

*Trabajo de Fin de Grado*  
Diseño e implementación de un módulo para la  
anotación de imágenes médicas

Carlos García Huerta (s10m009),  
Grado en Matemáticas e Informática,  
Escuela Técnica Superior de Ingenieros Informáticos (ETSIINF),  
Universidad Politécnica de Madrid

30 de junio de 2014

# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Objetivos</b>	<b>3</b>
2.1. Visión general del proyecto . . . . .	3
2.2. Objetivos específicos de este proyecto de Fin de Grado . . . . .	4
<b>3. Estado del arte</b>	<b>6</b>
3.1. Segmentación de imágenes . . . . .	6
3.2. Anotación de imágenes . . . . .	9
3.2.1. Clasificadores binarios . . . . .	10
3.2.2. Anotación múltiple usando métodos Bayesianos . . . . .	12
<b>4. Metodología</b>	<b>13</b>
4.1. Selección de algoritmo . . . . .	13
4.2. Modelo basado en forma . . . . .	14
4.2.1. Fundamentos teóricos . . . . .	14
4.3. Datasets utilizados . . . . .	16
4.4. Procedimiento y método de evaluación . . . . .	18
<b>5. Resultados</b>	<b>19</b>
<b>6. Conclusión</b>	<b>25</b>

# Índice de figuras

4.1. Ejemplo de radiografía . . . . .	17
4.2. Máscaras de distintos órganos . . . . .	17
5.1. Resultados del pulmón derecho en la misma imagen con distinto pre- procesado . . . . .	23
5.2. Resultados del corazón en los tres tipos de imágenes . . . . .	23

# Índice de cuadros

5.1. Tabla de errores de imágenes 1-3 . . . . .	20
5.2. Tabla de errores de imágenes 4-6 . . . . .	21
5.3. Tabla de errores de imágenes 7-9 . . . . .	22

## Resumen

Se comenzó el trabajo recabando información sobre los distintos enfoques que se le había dado a la anotación a lo largo del tiempo, desde anotación de imágenes a mano, pasando por anotación de imágenes utilizando características de bajo nivel, como color y textura, hasta la anotación automática.

Tras entrar en materia, se procedió a estudiar artículos relativos a los diferentes algoritmos utilizados para la anotación automática de imágenes. Dado que la anotación automática es un campo bastante abierto, hay un gran número de enfoques. Teniendo las características de las imágenes en particular en las que se iba a centrar el proyecto, se fueron descartando los poco idóneos, bien por un coste computacional elevado, o porque estaba centrado en un tipo diferente de imágenes, entre otras cosas.

Finalmente, se encontró un algoritmo basado en formas (Active Shape Model) que se consideró que podría funcionar adecuadamente. Básicamente, los diferentes objetos de la imagen son identificados a partir de un contorno base generado a partir de imágenes de muestra, siendo modificado automáticamente para cubrir la zona deseada. Dado que las imágenes usadas son todas muy similares en composición, se cree que puede funcionar bien.

Se partió de una implementación del algoritmo programada en MATLAB. Para empezar, se obtuvieron una serie de radiografías del tórax ya anotadas. Las imágenes contenían datos de contorno para ambos pulmones, las dos clavículas y el corazón. El primer paso fue la creación de una serie de scripts en MATLAB que permitieran: - Leer y transformar las imágenes recibidas en RAW, para adaptarlas al tamaño y la posición de los contornos anotados - Leer los archivos de texto con los datos de los puntos del contorno y transformarlos en variables de MATLAB - Unir la imagen transformada con los puntos y guardarla en un formato que la implementación del algoritmo entendiera

Tras conseguir los ficheros necesarios, se procedió a crear un modelo para cada órgano utilizando para el entrenamiento una pequeña parte de las imágenes. El modelo obtenido se probó con varias imágenes de las restantes. Sin embargo, se encontró bastante variación dependiendo de la imagen utilizada y el órgano detectado.

## Abstract

The project was started by procuring information about the different approaches to image annotation over time, from manual image annotation to automatic annotation.

The next step was to study several articles about the different algorithms used for automatic image annotation. Given that automatic annotation is an open field, there is a great number of approaches. Taking into account the features of the images that would be used, the less suitable algorithms were rejected.

Eventually, a shape-based algorithm (Active Shape Model) was found. Basically, the different objects in the image are identified from a base contour, which is generated from training images. Then this contour is automatically modified to cover the desired area. Given that all the images that would be used are similar in object placement, the algorithm would probably work nicely.

The work started from a MATLAB implementation of the algorithm. To begin with, a set of chest radiographs already annotated were obtained. These images came with contour data for both lungs, both clavicles and the heart. The first step was the creation of a series of MATLAB scripts to join the RAW images with the annotation data and transform them into a format that the algorithm could read.

After obtaining the necessary files, a model for each organ was created using part of the images for training. The trained model was tested on several of the remaining images. However, there was much variation in the results from one image to another. Generally, lungs were detected pretty accurately, whereas clavicles and the heart gave more problems.

To improve the method, a new model was trained using half of the available images. With this model, a significant improvement of the results can be seen.

# Capítulo 1

## Introducción

Debido al crecimiento de las tecnologías digitales, a día de hoy los datos almacenados en forma de imagen son tan comunes como los que se encuentran en forma de texto. Por tanto, existe una necesidad urgente de herramientas para encontrar información visual según sea necesario.

En las últimas dos décadas, se han producido una gran cantidad de investigaciones sobre consulta de imágenes (del inglés image retrieval [IR]). En general se pueden distinguir tres enfoques:

- Anotación manual de imágenes. Las imágenes se anotan manualmente, para después ser consultadas de la misma forma que los documentos de texto. Sin embargo, es poco práctico debido al gran volumen de imágenes y la subjetividad y ambigüedad que los anotadores puedan aportar a las anotaciones.

- Consulta de imágenes mediante ejemplo (del inglés content-based image retrieval [CBIR]). Las imágenes se anotan en función de características de bajo nivel, como el color, la forma y la textura. Sin embargo, existe una diferencia significativa entre las características de bajo nivel y los conceptos semánticos usados por los seres humanos para interpretar imágenes. Además, al necesitar imágenes de ejemplo, es poco práctico para el usuario de a pie.

- Anotación automática de imágenes. La idea principal es usar técnicas para producir modelos semánticos a partir de una serie de imágenes de muestra, y después usar dichos modelos para anotar nuevas imágenes. Una vez anotadas, las imágenes pueden ser consultadas mediante el uso de palabras clave, similar a la consulta de documentos de

texto. La principal característica de la anotación automática es que ofrece búsqueda por palabras clave basadas en el contenido de la imagen y combina las ventajas de la anotación de textos y de la consulta de imágenes mediante ejemplo. Se partirá de este último enfoque para el desarrollo de este trabajo.



# Capítulo 2

## Objetivos

### 2.1. Visión general del proyecto

Este proyecto Fin de Grado se enmarca en uno de los proyectos de investigación que se están desarrollando en el Laboratorio de Análisis de Datos y Simulación (MIDAS) del Centro de Tecnología Biomédica de la UPM. El objetivo de este proyecto es la anotación de imágenes médicas junto con la anotación de notas clínicas.

El objetivo general de dicho proyecto es conseguir una herramienta que pueda ayudar a los médicos de los hospitales actuales y, por tanto, con el material actual; a estructurar, la información electrónica disponible de los pacientes de manera automática, para su posterior almacenamiento y consulta. Esta herramienta o herramientas están idealmente compuestas de varios módulos que deberían ser capaces de coordinarse entre sí para conseguir el objetivo anterior.

El proyecto consta entonces de las siguientes partes desde un punto de vista general:

- Anotación de imágenes
  - Delineación de las regiones de interés (ROI)
    - Delineación a baja resolución espacial usando imágenes PET
    - Delineación a alta resolución espacial usando imágenes CT a partir de lo obtenido en PET
  - Separación de las distintas ROI obtenidas

- Anotación de las ROI
  - Asignación de posición relativa de cada ROI en el cuerpo
  - Caracterización de cada ROI en base a los resultados obtenidos con cada módulo
- Integración con el módulo de anotación de notas clínicas
- Anotación de notas clínicas
  - Análisis de las notas clínicas del paciente/s
    - Clasificación de las partes del texto analizadas en base a algoritmos de aprendizaje entrenados con notas anteriores
    - Continuación de dicho entrenamiento para perfeccionar el modelo
  - Estructuración de los datos en una base de datos

Así esperamos que los especialistas tengan muchas menos dificultades y poder agilizar su trabajo sin necesidad de que tengan que actualizar su equipo ya que estarían utilizando tan sólo nuevo software. Los aparatos y dispositivos que llevan utilizando hasta ahora serán los mismos.

## 2.2. Objetivos específicos de este proyecto de Fin de Grado

Los objetivos generales del proyecto global han sido señalados en la sección anterior. Este trabajo de fin de grado se centra en la anotación de regiones de interés en las imágenes médicas.

Para conseguir este objetivo principal, se definen los siguientes objetivos específicos:

- O1:** Estudiar diferentes algoritmos de uso en anotación y escoger el que mejor se adapte a las imágenes de trabajo.
- O2:** Implementar el algoritmo o, en su defecto, adaptar una implementación ya existente para su uso con las imágenes de trabajo.
- O3:** Comprobar el correcto funcionamiento del módulo, e idear posibles mejoras

El objetivo de integración del submódulo con los demás módulos que componen el proyecto general a nivel de departamento no será incluido en la lista de objetivos y también quedará como objetivo adicional ya que los demás submódulos todavía no han llegado a ser desarrollados.

Se considera de mayor importancia que este trabajo se centre en revisar más técnicas existentes y evaluar la corrección de los resultados en lugar de cumplir el anterior objetivo adicional u otros posibles objetivos adicionales.

# Capítulo 3

## Estado del arte

En este capítulo se realizará una revisión de lo encontrado en la literatura existente referente al tema de anotación automática de imágenes en general, así como una descripción de las diferentes técnicas revisadas para realizar la anotación de las imágenes de trabajo.

Se puede considerar que el proceso de anotación de imágenes consta de dos partes. En primer lugar se trata de diferenciar las diferentes regiones de la imagen, para, en segundo lugar, realizar una clasificación de dichas regiones acorde con lo que representan en la realidad.

### 3.1. Segmentación de imágenes

La segmentación de imágenes es el proceso de dividir los píxeles de la imagen en diferentes regiones. Cada región tendrá asignadas una serie de características en función del algoritmo utilizado y de la información que se quiera obtener.

En la literatura se han propuesto una gran variedad de técnicas para segmentar imágenes. Sin embargo, no hay una técnica estándar que produzca resultados satisfactorios para todas las aplicaciones. La meta a la que se quiere llegar varía en función del objetivo del estudio a realizar y de los tipos de imágenes de que se disponen. Así, se utilizarán diferentes técnicas dependiendo de las características de las imágenes a analizar.

Como la segmentación automática de imágenes es una tarea complicada, en algunos casos se han utilizado varias técnicas para simplificar el proceso, tales como la división de la imagen en bloques. De estos bloques se extraen características visuales de la imagen. Aunque es poco exigente computacionalmente, no suele extraer los componentes semánticos de la imagen con total precisión, dado que cada bloque suele contener partes de varios objetos.[6]

A continuación, se expondrán brevemente algunos enfoques que se le ha dado a la segmentación de imágenes:

**Segmentación basada en contornos:** este tipo de algoritmos se basan en la utilización de los bordes de los objetos presentes en la imagen para realizar una segmentación de la misma. El método en general consiste en el uso de un operador sobre la imagen que permita encontrar los contornos. El siguiente paso es determinar cuáles de los bordes detectados se pueden utilizar para realizar la segmentación.

Sin embargo, en muchas ocasiones, los bordes encontrados no forman recintos cerrados, lo cual suele ser necesario para realizar una correcta segmentación. Para solucionar dicho problema, se suele hacer uso de técnicas de postprocesado que cierran dichos recintos. El problema está en realizar dicha operación de forma automática sin que resulten situaciones ambiguas. Debido a esto, será necesario recurrir a métodos semiautomáticos o incluso totalmente manuales.[8]

Cootes y Taylor [1] proponen un modelo basado en contornos generados a partir de imágenes de ejemplo, los llamados modelos basados en forma (del inglés active shape models [ASM]). Dichos modelos son entrenados a partir de un conjunto de imágenes anotadas manualmente, a partir de las cuales se genera un contorno base. Al aplicar el modelo a las imágenes a segmentar, el contorno base se deformará en función de los datos obtenidos de las imágenes de entrenamiento, para adaptarlo a la región de interés en la imagen a segmentar.

Este algoritmo no está exento de problemas. El principal es que es muy específico, esto es, la imagen a segmentar tiene que ser muy similar a las de prueba, haciendo viable el algoritmo únicamente en ámbitos determinados.

Por otro lado, dado que el modelo tiene un contorno base, suele detectar el objeto con relativa facilidad sin encontrar muchos problemas con la homogeneidad de los píxeles.

Por último, si lo que nos interesa es una zona determinada de la imagen, se pueden entrenar modelos para obtener el objeto presente en dicha zona, ignorando el resto de

la imagen, lo que también nos permite hacer segmentación y clasificación en un mismo paso.

**Segmentación basada en modelos estadísticos:** entre ellos, uno de los más usados es Blobworld. En este algoritmo, a cada píxel se le asigna un vector formado por características de color, textura y posición. Este vector contiene un total de ocho elementos. La distribución de píxeles en este espacio 8-dimensional se modela como una mezcla de dos a cinco distribuciones gaussianas. El número de las mismas así como el número de regiones se calcula utilizando el algoritmo esperanza-maximización (del inglés expectation maximization (EM)). Una vez que se ha seleccionado un modelo adecuado, se realiza una agrupación de píxeles conectados que pertenezcan al mismo grupo de color o textura. Uno de los problemas de este algoritmo es su coste computacional, debido a que el EM es un algoritmo de optimización. [4]

**Segmentación basada en teoría de grafos:** J. Shi y J. Malik [5] proponen una técnica para segmentar imágenes basada en la teoría de grafos, a la que denominan normalized cut (Ncut). Éste método representa la imagen en forma de grafo, siendo los vértices del grafo cada uno de los píxeles de la imagen y el peso de las aristas las características similares entre los mismos. A partir de ahí, la segmentación se convierte en un problema de partición de grafos.

Un grafo se puede dividir en dos particiones simplemente eliminando las aristas que conectan ambas partes. El grado de desemejanza entre ambas partes se puede representar como la suma de los pesos de las aristas borradas. En teoría de grafos a esa suma se la denomina corte. Así, una bipartición óptima de un grafo es aquella en la se minimiza el valor de dicho corte.

Sin embargo, al utilizar la medida de corte para partir el grafo, se suele tender a separar conjuntos muy pequeños de vértices. Para evitar dicho problema, Shi y Malik [5] proponen utilizar una nueva medida que consiste en calcular el corte en proporción al total de las conexiones en el grafo. Esta medida es lo que se llama Ncut.

Una vez definida dicha medida, el objetivo es dividir el grafo en subgrafos disjuntos de tal manera que la similitud total entre subgrafos distintos sea mínima. El principal problema del método es el siguiente: como la cantidad de píxeles en una imagen es elevada, es necesaria una gran capacidad de cómputo para encontrar la partición óptima. Para mejorar resultados, se han propuesto mejoras tales como una presegmentación de la imagen, utilizando las regiones obtenidas como vértices en el grafo.

**Segmentación por crecimiento de regiones:** las técnicas de crecimiento de regiones consisten en agrupar píxeles con características similares en lugar de buscar diferencias entre los mismos. El algoritmo comienza por un píxel llamado semilla (del inglés seed) a partir del cual se procederá a crecer la región.

La ventaja de este tipo de algoritmos es que se extraen los píxeles con características similares relativas a la intensidad en cada punto, pero además, se tienen en cuenta las características espaciales en la imagen. Por este motivo, las regiones obtenidas con estos procedimientos son siempre regiones conexas, lo cual sería necesario comprobar en otros tipos de métodos, en el caso de que se requieran necesariamente regiones conexas. Por otro lado, una desventaja que presentan es que se obtienen distintos resultados dependiendo de qué semilla se elija. Pero no sólo depende de la semilla, también depende en gran medida del criterio que se utilice para decidir si dos píxeles son homogéneos.

Dentro de dichos algoritmos, uno de los más usados es el llamado JSEG. La idea básica del método es la separación del proceso de segmentación en dos etapas. En primer lugar, los colores de la imagen se discretizan en diferentes clases representativas que posteriormente se utilizarán para diferenciar regiones de la imagen. Esta discretización se realiza sin tener en cuenta la distribución espacial de los colores. A continuación, se reemplaza el valor del color de los píxeles por la etiqueta correspondiente.[6]

En una segunda etapa, se realiza una segmentación espacial sobre el mapa de etiquetas, sin tener en cuenta las similitudes de color entre píxeles. Píxeles con vecinos homogéneos se interpretan como píxeles interiores de una región. Éstos se eligen como semillas, haciendo crecer la región a su alrededor. Como este método comprueba similitud tanto en el color como la textura, las regiones obtenidas tienen características muy homogéneas.

## 3.2. Anotación de imágenes

Una vez que se ha conseguido representar las imágenes en función de sus características de bajo nivel, como colores o texturas, se pueden aprender conceptos semánticos a partir de imágenes de muestra. En primeras aproximaciones se realizaba una revisión manual de los resultados obtenidos con los modelos para enseñar al modelo dichos conceptos. Sin embargo, este método tiene problemas similares a la anotación manual de imágenes, con lo que la tendencia actual es hacia una anotación completamente automática. Así, un algoritmo puede ser entrenado a partir de imágenes anotadas semánticamente para encontrar características coincidentes en nuevas imágenes, y así proceder

a su anotación. [6]

Se pueden considerar dos enfoques diferentes para anotación automática de imágenes. En primer lugar, utilizar clasificadores convencionales para asignar una sólo etiqueta a la imagen. En segundo lugar, asignar varios conceptos a una misma imagen utilizando métodos Bayesianos.

Hay que tener en cuenta que los métodos presentados se pueden aplicar a la imagen de forma global o a cada una de las regiones obtenidas tras la segmentación, permitiendo etiquetar cada zona por separado.

### 3.2.1. Clasificadores binarios

En este enfoque, tras extraer las características básicas de la imagen, éstas se introducen en un clasificador binario del que se obtiene una respuesta positiva o negativa. La salida del clasificador será el concepto utilizado para realizar la anotación de la imagen. Los clasificadores utilizados más habitualmente son máquinas de soporte vectorial (del inglés support vector machine [SVM]), redes neuronales y árboles de decisión.

**Máquinas de soporte vectorial:** son clasificadores supervisados, los cuales han resultado muy eficaces para clasificaciones de datos de muchas dimensiones, especialmente cuando el conjunto de datos de entrenamiento es pequeño. La principal ventaja de SVM frente a otros clasificadores es que maximiza la separación entre las distintas clases.

Un clasificador SVM funciona dividiendo un conjunto de muestras de entrenamiento mediante un hiperplano. Cada muestra se representa como un vector de características y una etiqueta de clase. La forma de aprender del algoritmo consigue separar la mayor cantidad posible de muestras de una clase del resto. SVM es, en resumen, un clasificador binario. Sin embargo, la anotación automática de imágenes requiere un clasificador multiclase. La forma de enfocar dicho problema es entrenar un clasificador para cada concepto, generando cada uno un valor de probabilidad, los cuales se fusionarán en la última fase para generar la etiqueta de clase para una imagen determinada.

Este enfoque funciona bien cuando hay pocos conceptos semánticos. Sin embargo, la calidad de la clasificación se degrada al aumentar el número de conceptos, debido al aumento de ruido y de desequilibrio entre clases. Para mejorar resultados, algunos enfoques introducen un nivel más de decisión, utilizando varios conjuntos de clasificadores



y fusionando la decisión de cada conjunto.

SVM ha demostrado un rendimiento considerable en su uso para anotación de imágenes. La principal ventaja es que SVM puede aprender con un conjunto pequeño de muestras. Sin embargo, SVM tiene problemas de rendimiento en clasificaciones con clases poco equilibradas, un fenómeno muy común en datos de imagen. [6]

**Redes neuronales:** son redes que pueden aprender de ejemplos y tomar decisiones relativas a una nueva muestra. La diferencia con los clasificadores habituales es que pueden aprender varios conceptos de una vez. Una red neuronal consiste en varias capas de nodos interconectados, también llamados neuronas. Las aristas entre neuronas de distintas capas son asignadas un peso. Cada neurona funciona como un elemento de procesamiento, y tiene asignada una función de activación que genera una salida dependiendo de las aristas que tiene conectadas y de la salida de las neuronas de las capas anteriores. Durante el entrenamiento, la red aprende los pesos de las aristas de manera que el error general se minimice. Al clasificar una muestra, cada neurona de salida genera una medida de confianza. La clase correspondiente a la medida de mayor valor indica la decisión para la muestra.

Una red neuronal tiene la ventaja de que las salidas de la última capa se determinan a partir de las capas anteriores, con lo que no necesita ajustes de parámetros o suposiciones sobre la distribución de características. Sin embargo, existen varios problemas con las redes neuronales. Para empezar, la precisión de la clasificación depende del número de capas intermedias y de neuronas. Además, dicho número no se suele justificar en los trabajos de investigación. Escoger las funciones de activación puede resultar problemático. Por último, el entrenamiento tarda mucho tiempo y puede caer en mínimos locales. [6]

**Árbol de decisión:** es una herramienta de toma de decisiones de varias etapas. Dependiendo del número de decisiones que se toman en cada nodo interno del árbol, un árbol de decisión puede ser binario o n-ario. Es diferente de otros clasificadores en el sentido de que mientras que en otros casos la relación entre entrada y salida puede ser difícil de describir, en un árbol de decisión se puede expresar con reglas fáciles de entender, es decir, del tipo si [situación] entonces [consecuencia].

Un árbol de decisión se entrena a partir de un conjunto de imágenes de entrenamiento ya anotadas. Las muestras se representan con una serie de atributos. Durante el entrenamiento, un árbol de decisión se construye dividiendo recursivamente las muestras en conjuntos disjuntos, descartando en cada paso el atributo usado para dividir.

El procedimiento continúa hasta que todas las muestras de un grupo pertenezcan a la misma clase o el árbol alcance su máxima profundidad, cuando no quedan atributos que usar para la división.

Comparado con otros métodos de aprendizaje, un árbol de decisión es fácil de interpretar, y puede aprender con pocas muestras. También es resistente a datos incompletos con con ruido. Sin embargo, los árboles de decisión suelen necesitar que los atributos tengan valores discretos. Aunque hay algoritmos particulares que funcionan con valores continuos, su rendimiento disminuye en gran medida. [6]

### **3.2.2. Anotación múltiple usando métodos Bayesianos**

Al contrario que los clasificadores binarios revisados anteriormente, los métodos con etiquetas múltiples anotan cada imagen con varios conceptos. El concepto de etiquetación múltiple está relacionado con el aprendizaje de ejemplo múltiple y etiqueta múltiple (del inglés multi-instance multi-label [MIML]). En MIML, cada imagen se representa como un conjunto de características o conjunto de regiones. La imagen se anota con una etiqueta conceptual si alguna de las regiones en el conjunto se asocia con la etiqueta. Así, cada imagen es anotada con varias etiquetas.

Un MIML se consigue utilizando herramientas probabilísticas tales como métodos Bayesianos. Dichos métodos funcionan encontrando la probabilidad de que una imagen pertenece a algún concepto en particular, dadas ciertas características de la imagen o región. Esto hace posible que a una imagen se le asignen varios conceptos y permite ordenar imágenes con el mismo concepto según las probabilidades obtenidas. [6]

# Capítulo 4

## Metodología

### 4.1. Selección de algoritmo

Para elegir los algoritmos que usar en la anotación de las imágenes de trabajo, se ha tenido en cuenta las características de las mismas. Las imágenes utilizadas proceden en su mayoría de tomografías computarizadas, o, en su defecto, de radiografías. Esto es, son en escala de grises y, debido a las características físicas del propio cuerpo humano, los objetos en las mismas pueden resultar difíciles de distinguir debido a pocas diferencias de contraste.

Tras esta revisión de los algoritmos de segmentación, se ha escogido un algoritmo basado en contornos. Esto es debido principalmente por su menor dependencia de los colores y las texturas de la imagen.

El algoritmo utilizado en particular es un modelo basado en forma (del inglés active shape model [ASM]), del que se explicará su funcionamiento en más profundidad en el siguiente apartado. Básicamente la segmentación de las zonas se genera a partir de un contorno inicial generado a partir de imágenes de entrenamiento. El contorno se deforma hasta rodear la zona deseada.

El único problema que se podía encontrar es que la posición inicial del contorno tiene que ser colocada cerca de la zona a segmentar. Sin embargo, debido a la regularidad de las imágenes de trabajo, la posición se puede aproximar con cierta facilidad.

Por último, se entrena un modelo por cada elemento a segmentar, por lo que la

anotación de los distintos objetos de la imagen es trivial.

## 4.2. Modelo basado en forma

Los modelos basados en forma (del inglés active shape models [ASM]), fueron introducidos en 1995 por Cootes y Taylor [1]. Estos algoritmos de segmentación se entrenan utilizando imágenes de ejemplo con contornos dibujados manualmente. De dichas imágenes se extraen la media y varianzas de la forma del objeto utilizando análisis de componentes principales (PCA). Así se consigue un contorno inicial que podrá ser deformado para delimitar la zona de interés.

### 4.2.1. Fundamentos teóricos

En este apartado se describirá la construcción de los modelos de forma y apariencia utilizados para representar los diferentes objetos en las imágenes. Como se ha mencionado, el modelo se entrena a partir de imágenes anotadas a mano. Analizando las variaciones de forma y apariencia del conjunto de entrenamiento, se construye un modelo que pueda imitar dichas variaciones. El algoritmo implementado está basado en definiciones extraídas de la tesis de Dirk-Jan Kroon de 2011 [2].

**Construcción del modelo de forma:** para comenzar a construir dicho modelo, el primer paso es dibujar un contorno alrededor del objeto de interés en todas las imágenes de entrenamiento. Tras ello, marcaremos  $n$  puntos destacados a lo largo de dicho contorno. Esquinas en los bordes del objetos o intersecciones con forma de 'T' son buenas elecciones para colocar dichos puntos. Sin embargo, no suelen ser suficientes, con lo cual, aumentaremos la lista marcando puntos espaciados regularmente entre puntos destacados.

Una vez tengamos las posiciones  $(x,y)$  de los puntos, los agruparemos en un vector columna de la siguiente forma:

$$x = (x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)^T \quad (4.1)$$

Si tenemos  $s$  imágenes de entrenamiento, obtendremos  $s$  vectores similares. Hay que tener en cuenta que las formas que buscamos en dichas imágenes estén todas en el mismo marco de coordenadas.

Los  $s$  vectores obtenidos formarán una distribución en el espacio dimensional  $2n$  al que pertenecen. Si podemos hacer un modelo de dicha distribución, podremos generar nuevos ejemplos. El primer paso será simplificar la dimensión de los datos usando PCA, permitiéndonos aproximar cualquiera de los puntos originales mediante un modelo con menos de  $2n$  parámetros. Aplicando PCA, podemos aproximar cualquier punto  $x$  usando:

$$x \simeq \bar{x} + Pb \quad (4.2)$$

Con  $P = (p_1 \dots p_t)$  conteniendo  $t$  vectores propios de la matriz de covarianza y  $b$  siendo un vector  $t$  dimensional de la siguiente forma

$$b = P^T(x - \bar{x}) \quad (4.3)$$

Restringimos  $b \pm m\sqrt{\lambda_i}$  entre 2 y 3, lo que nos permitirá restringir formas que sean posibles entre dos y tres veces la desviación estándar de la distribución de formas del modelo.

**Construcción del modelo de apariencia:** el modelo de forma descrito anteriormente no genera contornos de segmentación, sino que sólo delimita un contorno existente. Para segmentar una imagen también necesitamos un método para incluir la parte de apariencia en el ASM.

El método de Cootes y Taylor [1] comienza calculando vectores normales de todos los puntos destacados del contorno. Después se escoge una muestra de  $k$  puntos a lo largo de la normal espaciados de forma regular, tanto en sentido positivo como en negativo. Esto resulta en un perfil de grises  $s$  de longitud  $2k+1$  por cada punto destacado. Entonces se extrae una derivada de primer orden del perfil:

$$g(i) = s(i+2) - s(i), i = 1, 2, \dots, 2k-1 \quad (4.4)$$

Esto se normaliza con

$$g(i) = g(i)/l, l = \sum g(i) \quad (4.5)$$

A partir del perfil de gris  $g_1, g_2, \dots, g_n$  de cada punto se calcula la media y la matriz de covarianza  $S_g$ . Con estas se construye una función de coste basada en la distancia de Mahalanobis:

$$f(g_i) = (g_i - \bar{g})^T S_h^{-1} (g_i - \bar{g}) \quad (4.6)$$

Dicha función representa la distancia de un nuevo perfil  $g_i$  a la distribución gaussiana de los perfiles de entrenamiento

**Búsqueda de forma:** una vez generados ambos modelos, el algoritmo de búsqueda consiste en los siguientes pasos:

- Se coloca el contorno inicial generado por el modelo en una zona cercana al objeto que queremos delimitar.
- Se toman muestras de variaciones de los niveles de gris en los puntos del contorno, en el sentido de la normal.
- Se busca en la muestra de nivel de gris el punto donde se minimice la distancia de Mahalanobis.
- Cada punto del contorno se coloca en la posición hallada en el paso anterior.
- Se normalizan las nuevas coordenadas del punto, se convierten en parámetros  $b$  del modelo y se limitan los parámetros  $a$
- Se reconvierte  $b$  a coordenadas  $x, y$  de puntos del contorno
- Se repite desde el paso 2 hasta que el desplazamiento de los puntos sea menor que un umbral prefijado o hasta que se alcance el número máximo de iteraciones.

Hay que tener en cuenta que este algoritmo de búsqueda sólo funciona correctamente si la posición del contorno inicial es lo suficientemente cercana al objeto a delimitar.

### 4.3. Datasets utilizados

Para realizar las pruebas del modelo se ha utilizado una serie de radiografías del pecho proporcionadas por la Sociedad Japonesa de Tecnología Radiológica (JSRT). Dichas radiografías están almacenadas en una base de datos creada por la JSRT en cooperación con la Sociedad Japonesa de Radiología (JSR) en 1998 [3]. La base de datos con las imágenes está disponible públicamente para investigación.

En la base de datos hay disponibles un total de 247 imágenes, divididas en dos grupos: 154 imágenes con nódulos en los pulmones y 93 sin ellos. En este caso esta división no es relevante, ya que se trabajará a nivel de órganos.

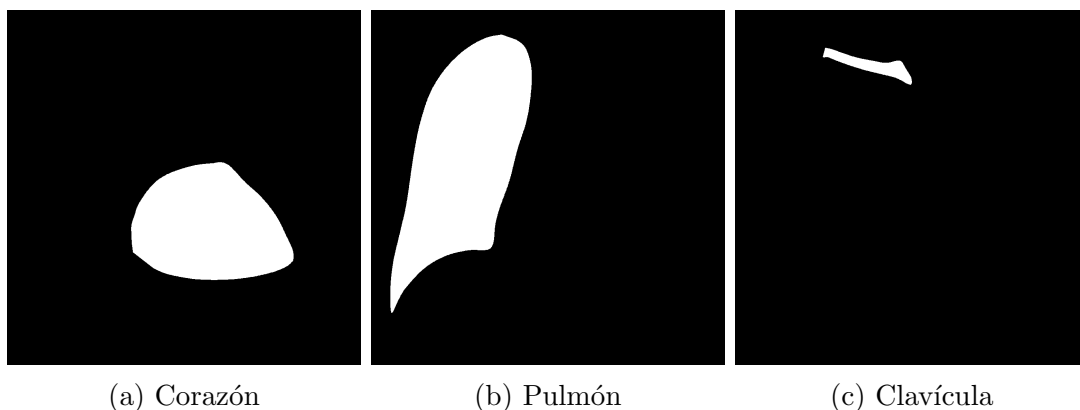
Por último, las imágenes están en formato universal raw, en escala de grises de 12 bits y una resolución de 2048 x 2048 píxeles. Para trabajar con los contornos de que se disponían, se ha reducido la resolución de las imágenes a la mitad, es decir, a 1024 x 1024 píxeles.



Figura 4.1: Ejemplo de radiografía

Con respecto a los contornos utilizados, para cada radiografía se dispone de datos de contorno de cinco zonas de la radiografía: pulmón derecho, pulmón izquierdo, clavícula derecha, clavícula izquierda y corazón.

Para cada órgano se dispone de: datos de puntos destacados del contorno, datos del resto de puntos del contorno y una máscara de imagen con el interior del contorno marcado en blanco y el resto en negro. Todos los datos de contorno mencionados están anotados sobre una imagen de resolución 1024 x 1024 píxeles, siendo ese el motivo de la reducción del tamaño de las imágenes de la base de datos



(a) Corazón

(b) Pulmón

(c) Clavícula

Figura 4.2: Máscaras de distintos órganos

## 4.4. Procedimiento y método de evaluación

El procedimiento seguido para la realización de las diferentes pruebas se divide en varias etapas

**Obtención de las imágenes:** las imágenes de prueba mencionadas en el apartado anterior se descargan de la base de datos.

**Generación de archivos necesarios para el algoritmo:** se realiza una lectura de los archivos de texto con los puntos de contorno y junto con las imágenes descargadas se genera un archivo .mat que la implementación del algoritmo sea capaz de leer. Esta operación se realiza por todas las imágenes una vez por cada elemento a segmentar, en nuestro caso cinco

**Entrenamiento del modelo** utilizando la mitad de las imágenes de que se dispone, se genera un modelo por cada zona de la imagen que sea necesario segmentar

**Ejecución del modelo:** se escogen las imágenes de prueba y se ejecutan todos los modelos con cada imagen, obteniendo en cada caso una máscara de imagen por cada modelo utilizado.

**Generación de los datos a evaluar:** se realiza una comparación entre las máscaras obtenidas con la ejecución de los modelos, obteniendo una imagen que representa la superposición de las máscaras y datos indicando la cantidad de píxeles que coinciden y que no entre ellas.

**Evaluación de los datos:** se realiza una evaluación de coincidencia siguiendo el método explicado a continuación

Para comprobar la variación de resultados, el procedimiento se ha realizado tres veces, siendo la diferencia entre ellas un preprocesado de las imágenes para intentar mejorar los resultados. En el primer caso, las imágenes no se han tocado. En un segundo caso, se ha realizado una ecualización del histograma de la imagen de forma global. Por último, en el tercer caso se ha realizado también una ecualización del histograma, esta vez de forma local.

Para evaluar el algoritmo se ha realizado una comprobación de los píxeles que no coinciden entre las máscaras obtenidas a partir de la ejecución del algoritmo y las máscaras obtenidas por la anotación manual. Cuanto menor diferencia exista entre ambos números, los resultados obtenidos son mejores.



# Capítulo 5

## Resultados

En este capítulo se muestran los resultados obtenidos en la ejecución del algoritmo utilizando el algoritmo decidido y explicado en el apartado anterior siguiendo el proceso descrito en la sección 4.3. Como se ha explicado, se ha entrenado un modelo para cada uno de los tres tipos de imagen: limpia, con ecualización del histograma global y con ecualización del histograma local. Dado que las tablas de datos generadas son de un tamaño considerable, se van a mostrar tablas de resultados de nueve pruebas distintas. En cada una se ha realizado la segmentación de los cinco órganos de que se disponía contorno.

En las siguientes páginas se mostrarán una serie de tablas indicando los resultados obtenidos para cada uno de los tres tipos de imagen. La primera columna de la tabla indicará el órgano segmentado junto con el nombre de la imagen. Cada imagen ocupa dos filas, siendo la superior los falsos positivos, o píxeles por exceso, y la inferior los falsos negativos o píxeles por defecto.

También se mostrarán varios ejemplos de los resultados de superposición de las máscaras. La forma de interpretar dichas imágenes es la siguiente: la zona de la imagen de color verde representa la máscara extraída de la anotación manual del contorno, la zona de color rojo representa la máscara obtenida con el algoritmo y la zona de color amarillo respresenta la superposición de ambas.

A continuación, se muestran las tablas de datos:

Se puede comprobar sobre todo en las imágenes que los resultados para los pulmones son similares a los contornos anotados de forma manual. También se puede apreciar que

		Normal	Histograma	Histograma Local
PulmonDer JPCLN001	Falsos positivos	48646	8867	24614
	Falsos negativos	1683	5554	936
PulmonIzq JPCLN001	Falsos positivos	15229	4956	12171
	Falsos negativos	9747	23417	3736
ClavicDer JPCLN001	Falsos positivos	10573	9106	7978
	Falsos negativos	8714	8821	8783
ClavicIzq JPCLN001	Falsos positivos	4374	4493	3788
	Falsos negativos	7309	7716	6549
Corazon JPCLN001	Falsos positivos	32826	903	88
	Falsos negativos	36860	26940	42892
PulmonDer JPCLN002	Falsos positivos	67299	7779	55152
	Falsos negativos	10148	12320	3278
PulmonIzq JPCLN002	Falsos positivos	37525	14919	28340
	Falsos negativos	3546	8991	2469
ClavicDer JPCLN002	Falsos positivos	1568	2650	3206
	Falsos negativos	754	3973	2451
ClavicIzq JPCLN002	Falsos positivos	3315	3299	2146
	Falsos negativos	1646	4339	1422
Corazon JPCLN002	Falsos positivos	48983	25017	32778
	Falsos negativos	39310	9056	35578
PulmonDer JPCLN003	Falsos positivos	18403	18133	18739
	Falsos negativos	3572	6149	3300
PulmonIzq JPCLN003	Falsos positivos	13635	16533	17541
	Falsos negativos	9082	456	333
ClavicDer JPCLN003	Falsos positivos	782	331	4350
	Falsos negativos	3522	2172	2857
ClavicIzq JPCLN003	Falsos positivos	10561	3317	3756
	Falsos negativos	344	617	751
Corazon JPCLN003	Falsos positivos	41091	34250	21766
	Falsos negativos	33196	36899	26561

Cuadro 5.1: Tabla de errores de imágenes 1-3

		Normal	Histograma	Histograma Local
PulmonDer JPCLN004	Falsos positivos	17521	22335	22688
	Falsos negativos	16410	8684	10406
PulmonIzq JPCLN004	Falsos positivos	1262	5571	2633
	Falsos negativos	15718	4523	20890
ClavicDer JPCLN004	Falsos positivos	3345	1345	1856
	Falsos negativos	3635	4242	4145
ClavicIzq JPCLN004	Falsos positivos	6031	1858	4648
	Falsos negativos	2968	3304	3624
Corazon JPCLN004	Falsos positivos	30773	31173	18949
	Falsos negativos	17533	21589	24618
PulmonDer JPCLN005	Falsos positivos	14387	14577	14021
	Falsos negativos	1004	901	2423
PulmonIzq JPCLN005	Falsos positivos	22155	25381	14544
	Falsos negativos	1209	1079	3950
ClavicDer JPCLN005	Falsos positivos	2375	2344	1046
	Falsos negativos	5680	4850	4788
ClavicIzq JPCLN005	Falsos positivos	2092	1796	544
	Falsos negativos	1398	519	4028
Corazon JPCLN005	Falsos positivos	9084	10368	3006
	Falsos negativos	41107	43571	25142
PulmonDer JPCLN006	Falsos positivos	5269	3846	5407
	Falsos negativos	4788	5077	3797
PulmonIzq JPCLN006	Falsos positivos	4359	2276	6614
	Falsos negativos	20226	26994	13905
ClavicDer JPCLN006	Falsos positivos	866	2361	1108
	Falsos negativos	2178	1898	4105
ClavicIzq JPCLN006	Falsos positivos	4021	6032	3188
	Falsos negativos	987	3229	4038
Corazon JPCLN006	Falsos positivos	67269	66527	56870
	Falsos negativos	9275	0	8016

Cuadro 5.2: Tabla de errores de imágenes 4-6

		Normal	Histograma	Histograma Local
PulmonDer JPCLN007	Falsos positivos	21368	4070	16642
	Falsos negativos	19431	19440	23806
PulmonIzq JPCLN007	Falsos positivos	8882	2663	5016
	Falsos negativos	11555	33717	20731
ClavicDer JPCLN007	Falsos positivos	2152	5788	2328
	Falsos negativos	2374	381	1346
ClavicIzq JPCLN007	Falsos positivos	7037	7833	4155
	Falsos negativos	938	567	4866
Corazon JPCLN007	Falsos positivos	0	10992	5291
	Falsos negativos	55056	37334	50614
PulmonDer JPCLN008	Falsos positivos	783	2366	1807
	Falsos negativos	53256	49937	48564
PulmonIzq JPCLN008	Falsos positivos	1189	1848	2752
	Falsos negativos	41964	40588	34818
ClavicDer JPCLN008	Falsos positivos	1062	1302	27
	Falsos negativos	6817	6608	6908
ClavicIzq JPCLN008	Falsos positivos	4628	4092	588
	Falsos negativos	7099	7496	5769
Corazon JPCLN008	Falsos positivos	27233	38468	19954
	Falsos negativos	31706	32951	26283
PulmonDer JPCLN009	Falsos positivos	8950	7298	8955
	Falsos negativos	4393	3927	3030
PulmonIzq JPCLN009	Falsos positivos	5871	5228	7912
	Falsos negativos	7048	7640	4546
ClavicDer JPCLN009	Falsos positivos	503	375	3376
	Falsos negativos	4546	4790	6731
ClavicIzq JPCLN009	Falsos positivos	3401	1220	1383
	Falsos negativos	1578	5123	1782
Corazon JPCLN009	Falsos positivos	13617	18059	1388
	Falsos negativos	17002	23349	7857

Cuadro 5.3: Tabla de errores de imágenes 7-9

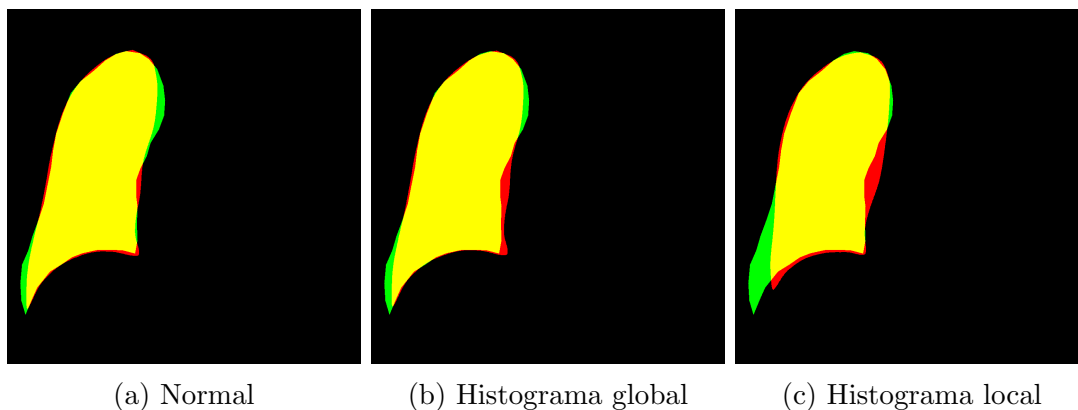


Figura 5.1: Resultados del pulmón derecho en la misma imagen con distinto preproceso

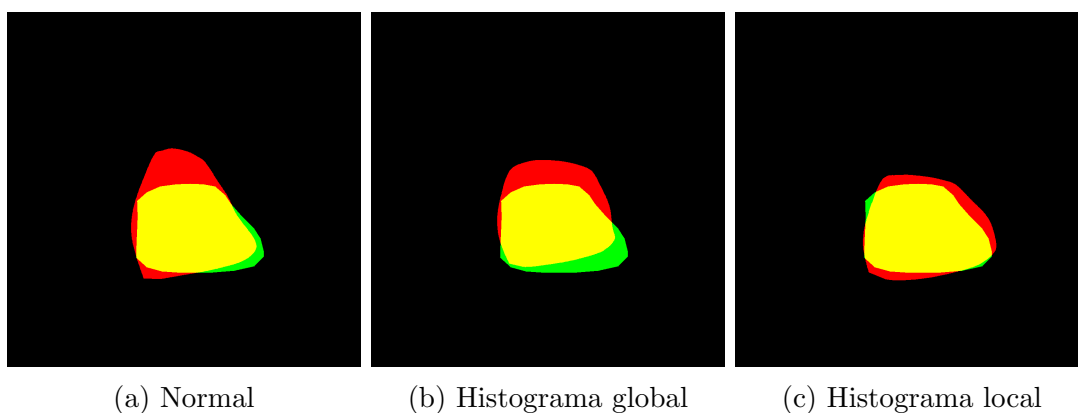


Figura 5.2: Resultados del corazón en los tres tipos de imágenes

la diferencia entre los tres tipos de imágenes utilizados es, en la mayoría de los casos, muy pequeña. Sin embargo, sí es cierto que una observación de las tablas nos puede llevar a la conclusión de que los mejores resultados se obtienen en una imagen sin preprocesar. Esto es debido probablemente a que ya existía una diferencia elevada de contraste entre zonas en la original

En el caso de las clavículas y el corazón, los resultados son menos precisos, debido a esa falta de contraste que sí tiene los pulmones. Tras la elaboración de las tablas de diferencias de píxeles, se puede inferir que los mejores resultados para el corazón se obtienen al realizar una ecualización del histograma de forma local, como se puede apreciar en la imagen de ejemplo. En el caso de las clavículas sucede lo mismo, ya que se consigue un mayor contraste al ecualizar de forma local.

# Capítulo 6

## Conclusión

La anotación automática de imágenes no es un tema trivial. Aunque es una línea de investigación relativamente reciente, con el aumento de la cantidad de archivos en formato gráfico que se generan cada día, cada vez está cobrando una mayor importancia.

Hasta la fecha, se han realizado numerosos estudios sobre algoritmos para intentar resolver, al menos en parte, el problema de la anotación automática. Sin embargo, todavía es un campo abierto, y como tal, no existe una solución correcta.

Dado que de momento es imposible afrontarlo como un problema global, debido a las muchas diferencias que existen de unas imágenes a otras, por el momento, la mejor manera de afrontarlo es dividiendo el problema en partes, en nuestro caso tipos de imágenes, e ir resolviendo los problemas que nos resulten de mayor importancia.

En nuestro caso, se ha decidido realizar una herramienta para conseguir clasificar imágenes médicas, habiendo trabajado con radiografías y tomografías. Dichas imágenes tienen unas características determinadas, con lo que el método para resolverlo será distinto que, por ejemplo, para imágenes RGB de satélite.

Las imágenes con las que trabajamos, por ejemplo, son en escala de grises. Es decir, cualquier algoritmo que utilice variaciones de color como su baza principal en este caso será ineficiente o, en el peor de los casos, no funcional.

También hay que tener en cuenta la composición de las diferentes imágenes que hay. La ventaja que tenemos en nuestro caso, es que son imágenes médicas y, como tal, todas ellas son obtenidas de la misma manera. Por lo tanto la composición y la forma de los

objetos se mantiene aproximadamente constante.

Dadas estas características, se ha concluido en este caso que el modelo de anotación basado en forma es de las mejores opciones para conseguir una anotación precisa de nuestras imágenes.

Si bien es cierto que está lejos de ser perfecto, y que quedan errores por pulir, los resultados obtenidos hasta el momento son bastante satisfactorios. Además, viendo el buen funcionamiento que tiene con las radiografías, es sencillo exportarlo a otro tipo de imágenes médicas.

En resumen, el algoritmo es prometedor, aunque es necesario continuar investigando formas de mejorar los resultados.

# Bibliografía

[1] T. Cootes, C. Taylor, D. Cooper, and J. Graham, “Active Shape Models – Their Training and Application,” *Computer Vision, Graphics and Image Processing*, vol. 61(1), pp. 38–59, January 1995.

[2] Dirk-Jan Kroon, “Segmentation of the mandibular canal in cone-beam CT data,” Signals Systems group, EEMCS Faculty, University of Twente, 2011

[3] Shiraishi J, Katsuragawa S, et al: Development of a digital image database for chest radiographs with and without a lung nodule: Receiver operating characteristic analysis of radiologists’ detection of pulmonary nodules. *AJR* 174; 71-74, 2000

[4] Carson, Chad, et al. ”Blobworld: Image segmentation using expectation-maximization and its application to image querying.” *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.8 (2002): 1026-1038.

[5] Shi, Jianbo, and Jitendra Malik. ”Normalized cuts and image segmentation.” *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.8 (2000): 888-905.


[6] Dengsheng Zhang , Md. Monirul Islam , Guojun Lu, A review on automatic image annotation techniques, *Pattern Recognition*, v.45 n.1, p.346-362, January, 2012

[7] Cootes, Tim, E. R. Baldock, and J. Graham. .An introduction to active shape models. *Image Processing and Analysis* (2000): 223-248.

[8] Isaac Bankman and Joseph Bronzino. *Handbook of Medical Imaging Processing and Analysis*, chapter II Segmentation. Academic Press, 2000.



Este documento esta firmado por

	<b>Firmante</b>	CN=tfgm.fi.upm.es, OU=CCFI, O=Facultad de Informatica - UPM, C=ES
	<b>Fecha/Hora</b>	Tue Jul 01 00:00:04 CEST 2014
	<b>Emisor del Certificado</b>	EMAILADDRESS=camanager@fi.upm.es, CN=CA Facultad de Informatica, O=Facultad de Informatica - UPM, C=ES
	<b>Numero de Serie</b>	630
	<b>Metodo</b>	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sh1 (Adobe Signature)