

# Actualización del SIG citrícola de la Comunidad Valenciana mediante métodos automáticos supervisados

## Semi-supervised scheme to update the citric GIS of the Comunidad Valenciana region

E. Izquierdo<sup>1</sup>, J. Amorós<sup>1</sup>, L. Gómez<sup>1</sup>, J. Muñoz<sup>1</sup>, J. Z. Rodríguez<sup>2</sup>, G. Camps<sup>1</sup> y J. Calpe<sup>1</sup>  
emizver@alumni.uv.es

<sup>1</sup>Dept. Ingeniería Electrónica, IPL, Universidad de Valencia, Valencia, España

<sup>2</sup>Instituto Cartográfico Valenciano, Valencia, España

Recibido el 20 de Mayo de 2008, aceptado el 30 de Mayo de 2008

### RESUMEN

La actualización de los Sistemas de Información Geográfica (SIG) es una necesidad por parte de las administraciones públicas para poder desarrollar políticas de reestructuración de cultivos, gestión eficaz de ayudas o control de superficies y producción de cultivos. Los problemas de la actualización de los SIG son el coste económico y de tiempo que requieren, ya que tradicionalmente se han realizado con técnicas de fotointerpretación y visitas al campo.

En el presente trabajo se ha desarrollado la actualización del SIG citrícola de la Comunidad Valenciana (España) mediante técnicas automáticas a partir de ortoimágenes aéreas de alta resolución con el fin de reducir la intervención humana y optimizar los recursos. La metodología propuesta realiza un análisis orientado a objetos que define los recintos catastrales como entidades individuales, extrayéndose las características principales de cada parcela y clasificándola posteriormente combinando árboles de decisión, máquinas de vectores soporte y redes neuronales artificiales. Por medio de la clasificación automática se ha podido clasificar el 87% de las parcelas procesadas de la Comunidad con un acierto superior al 92% en las tres provincias.

**PALABRAS CLAVE:** Sistema de Información Geográfica (SIG), detección de árboles, extracción/selección de características, clasificación, Árboles de decisión, Máquinas de Vectores Soporte, Redes Neuronales

### ABSTRACT

Updating Geographic Information Systems (GIS) databases is a requirement for public administrations in order to develop crop restructuring policies, funding systems, or crop management.

The main difficulties to update the GIS are the economic and time cost involved, since the update has traditionally relied on human-based photointerpretation and visiting crop fields. This paper presents the strategy that has been developed in order to update the citric SIG of the Comunidad Valenciana (Spain) introducing automatic methods in order to reduce the photointerpretation and field tasks. The proposed methodology carries out an object-oriented analysis that defines the cadastral precinct as the main working entity, the main features are extracted from each precinct and they are then classified with decision trees, support vector machines and multi-layer perceptrons. Using automatic classification schemes, an 87% of processed parcels have been labelled with classification accuracy higher than 92% in all three provinces.

**KEYWORDS:** Geographic Information System (GIS), tree detection, feature extraction/selection, classification, decision tree, support vector machine, neural networks.

## INTRODUCCIÓN

Los Sistemas de Información Geográfica (SIG) consisten en una base de datos (DeMers, 1999) donde figura la información geográfica que mediante un identificador común los asocia a imágenes aéreas o mapas digitales. Están diseñados para analizar la información que contienen de forma que se pueda realizar acciones de planificación y gestión. Se pueden encontrar diversos tipos de SIG, en función del tipo de cultivo del que se quiera obtener información, de los que cabe destacar: SIG cítrica, SIG vinícola, SIG oleícola (Peedell *et al.*, 1998), SIGA (Sistema de Información Agraria) o SIGPAC (Sistema de Información Geográfica de Parcelas Agrícolas), entre otros. En el caso del SIG cítrica, los atributos de tipo agrario que contiene incluyen la clasificación parcelaria indicando si el tipo de cultivo es cítrico o no, aunque también puede incluir información adicional de las parcelas pertenecientes a un territorio como la cota, superficie, etc.

Debido al impacto económico que tiene el cultivo de cítricos en la Comunidad Valenciana (España), el gobierno regional determinó la necesidad de elaborar el SIG cítrico de esta comunidad con objeto de poder desarrollar acciones administrativas como: la gestión eficaz de ayudas, control de la superficie del cultivo de los cítricos, actualización del catastro, diseño de políticas de reestructuración de cultivos, etc. El mantenimiento del SIG cítrico requiere una actualización continua para poder ofrecer un conocimiento de la realidad de este sector a los agentes económicos. El principal problema de la actualización del SIG cítrico radica en la recopilación de datos, debido a que las parcelas cítricas cambian constantemente ya sea porque se abandonan los campos, se cambia el tipo de cultivo o la zona se urbaniza. Por tanto, se necesita recopilar una gran cantidad de información cada poco tiempo.

Tradicionalmente, la generación o actualización de los SIG se realizaba mediante fotointerpretación, y realizando visitas al campo cuando era necesario, haciendo que el proceso fuese muy lento y requiriese gran cantidad de recursos. Actualmente, debido al desarrollo de la tecnología y el aumento de información y resolución de los sensores, las técnicas automáticas de análisis de imágenes aéreas o de satélite comienzan a emplearse en la extracción de información geográfica como la determinación de cultivos o los cambios producidos respecto a la base de datos anterior. Otros factores importantes para su empleo

son la reducción de tiempo y de coste empleado que proporciona el uso de técnicas automáticas en la generación o actualización de los SIG.

El objetivo de este trabajo es la actualización del SIG cítrico de la Comunidad Valenciana (CV) a escala 1:10000, mediante un sistema de clasificación automático de cada parcela catastral. Las administraciones territoriales exigen una alta fiabilidad en la clasificación parcelaria al ser éste un factor crítico en el desarrollo de políticas de reestructuración de cultivos, la gestión eficaz de ayudas o el control de superficies y producción de cultivos. La aproximación utilizada combina las técnicas de procesamiento digital de imagen y de aprendizaje basado en muestras (*machine learning*) con técnicas de fotointerpretación para asegurar la fiabilidad requerida mientras se reduce el número de parcelas fotointerpretadas en función de los recursos humanos disponibles y un grado de error asumible. En una primera etapa las parcelas se clasifican asignándoles un grado de fiabilidad y aquellas que no lo superan son etiquetadas para revisar. En la segunda etapa, el Instituto Cartográfico Valenciano (ICV) realiza la fotointerpretación de las parcelas a revisar.

La primera sección describe el área de estudio y el material utilizado en este trabajo. La siguiente sección detalla la metodología empleada, donde se incluye la descripción de las tipologías de parcela más comunes; las técnicas y selección de características empleadas; y en último lugar, la metodología de clasificación de las parcelas cítricas. En la tercera sección se detallan los resultados obtenidos en la actualización del SIG de la CV, finalizando con las conclusiones.

## MATERIAL

### Área de estudio

El área de estudio es la Comunidad Valenciana, comunidad autónoma situada al este de la Península Ibérica (España), contiene 542 municipios divididos en tres provincias Castellón, Alicante y Valencia. En 2006, la CV tenía 598.905 ha de cultivo, datos obtenidos del documento Informe del Sector Agrario Valenciano del año 2006 realizado por Conselleria d'Agricultura, Pesca i Alimentació de la Generalitat Valenciana, de las cuales el 31% estaban destinadas al cultivo de cítrico repartidas en las tres provincias, aunque más de la mitad de las hectáreas se encontraban en la provincia de Valencia (Tabla 1).

CULTIVO	Alicante	Castellón	Valencia	Comunidad Valenciana
Árboles secano	102.348 ha	77.595 ha	128.894 ha	308.837 ha
Cítricos	38.101 ha	41.984 ha	104.737 ha	184.822 ha
Herbáceos	20.637 ha	13.841 ha	41.780 ha	76.258 ha
Frutales	5.950 ha	3.096 ha	19.569 ha	28.615 ha
Total	167.036 ha	136.516 ha	294.980 ha	598.532 ha

**Tabla 1.** Hectáreas dedicadas a los principales cultivos de las provincias de la Comunidad Valenciana entre 1993 y 2005.

### Datos utilizados

El material utilizado en el presente trabajo para la clasificación y posterior actualización del SIG cítrico de la CV ha sido facilitado por el Instituto Cartográfico Valenciano (ICV). La información proporcionada es:

- Ortoimágenes aéreas en el visible (rojo, verde y azul, RGB) y en el infrarrojo cercano (CIR), georeferenciadas y ortorectificadas correspondientes a los recintos cítricos de la CV con una resolución de 0.5m.
- Archivo vectorial con la información catastral parcelaria a escala 1:10.000.
- Clasificación del SIG cítrico de las tres provincias de la CV del año 2001 y la clasificación del SIG cítrico de la provincia de Castellón del año 2006.
- Archivo vectorial de las estadísticas del tipo de cultivo por municipio a escala 1:10.000.

### METODOLOGÍA

En la primera fase del estudio se caracterizaron las tipologías de parcela más comunes según la variedad de los árboles contenidos en la parcela y según el cultivo mayoritario del municipio para conocer la variabilidad de las parcelas con las que se trabaja. En la siguiente etapa se realizó la extracción de características, obteniéndose 34 características inicia-

les a nivel parcelario, de las que posteriormente se realizó una selección para entrenar los clasificadores automáticos. Finalmente, en la clasificación final de las parcelas se utilizaron varios conjuntos de entrenamiento con tres métodos de clasificación. Mediante una combinación lineal de las salidas de los clasificadores se obtuvo la clase final de las parcelas y un grado de fiabilidad. A continuación se describen los distintos pasos utilizados para la clasificación.

### Caracterización de la tipología de las parcelas de cítricos

El estudio de las parcelas más comunes de la CV se realizó para caracterizar las tipologías de las parcelas. Se hizo una primera clasificación atendiendo al tamaño del árbol de la parcela dividiéndolos en árboles adultos y árboles jóvenes. En ambos casos, las parcelas más comunes se diferencian según la geometría de los árboles dentro de la parcela (lineal, regular, irregular y diagonal), según la masa vegetativa de los árboles (más o menos densos) y si el suelo de la parcela tiene o no cubierta vegetal. En cualquier caso, pueden existir parcelas que sean una mezcla de las parcelas más comunes o parcelas que contengan tipos de cítricos diferentes al común.

El ICV realizó una división de los municipios de la CV según el tipo de cultivo mayoritario del municipio, basándose en el documento Informe del Sector Agrario Valenciano del año 2006 (Conselleria d'Agricultura, Pesca i Alimentació), en árboles de secano, cítricos, herbáceos y frutales (Tabla 2).

Tipología	Alicante	Castellón	Valencia
Árboles Secano	36	42	75
Cítricos	18	14	94
Herbáceos	20	1	33
Frutales	8	–	6

**Tabla 2.** Número de municipios según la tipología del municipio de los 337 municipios procesados.

En el grupo de árboles de secano, no sólo se incluyen la vid y el olivo sino también el almendro, el algarrobo, el nogal y el avellano. El grupo de herbáceos está formado por las hortalizas, cereales, leguminosa, tubérculos, cultivo forrajero y viveros. Finalmente, en el grupo de frutales incluimos el resto de frutales.

La mayoría de municipios con cultivo mayoritariamente cítrico se suelen situar a baja altitud para prevenir el riesgo de heladas. Sin embargo, en zonas de mayor altura podemos encontrar cítricos en lugares más protegidos como por ejemplo entre valles de montañas o zonas con más humedad como la orilla de los ríos. Por ese motivo, casi todos los municipios con cultivo mayoritariamente cítrico se encuentran cerca de la costa mediterránea, salvo la Albufera y la Huerta de la provincia de Valencia, donde el cultivo mayoritario es herbáceo (Figura 1).

### Extracción de características

La extracción de características se desarrolló a nivel parcelario tratándose cada parcela como un objeto representado por las características extraídas. En los casos donde la parcela era heterogénea se trabajó a un nivel inferior que permitiese obtener una

identificación de las áreas parciales de cultivo, separando las áreas homogéneas en diferentes subgrupos dentro de esta parcela.

La localización de los árboles se obtuvo a partir de la característica del NDVI (Normalized Difference Vegetation Index). El NDVI (Tucker, 1979) es una medida de la diferencia normalizada entre las reflectancias del rojo y del infrarrojo cercano, mostrando valores altos cuando la vegetación es densa y sana, y valores bajos en zonas de vegetación pobre y dispersa, o en ausencia de ella. Antes de obtener los centros de los árboles, se filtró la característica del NDVI con un filtro de suavizado para eliminar el ruido del fondo y con un filtro de convolución circular para realzar la localización de los árboles. La detección de los árboles se realizó buscando máximos locales para que los centros detectados coincidiesen con los centros de las copas de los árboles y no estuviesen desplazados (figura 2).

Posteriormente a la detección de los árboles, se realiza una fusión de los centros localizados a muy corta distancia, en zonas homogéneas, en un solo árbol.

A partir de los árboles detectados se extrajeron 34 características de cada parcela catastral que podemos clasificar atendiendo a características estructurales,

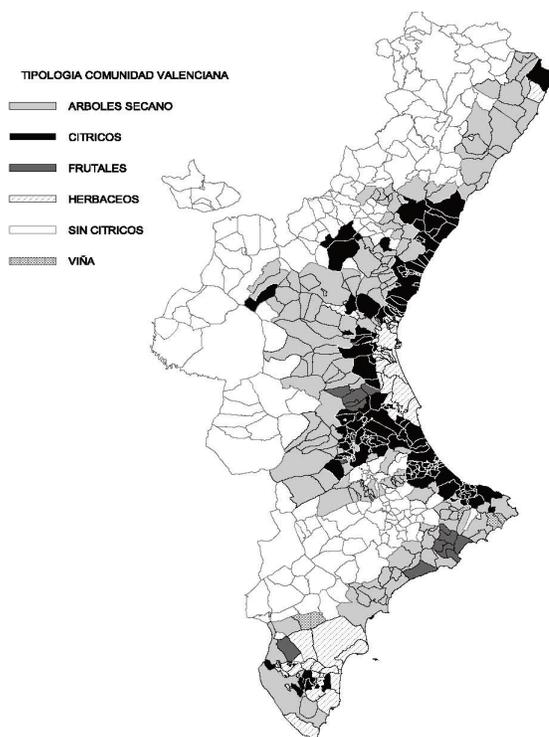


Figura 1. Comunidad Valenciana según la tipología de los municipios procesados.

espectrales y de textura.

*Estructurales (obtenidas a partir de las posiciones de los árboles detectados):*

- Media y desviación estándar de la distancia al vecino más próximo.
- Máximo y mínimo de los histogramas de vectores de unión entre los vecinos más próximos.
- Medida de distribución de los árboles o grid (regular o irregular).
- Distancia máxima y mínima de las direcciones principales (separación entre líneas).

*Espectrales (obtenidas a partir de los valores radiométricos de las 4 bandas espectrales):*

- Media y desviación estándar del NDVI de la copa, de los centros de la copa y del fondo.
- Media y desviación estándar del índice de verde de la copa.
- Media y desviación estándar del tono y saturación de la copa y de los centros de la copa.

*De textura (obtenidas a partir de la región que delimita cada grupo de árboles) (Gonzalez et al., 2002):*

- Energía: definida como la suma del cuadrado de los elementos de la matriz de coocurrencia.
- Correlación: definida como la medida de cómo se correlacionan los píxeles con sus vecinos sobre toda la imagen.
- Contraste: definida como la medida de la intensidad de contraste entre un píxel y su vecino sobre toda la imagen.
- Homogeneidad: definida como el valor que mide la proximidad de la distribución de elementos de la matriz de coocurrencia con la diagonal de la matriz.

### Selección de características

La selección de características se realizó con diversos métodos de selección, asignándole un voto a

cada característica seleccionada por cada método. Finalmente, las características más votadas fueron las elegidas para entrenar los clasificadores.

Los métodos utilizados son (Blum et al., 1998; Kohavi et al., 1997):

- Coeficiente de correlación de cada característica con la salida.
- Mínima redundancia máxima relevancia (Peng et al., 2005).
- Criterio discriminante de Fisher (Bishop, 2005).
- Selección mediante estadístico t (versión no sesgada de Fisher) (Theodoridis et al., 1999; Liu et al., 1998).
- Selección mediante la cota de clasificación de Chernoff (Theodoridis et al., 1999; Liu et al., 1998).

### Clasificadores

La gran cantidad de tipologías de parcelas que se pueden encontrar en la CV hace que exista una gran variabilidad en los datos de entrada a analizar, y a su vez establece la necesidad de utilizar más de un clasificador para asegurar un alto grado de confianza en la clasificación. Los métodos de clasificación utilizados se eligieron por su robustez y por su fácil interpretación, además de ser modelos con una alta fiabilidad. A continuación se describen los tres modelos de clasificación utilizados.

#### • Árboles de decisión

Los árboles de decisión (Breiman et al., 1984) son unos potentes y populares algoritmos de clasificación cuyo aprendizaje es fácil de implementar. Son utilizados para predecir y clasificar mediante procedimientos estadísticos, donde la relación entre el conjunto de entrenamiento y la respuesta del árbol viene dada por una regresión no paramétrica.

Un árbol de decisión considera como entrada un objeto o situación descrita por un conjunto de atribu-



**Figura 2.** Detección de los árboles en la imagen del NDVI. Imagen RGB del área seleccionada de una parcela (a). Imagen del NDVI con los árboles detectados marcados con un punto negro (b).

tos o características y devuelve una decisión ‘verdadero/falso’. El algoritmo denominado “*Classification and Regression Tree*” (CART) ideado por Breiman et al. (1984), consiste en una secuencia de nodos cada uno de los cuales se separa en dos ramas. Las divisiones se realizan de forma que los nodos hijos separan mejor las clases que los nodos padres sin sobreajustar el modelo.

Las ventajas del uso de CART frente a otros métodos son la elevada generalización, extracción sencilla de reglas e interpretabilidad del modelo.

- *Perceptrón multicapa (MultiLayer Perceptron, MLP)*

EL MLP es un modelo de red neuronal artificial perteneciente al grupo de las redes estáticas (Hush *et al.*, 1993; Haykin, 1999), se trata de sistemas sin memoria que no dependen de entradas o salidas pasadas o futuras. Los MLP están constituidos por neuronas interconectadas parcial o totalmente, con recurrencias o no, y que se encuentran organizadas en capas de tal forma que las señales que lo alimentan (señales de entrada) se propagan desde los nodos de entrada hacia los de salida.

La metodología del MLP se realiza mediante un mapeo complejo entre la entrada y la salida donde los parámetros son pesos sinápticos del modelo. El entrenamiento de la red se suele realizar empleando el conocido algoritmo ‘backpropagation’ o de retropropagación (Rumelhart *et al.*, 1986).

El MLP entrenado con dicho algoritmo es el más utilizado en multitud de campos, tanto en clasificación de patrones como en predicción de series temporales.

- *Máquinas de Vectores Soporte (Support Vector Machines, SVM)*.

Las máquinas de Vectores Soporte (Vapnik, 1998) son herramientas avanzadas para el aprendizaje en máquinas. Su inspiración es bastante diferente a las redes neuronales artificiales. Proporcionan cuatro características básicas:

- Elevadas capacidades de generalización: La metodología de clasificación en la SVM, se realiza trazando hiperplanos de máximo margen para separar las distintas clases en las que se dividen los datos de entrada (Camps-Valls *et al.*, 2005).
- ‘Slack variables’: Cuando los datos no son linealmente separables, las SVM relajan las restricciones asociadas a cada muestra (cada muestra tiene como restricción impuesta que pertenezca a una u otra clase) introduciendo términos positivos de relajación (slack variables) para cada muestra.

- Espacio de características: Las SVM pueden construir funciones de decisión no lineales transformando los datos de entrada en un espacio de alta dimensionalidad llamado espacio de características donde los datos si son separables linealmente.
- Vectores soporte: La frontera de decisión se calcula como una combinación lineal de pocas muestras del espacio de entrada llamadas vectores soporte.

Si las distintas clases no pueden ser separadas de forma lineal, la SVM utiliza el método de kernel (Gualtieri et al., 2000 y Schölkopf *et al.*, 2002).

## Clasificación

El entrenamiento de los distintos métodos de clasificación requiere un conjunto de entrenamiento (clasificación supervisada) que se obtuvo mediante el etiquetado de parcelas por fotointerpretores expertos. Este conjunto de entrenamiento fue realizado por el ICV y está compuesto por 109 municipios con un total de 72.950 parcelas etiquetadas de la CV. Se seleccionaron parcelas de diversos municipios cubriendo todas las tipologías de cultivo (Tabla 2): 41 municipios de árboles secano, 43 de cítricos, 10 de frutales y 15 de herbáceos.

Con todas estas muestras etiquetadas se entrenaron tres clasificadores globales (TREE, SVM y MLP) para cada provincia y tres más para cada tipología de cultivo (TREE, SVM y MLP) de cada provincia. Independientemente de la provincia donde se encuentre el municipio a clasificar, cada parcela estará clasificada por 6 clasificadores (los tres globales y los tres según la tipología que tenga el municipio). Se realizó una combinación lineal (Hashem, 1997) entre el resultado de los distintos clasificadores y la clasificación del SIG citrícola anterior. Con la combinación lineal se obtuvo la clase final de las parcelas y un grado de fiabilidad definido por la discrepancia entre los distintos clasificadores y la clase del SIG anterior. Las parcelas con bajo grado de fiabilidad son etiquetadas como “Revisar” para la posterior fotointerpretación por parte de ICV. Esto permite aumentar la fiabilidad en la actualización del SIG citrícola de la comunidad al fotointerpretarse las parcelas con mayor probabilidad de error en la clasificación automática.

## RESULTADOS

La actualización citrícola de la CV se realizó en

dos etapas, por lo que el conjunto de entrenamiento utilizado es distinto para cada una de las clasificaciones. En la primera etapa se clasificó la provincia de Valencia (Septiembre de 2007), utilizándose solo las muestras etiquetadas de esta provincia como conjunto de entrenamiento (53.796 parcelas). En la siguiente fase se clasificaron las provincias de Alicante y Castellón (Marzo de 2008) utilizando, en Alicante las muestras etiquetadas de la provincia de Alicante (11.577 parcelas), además de las que se tenían de la provincia de Valencia y para la clasificación de la provincia de Castellón las muestras fotointerpretadas de Castellón (7.577 parcelas), junto con las de Valencia.

Se ha actualizado el SIG citrícola del año 2001 de 322 municipios de los 542 municipios que tiene la CV y además se ha generado el SIG citrícola en 15 municipios nuevos. Por provincias, se han clasificado 82 municipios en Alicante, 57 en Castellón y 208 en Valencia, con un total de 1.636.126 parcelas. Considerando los recursos humanos disponibles, el tiempo y el alto grado de fiabilidad exigido, se han clasificado automáticamente el 87% de las parcelas de la CV y se han etiquetado para revisar posteriormente por fotointerpretación el 13%.

Dentro de las parcelas a revisar, además de las que

tienen el grado de fiabilidad más bajo, también se incluyen las parcelas que presentan cultivos con árboles excesivamente pequeños definidos como plantones. Muchas de estas parcelas presentan dificultades para ser clasificadas mediante fotointerpretación y deberán comprobarse mediante visitas al campo.

El número de parcelas etiquetadas a revisar está relacionado con el grado de fiabilidad de la clasificación de las parcelas, pudiéndose aumentar o disminuir en función de los medios y recursos disponibles para fotointerpretar estas parcelas. En la figura 3 vemos que si se aumenta la probabilidad fijada para clasificar la provincia de Valencia se obtendrán más parcelas a revisar en esa provincia. Esta probabilidad se puede fijar para todas las parcelas de la provincia o por tipologías. En el presente trabajo se han clasificado las tres provincias con sus respectivos modelos de clasificación y con el mismo umbral en la probabilidad de acierto, fijado en el 35,30%.

Los resultados obtenidos se detallan atendiendo a cada provincia de la CV ya que los clasificadores se entrenaron por provincias. En la tabla 3 puede verse que se ha clasificado automáticamente el 80% en Alicante, y el 87% en Castellón y Valencia.

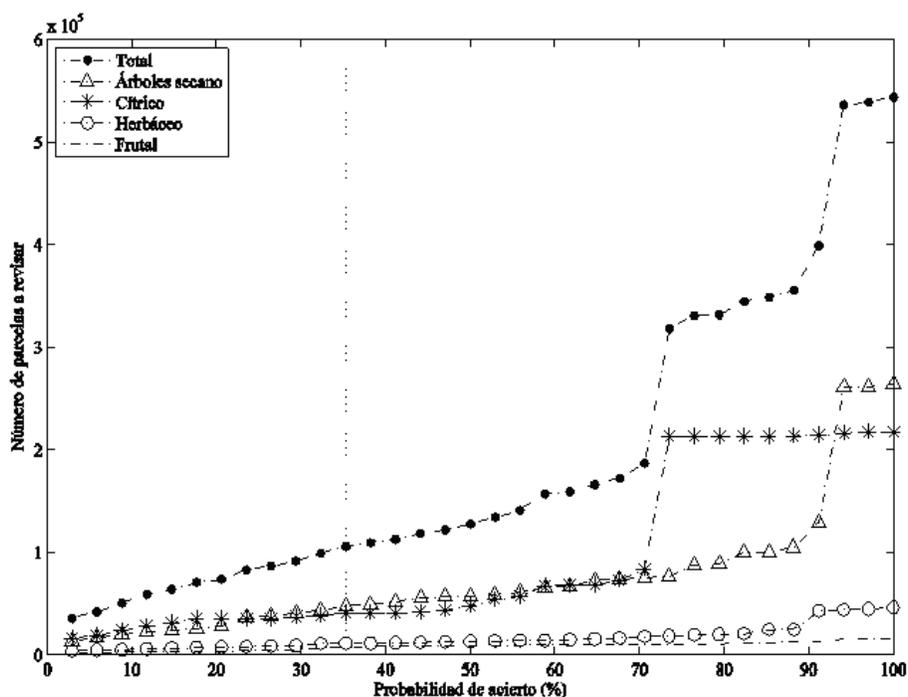


Figura 3. Relación entre el número de parcelas etiquetadas a revisar y la probabilidad fijada para clasificar la provincia de Valencia.

Tipología	ALICANTE (%)				CASTELLÓN (%)				VALENCIA (%)			
	cit	nocit	plantón	Rev	cit	nocit	plantón	Rev	cit	nocit	plantón	Rev
Árbol Secano	11,18	75,08	2,65	11,10	5,59	82,8	4,32	7,28	11,4	76,42	1,70	10,48
Citrico	29,8	53,08	1,58	15,54	34,49	50,39	2,60	12,53	47,06	39,92	1,76	11,26
Herbáceo	14,02	55,76	6,54	23,69	13,77	67,14	4,80	14,29	24,36	64,14	1,43	10,07
Frutal	16,32	51,75	1,32	30,61	--	--	--	--	22,09	51,93	3,35	22,63
<b>TOTAL</b>	<b>15,87</b>	<b>64,02</b>	<b>3,12</b>	<b>16,99</b>	<b>18,11</b>	<b>68,62</b>	<b>3,74</b>	<b>9,53</b>	<b>26,86</b>	<b>60,21</b>	<b>1,76</b>	<b>11,17</b>

**Tabla 3.** Estadísticas de la clasificación en % para las diferentes tipologías en las provincias de la CV (parcelas cítricas (cit), no cítricas (nocit), con plantones (plantón) y parcelas etiquetadas a revisar (Rev))

La provincia de Valencia es la que más parcelas dedica al cultivo de cítricos y por tanto es la que mayor porcentaje de parcelas clasificadas como cítrico tiene en todas sus tipologías. Observamos también que la tipología frutal es la que contiene más parcelas etiquetadas como revisar, siendo el 32% en la provincia de Alicante y el 26% en Valencia.

En la tabla 4 se muestran los resultados obtenidos en la clasificación automática de la Comunidad Valenciana (acierto e índice estadístico Kappa propuesto por Cohen (Congalton *et al.*, 1999)) usando la misma probabilidad para las tres provincias respecto de las muestras fotointerpretadas según la tipología (no se incluyen las parcelas a revisar).

Los resultados muestran una gran fiabilidad en la clasificación obtenida, siendo la Kappa total superior al 0.92 en las tres provincias. En la tabla 4 se observa que los mejores resultados se han obtenido para la provincia de Castellón. Esto es debido a que esta provincia tiene una tipología parecida a la de la provincia de Valencia y su modelo de entrenamiento se ha obtenido con la combinación de muestras eti-

quetadas de ambas provincias. Además, la combinación lineal de los clasificadores utiliza la clasificación cítrica anterior y en Castellón es donde se tenía más actualizada (2006). En cambio, el acierto en Valencia es menor ya que se generaron los modelos de clasificación con menos muestras (solo con las etiquetadas en Valencia) y la clasificación cítrica anterior utilizada era más antigua (2001).

La clasificación más baja se obtiene en la provincia de Alicante. Este resultado se podría justificar debido a que, aunque se generaron los modelos de clasificación con las muestras de Alicante y Valencia, las especies cítricas de Alicante son bastante diferentes a las de Valencia. En Valencia se cultiva principalmente la especie de naranjo dulce (*Citrus sinensis*), mientras que en Alicante predominan plantaciones de limoneros (*Citrus limonum risso*) en sur de la provincia (comarcas de La Vega Baja y Baix Vinalopó). Estas comarcas contienen municipios mayoritariamente herbáceos, lo cual disminuye la fiabilidad en la clasificación de esta tipología en Alicante (Kappa = 0.87).

Tipología	Alicante		Castellón		Valencia	
	Acierto	Kappa	Acierto	Kappa	Acierto	Kappa
Secano	99,15%	0,97	98,35%	0,97	97,90%	0,94
Citricos	97,15%	0,95	98,69%	0,97	97,05%	0,93
Herbáceo	93,33%	0,87	--	--	96,81%	0,93
Frutal	93,04%	0,86	--	--	95,08%	0,89
<b>Total</b>	<b>96,54%</b>	<b>0,92</b>	<b>98,57%</b>	<b>0,97</b>	<b>97,17%</b>	<b>0,94</b>

**Tabla 4.** Resultado de comparar las parcelas fotointerpretadas respecto las clasificadas en las tres provincias según la tipología del municipio de la Comunidad Valenciana.

En general, la tipología frutal es la que menor fiabilidad presenta, alcanzando consecuentemente, la mayor proporción de parcelas a revisar (tabla 3). Esto es debido principalmente a ser la tipología más parecida a la cítrica en cuanto al tamaño y forma de los árboles, vigorosidad de la vegetación (color y densidad) y estructura de la parcela. Además, el número de municipios de esta tipología es el más bajo de la comunidad (Figura 1) y, por tanto, el conjunto de entrenamiento para los clasificadores es bastante reducido.

En la tabla 5 se observa el acierto y la Kappa de los clasificadores de las tres provincias de las parcelas clasificadas como 'Revisar' que han sido fotointerpretadas por el ICV, es decir, parcelas definidas por los clasificadores con poco grado de fiabilidad.

Alicante		Castellón		Valencia	
Acierto	Kappa	Acierto	Kappa	Acierto	Kappa
47,37	-0,06	48,72	0,04	63,37	0,26

**Tabla 5.** Resultados de la validación de las parcelas etiquetadas a revisar.

Se observa al aplicar los clasificadores a estas parcelas, que si se hubiera clasificado automáticamente se habría obtenido un acierto bastante bajo en las tres provincias. Este resultado pone de manifiesto la necesidad de realizar la actualización del SIG cítrico de forma conjunta mediante técnicas de automatización y de fotointerpretación.

## CONCLUSIONES

En este trabajo se ha realizado la actualización de los recintos cítricos de la Comunidad Valenciana clasificando automáticamente el 87% de las parcelas procesadas con un alto grado de confianza. Es decir, en la segunda etapa se ha reducido el trabajo de fotointerpretación al 13% y por consiguiente la necesidad de visitas al campo.

La metodología propuesta permite clasificar los recintos parcelarios en cítricos o no cítricos con gran precisión, como demuestra el acierto obtenido del 96.54% en Alicante, 98.57% en Castellón y 97.17% en Valencia respecto de las parcelas fotointerpretadas.

La aproximación seguida en el presente estudio para la provincia de Valencia se determinó en función de las fechas de entrega y la disponibilidad de los datos, pudiéndose mejorar la clasificación generando los modelos de clasificación usando las mues-

tras fotointerpretadas de toda la Comunidad Valenciana, ya que esta provincia colinda con las otras dos, y por consiguiente, tiene características muy similares al norte de Alicante y al sur de Castellón.

Las dificultades encontradas en la diferenciación entre árboles cítricos y frutales, podrían no presentarse si se dispusiera de las imágenes de invierno ya que los cítricos son árboles de hoja perenne, mientras que los frutales de la CV son de hoja caduca.

Los resultados muestran que el acierto obtenido en la clasificación final esta totalmente condicionado por la calidad de las características extraídas. De hecho, esta sería la primera etapa a mejorar ya que influye directamente en la separación de las clases a determinar. El primer paso sería mejorar el algoritmo de localización de la posición de los árboles del cual dependen la mayoría de características, que se debería hacer más robusto frente a la cubierta vegetal de fondo, incluir la posibilidad de reposicionar los centros para mejorar las estructuras regulares, detectar faltas en estas estructuras y compensarlas, etc. Finalmente, la utilización de métodos automáticos junto a técnicas tradicionales de análisis de imágenes mediante fotointerpretación proporciona unos buenos resultados a la vez que reduce el tiempo y el coste de actualizar del SIG cítrico de forma tradicional.

## AGRADECIMIENTOS

Este proyecto ha sido subvencionado por el Ministerio de Educación y Ciencia bajo el proyecto DATASAT (ESP2005-07724-C05-03) y diversos contratos entre el Instituto Cartográfico Valenciano y la Universidad de Valencia en 2007 y 2008.

Agradecemos la colaboración del Instituto Cartográfico Valenciano por el cual se ha tenido acceso a los datos utilizados para la realización del presente trabajo.

## BIBLIOGRAFIA

- BISHOP, C. M., 2005. Pattern Recognition and Machine Learning. Clarendon Press—Oxford, 2005.
- BLUM, A., LANGLEY, P., 1998. Selection of Relevant Features and Examples in Machine Learning. *Artificial Intelligence* 97, 1997, pp. 245-271.
- BREIMAN, L., FRIEDMAN, J., OLSHEN, R., STONE, C., 1984. Classification and Regres-

- sion Trees. *Wadsworth and Brooks*, 1984.
- CAMPS-VALLS, G., BRUZZONE, L., 2005. Kernel-based Methods for Hyperspectral Image Classification. *IEEE Transactions on Geoscience and Remote Sensing*, 2005, Vol. 43, Issue 6, pp. 1351- 1362.
- CONGALTON, R. G., GREEN, K., 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. *CRC Press*, 1999.
- CONSELLERIA D'AGRICULTURA, PESCA I ALIMENTACIÓ, GENERALITAT VALENCIANA, 2006. Informe del Sector Agrario Valenciano. Anexo 2, 23-34.
- DEMERS, M. N., 1999. Fundamentals of Geographic Information Systems. *John Wiley & Sons*, 1999. 498 pages.
- GONZALEZ, R.C., WOODS, R.E., 2002. Digital Image Processing (2nd Edition). *Prentice Hall*, 2002.
- GUALTIERI, J. A., CHETTRI, S., 2000. Support vector machines for classification of hyperspectral data. *IEEE 2000 International*, 2000, Vol.2, pp. 813-815.
- HASHEM, S., 1997. Optimal Linear Combinations of Neural Networks. *Neural Networks* 10, 1997, pp. 599-614.
- HAYKIN, S., 1999. Neural Networks: A Comprehensive Foundation. *Prentice Hall*, 1999.
- HUSH, D.R., HORNE, B.G., 1993. Progress in supervised neural networks. *Signal Processing Magazine, IEEE*, 1993, Vol.10, pp. 8-39.
- KOHAVI, R., JOHN, G.H., 1997. Wrappers for features subset selection. *Int. J. Libr. 1*, 1997, pp. 108-121.
- LIU, H., MOTODA, H., 1998. Feature Selection for Knowledge Discovery and Data Mining. *Kluwer Academic Publishers*, 1998.
- PEEDEL, S., KAY S., GIARDINO G., 1998. Seeing the (olive) trees from the wood - using GIS in Europe for olive tree identification. *ESRI User Conference*, San Diego, 27th-31st July 1998.
- PENG, H., LONG F., DING C., 2005. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *Pattern Analysis and Machine Intelligence*, August 2005 Vol. 27, No. 8 pp. 1226-1238.
- RUMELHART, D.E., McCLELLAND, J.L., 1986. Parallel Distributed Processing: Explorations in the microstructure of cognition. *MA :MIT Press*, 1986, Vol., 1.
- SCHÖLKOPF, B., SMOLA, A., 2002. Learning with Kernels – Support Vector Machines, Regularization, Optimization and Beyond. *MIT Press Series*, 2002.
- THEODORIDIS, S., KOUTROUMBAS, K., 1999. Pattern Recognition, *Academic Press*, 1999, pp. 341-342.
- TUCKER, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote, Sens. Environ.*, 1979, Vol. 8, pp. 127-150.
- VAPNIK, V.N., 1998. Statistical Learning Theory. *John Wiley & Sons*, 1998.