. Measuring the spatial homogeneity in corneal endotheliums by means of a randomization test. In A. Kuba, Šámal, and A. Todd-Pokropek, editors, *Information Processing in Medical Imaging*, number 1613 in Lecture Notes in Computer Science, pages 418–423. Springer Verlag, 1999.
- [12] M.E. Díaz, G. Ayala, S. Quesada, and L. Martínez-Costa. Testing abnormality in the spatial arrangement of cells in the corneal endothelium by using spatial point processes. *Statistics in Medicine*, 20(22):3429–3439, November 2001.
- [13] P.J. Diggle. *Statistical Analysis of Spatial Point Patterns*. Academic Press, London, 1983.
- [14] J. Domingo, G. Ayala, and M.E. Díaz. Morphometric analysis of human corneal endothelium by means of spatial point patterns. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(2):127–143, 2002.
- [15] D. Dubois and H. Prade. The mean value of a fuzzy number. *Fuzzy Sets and Systems*, 24:279–300, 1987.
- [16] Thomas R. Fleming and David P. Harrington. A class of hypothesis tests for one and two sample censored survival data. *Communications in Statistics. Theory and Methods. A*, 10(8):763–794, 1981.
- [17] J. Flynn. Automatic vessel extraction in digital images. Undergraduate Thesis, Department of Electronic and Electrical Engineering, University College Dublin, Belfield, Dublin, Ireland, 2000.
- [18] C.A. Glasbey and G.W. Horgan. *Image analysis for the biological sciences*. Statistics in Practice. John Wiley and Sons, 1994.
- [19] Michael H. Goldbaum, Norman P. Katz, Subhasis Chaudhuri, Mark Nelson, and Paul Kube. Digital image processing for ocular fundus images. *Ophthalmology Clinics of North America*, 3(3):447–466, September 1990.
- [20] J. Goutsias and S. Batman. *Handbook of Medical Imaging. Volume 2. Medical Image Processing and Analysis*, chapter 4, pages 175–272. SPIE Optical Engineering Press, 2000.

- 
- [21] J. Goutsias, R.P.S. Mahler, and H.T. Nguyen, editors. *Random Sets Theory and Applications*, volume 97 of *The IMA Volumes in Mathematics and its Applications*. Springer, 1997.
- [22] C. Hartmann and L. Bergmann. Specular microscopy: from speculative to spectacular microscopy. *German J. Ophthalmol*, 5:496–503, 1997.
- [23] C. Heneghan, J. Flynn, M. O’Keefe, and M. Cahill. Characterization of changes in blood vessel width and tortuosity in retinopathy of prematurity using image analysis. *Medical Image Analysis*, 6:407–429, 2002.
- [24] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piece-wise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, March 2000.
- [25] L. Kaufman and Peter J. Rousseeuw. *Findings groups in data. An introduction to cluster analysis*. John Wiley and Sons, 1990.
- [26] Elisa T. Lee. *Statistical methods for survival data analysis*. John Wiley, New York, 1992.
- [27] J.M. Lester, J.L. McFarland, R.A. Bursell, S.E. Laing, and J.F. Brenner. Automated morphometric analysis of corneal endothelial cells. *Invest Ophthalmol Vis Sci*, 20(3):407–10, Mar 1981.
- [28] T. Lewis, R. Owens, and A. Baddeley. Averaging feature maps. *Pattern Recognition*, 32:1615–1630, 1999.
- [29] N. Mantel and W. Haenszel. Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22:719–748, 1959.
- [30] G. Matheron. *Éléments pour une théorie des milieux poreux*. Masson & Cie. Éditeurs, 1967.
- [31] G. Matheron. *Random sets and Integral Geometry*. Wiley, London, 1975.
- [32] M. Matsuda, R.V. Yee, and H.F. Edelhauser. Comparison of the corneal endothelium in an american and japanese population. *Arch Ophthalmol*, 103:68–70, 1985.
- [33] Rupert G. Miller, Gail Gong, and Álvaro Muñoz. *Survival analysis*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, 1981.

- [34] P.Ñucci, R. Brancato, M.B. Mets, and S.K. Shevell. Normal endothelial cell density range in childhood. *Arch Ophthalmol*, 108(2):247–8, Feb 1990.
- [35] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-9(1):62–66, January 1979.
- [36] D. Ralescu. Average level of a fuzzy set. In *Proceedings Eighth International Conference IPMU Information Processing and Management of Uncertainty in Knowledge-based Systems*, volume I, pages 190–194. Universidad Politécnica de Madrid, July 2000.
- [37] S.K. Rao, P. Ranjan Sen, R. Fogla, S. Gangadharan, P. Padmanabhan, and S.S. Badrinath. Corneal endothelial cell density and morphology in normal indian eyes. *Cornea*, 19(6):820–3, Nov 2000.
- [38] E.R. Dougherty S. Batman and F.Sand. Heterogeneous morphological granulometries. *Pattern Recognition*, 33:1047–1057, 2000.
- [39] G.A.F. Seber. *Multivariate Obseervations*. John Wiley and Sons, 1984.
- [40] J. Serra. *Image Analysis and Mathematical Morphology*. Academic Press, 1982.
- [41] C. Sinthanayothin, J.F. Boyce, H.L. Cook, and T.H. Williamson. Automated localisation of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *British Journal of Ophthalmology*, 83:902–910, 1999.
- [42] P. Soille. *Morphological Image Analysis. Principles and Applications*. Springer-Verlag, 1999.
- [43] D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and its Applications*. Wiley, Berlin, second edition, 1995.
- [44] D. Stoyan and H. Stoyan. *Fractals, Random Shapes and Point Fields. Methods of Geometrical Statistics*. Wiley, 1994.
- [45] G.D. Sturrock, E.S. Sherrard, and N.S. Rice. Specular microscopy of the corneal endothelium. *Br J Ophthalmol*, 62(12):809–14, Dec 1978.
- [46] S. Theodoridis and K. Koutroumbas. *Pattern Recognition*. Academic Press, 1999.



- [47] Y.A. Tolias and S.M. Panas. A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Transactions on Medical Imaging*, 17(2):263–273, April 1998.
- [48] L. Vincent. *Stochastic Geometry Likelihood and Computation*, volume 80 of *Monographs on Statistics and Applied Probability*, chapter 6, pages 199–283. Chapman & Hall /CRC, 1999.
- [49] R.W. Yee, M. Matsuda, R.O. Schultz, and H.F. Edelhauser. Changes in the normal corneal endothelial cellular pattern as a function of age. *Curr Eye Res*, 4(6):671–8, Jun 1985.
- [50] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
- [51] F. Zana. *Une Approche Morphologique pour les detections et Bayesienne pour le recalage d'images multimodales: Application aux images retiniennes. Analyse Automatique des Images Angiofluorographiques au cours de la Rétinopathie Diabétique*. PhD thesis, Ecole Nationale Supérieure des Mines de Paris, 1999.
- [52] F. Zana and J.C. Klein. Segmentation of vessel-like patterns using mathematical morphology and curvature evaluation. *IEEE Transactions on image processing*, 10(7):1010–1019, July 2001.
- [53] V. Zapater, L. Martínez-Costa, G. Ayala, and J. Domingo. Classifying human endothelial cells based on individual granulometric size distributions. *Image and Vision Computing*, 20(11):783–791, September 2002.
- [54] L. Zhou, M.S. Rzeszutarski, L.J. Singerman, and J.M. Chokreff. The detection and quantification of retinopathy using digital angiograms. *IEEE Transactions on Medical Imaging*, 13(4):619–626, December 1994.