FETEX: A feature extraction software tool for object-based image analysis

Ruiz Fernández, Luis Ángel (laruiz@cgf.upv.es)
Recio Recio, Jorge Abel (jrecio@cgf.upv.es)

GeoEnvironment Cartography and Remote Sensing Group (CGAT)

Department of Cartographic Engineering, Geodesy and Photogrammetry

UNIVERSIDAD POLITECNICA DE VALENCIA

1 Introduction

Image classification techniques are used in the field of remote sensing in a large number of applications, one of which is the creation of databases of occupation and land use. Traditionally, these techniques have been based on the analysis of independent pixels, using multiple spectral bands and, occasionally, additional information related to the spatial distribution of intensity values around each pixel, such as textures. However, nowadays methods for the classification of objects are developed, with a semantic and practical level superior to that of the pixel.

Additionally, much attention has been paid to the development of new methods for the creation of decision functions, in order to improve the results of the classification, but not so much to the extraction of valid and exhaustive information of the image. However, a crucial step prior to classification is the calculation of the attributes that describe the elements of the image to be classified, which is known as the feature extraction process.
FETEX is a software for the automatic extraction of descriptive characteristics of the texture, spectral, shape and structural properties of image fragments or image-objects.

It is oriented to the characterization of cartographic objects (plots, subplots, polygons, etc.), defined from vector files (shapefiles), as well as high-resolution satellite images (QuickBird, Ikonos, WorldView, etc.) and photographs CIR aerial. Thus, for example, with a multispectral image and the plot defined by the cadastral cartography as input data, an ASCII table (output) is generated with the values of the characteristics or attributes calculated automatically from the images for each of the cadastral parcels. This result can then be entered into classification software.

Many of the characteristics obtained are novel and are based on advanced calculation methods, especially the structural characteristics, which allow to describe numerically the spatial distribution of the elements present in an object / plot.

2 Objectives

- To understand the operation of Fetex software for the extraction of descriptive attributes of cartographic objects

- To apply Fetex to a simple example using cadastral plots and a CIR aerial image and to obtain and interpret numerical results and graphs.

3 General description of the software tool

The user interface is divided into three sections (see Figure 1):

- *Input files*: This section facilitates the loading of input data in the application: (i) high resolution multispectral images and (ii) vector files in shapefile format that contain the limits of the objects or polygons to be characterized.

- *Feature extraction and parameters*: Control panel that allows selecting the characteristics that will be extracted, as well as certain parameters for its calculation (Figure 2).

- *Output files*: It allows the definition of the output files that contain the descriptive characteristics for each object in several different formats. In addition, several types

of files can be generated with graphic information referring to some of the calculated characteristics (Figure 3).
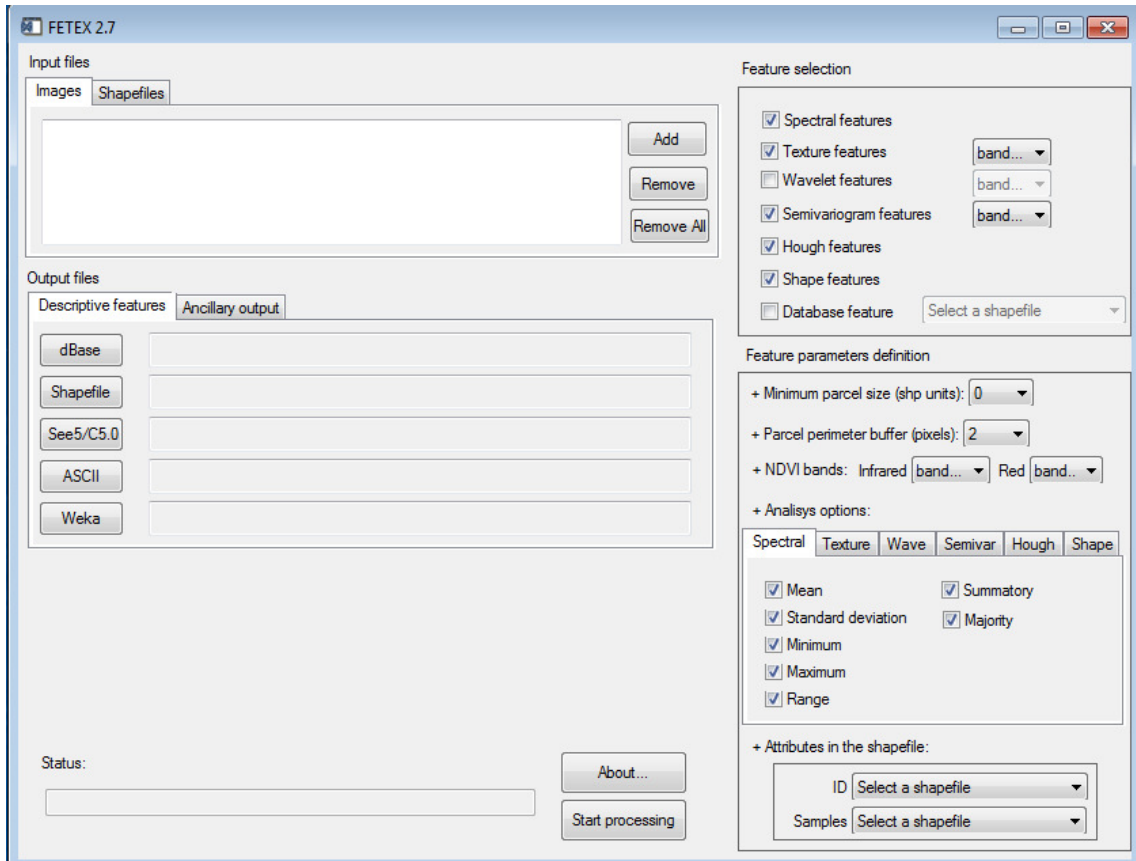

Figure 1.- Fetex 2.7 Interface.

Input files are loaded through the "Images" and "Shapefiles" tabs:

- *Images*: Dialog box for the selection and loading of the raster files in any geo-referenced format accepted by ENVI: * .tif, binary, ERDAS * .img, * .ecw, jpeg2000, etc.

- *Shapefiles*: Dialog box to select and load vector files in ESRI shapefile format (* .shp) that contain the boundaries of the parcels to be processed. In practical applications, it can be a layer of the Cadastre database, SIGPAC, SIOSE, etc.

Both allow the loading of one or several files. In case a polygon or plot is divided among several images, the program will automatically make the mosaic.

4 Feature extraction

The extraction of characteristics or descriptive attributes is one of the fundamental parts of an object-oriented classification process. In the Fetex program, it is done through the options on the right side of the user interface (Figure 1). The descriptive characteristics must be chosen and some parameters adjusted for its calculation. First, you must define the types of attributes that you want to calculate and, in the

case of texture variables, wavelets and derivatives of the semivariogram, also the band from which they are calculated. The following parameters are defined below:

- Minimum size of parcels (in pixels): those parcels with a number of pixels lower than that indicated will not be processed.

- Buffer of the perimeter of the plot (in pixels): During the processing, an outline with the thickness of the indicated pixels is excluded to avoid geometric mismatches due to georeferencing differences between the image and the shapefile files.

- NDVI Bands: It is necessary to indicate which bands correspond to the NIR and red to be able to calculate the normalized difference vegetation index (NDVI).

The groups of features that can be calculated by the tool are (Figure 2):

a) _Spectral_: Spectral statistics features that inform about the radiometric response of a plot.

b) _Texture_: texture features that describe the spatial distribution of the gray levels corresponding to the pixels within a plot or polygon. The roasts can be chosen in the intensity of edges, the shape of the histogram and the derivatives of the matrix of concurrency levels of gray (GLCM) (Ruiz et al., 2011). It is necessary to define the number of gray levels or intensity to be considered (normally 16, 32 or 64 is sufficient).

c) _Wavelets_: Texture features calculated from the details of the wavelet decomposition of the image, where the GLCM matrix is also used. The wavelet family and the number of gray levels must be defined for the calculation of the GLCM.

d) _Semivariogram_: Structural features calculated from the semivariogram, which describes the regularity of the internal elements of a plot. A set of variables described in Balaguer et al. (2010) and specified in Ruiz et al. (2011). As parameters, you must specify the maximum size of the lag (maximum distance between pixels) and the percentage used for its calculation (it is recommended to use the default value of 15%). This parameter allows to reduce substantially the calculation time without significantly influencing the result (Balaguer et al., 2011).

e) _Hough_: These are structural features calculated from the Hough transform and the positions of the trees in a plot for the characterization of plantation frames or other structures with geometric patterns, as described in Ruiz et al. (2011). It is necessary to define several parameters for the calculation of local maxima, such as NDVI and infrared thresholds and the minimum and maximum diameters, which depend on the radiometry of the images and the type of landscape elements (distances between trees, etc.) , so they should normally be adjusted by preliminary tests on a small area representative of the study area.

f) _Shape_: Shape features of the object based on relationships between the perimeter and the area. Describe the geometric properties of objects, complexity or elongation of their contour, etc.

g) _Database features_: auxiliary features directly taken from the information of the database analyzed, such as use or previous coverage of that plot (Recio et al., 2011).

Figure 2.- Interface for defining the characteristics and their parameters; a) spectral, b) texture, c) wavelets, d) semivariogram, e) Hough, and f) shape.

5 Output files

In section *Output files* there are two tabs:

- *Descriptive features*: The output format of the file that will contain the generated features is defined. This output file contains a table where each of the plots or study polygons is represented in the rows, with their corresponding identifier, and the columns contain the values of each of the calculated features. The possible output formats are: dBase, Shapefile, See5/C5.0, ASCII and Weka. See5 is a program of classification by decision trees, while Weka (http://www.cs.waikato.ac.nz/ml/weka/) is a free data mining program that presents numerous classification algorithms. Both can be used later to classify the plots from the descriptive characteristics generated by *Fetex*.

- *Ancillary output*: Optionally, graphical outputs of the features can be generated by plot, both in black and white and in color. Figure 3 shows an example of these graphic outputs. Additionally, a file in DBF format containing the semivariograms of all the plots can be obtained. All this contributes to an exhaustive interpretation of the results.

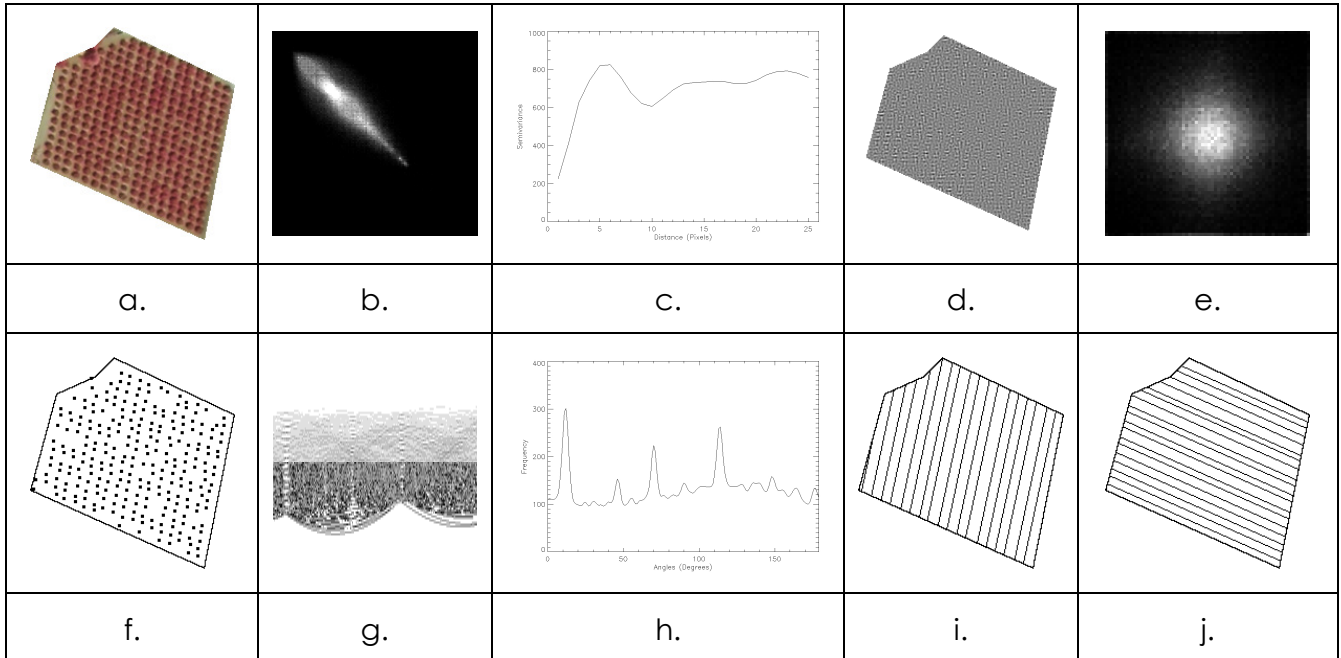|				
a.	b.	c.	d.	e.
f.	g.	h.	i.	j.

Figura 3. Algunas salidas gráficas generadas por FETEX para un objeto individual: imagen de falso color (a) e imagen de los detalles wavelets (d), representaciones de la GLCM de la imagen original (b) y de los detalles wavelets (e); curva del semivariograma del objeto-imagen (c); imagen binaria de la localización de los árboles (f); transformada de Hough (g) e histograma de direcciones (h); imágenes binarias de las alineaciones de los árboles detectados: (i) y (j).

7 Application example

As a simple example of application we will use a small fragment of an image with three spectral bands (NIR, Red and Green), a shapefile file with the limits of 8 cadastral plots and the trial version of the Fetex program, available at (http: // cgat.webs.upv.es/software/). This will allow us to interpret the results obtained.

Start and data loading:

- ccess the website http://cgat.webs.upv.es/software/ and download the compressed file "fetex2.rar" and the sample data "sample dataset" in the directory "mydirectory" (eg, "c : \ test_fetex").

- Start the program ENVI + IDL and, in the IDL command line type: cd, "mydirectory" and press "Enter". Then write fetex2.

- When the "Images" tab is activated, press "Add", search for the data and load the "Wave.tif" image (it contains three spectral bands: 1-NIR, 2-Red, 3-Green). Activate

the flap "Shapefiles" press "Add" and load the file with the limits of the test plots "parcelas.shp".

Definition of features and parameters:

- In the "Feature selection" dialog box, select all feature groups except, for example, "Wavelet features" and "Database feature". Also select the band for the calculation of the texture characteristics and the semivariogram (eg, 1, which corresponds to the NIR band).

- In the "Feature parameters definition" dialog box, select the minimum plot size (eg, 10 pixels), the buffer to be excluded from the perimeter of the plot (eg, 2 pixels), and the bands corresponding to Infrared and Red (in this image file the bands 1 and 2, respectively).

- Select all spectral features (Mean, Std. Dev., Etc.); the 3 groups of texture variables, taking 32 gray levels for the calculation of the GLCM; a maximum lag size in the semivariogram of 25 pixels (this value expresses the maximum distance between pixels to be considered in the calculation of the semivariogram of each plot), and a random percentage of pixels and distances for the calculation of each plot of 15%; we leave the parameters of the features derived from the Hough transform by default; and we select the 5 shape variables.

Output files:

- In the section "Output files" and with the tab "Descriptive features" active select the option "dBase", the path or directory and write the name of the output characteristics file (eg, "c:\test_fetex\output.dbf ").

- Activate the "Ancillary output" tab and select the "Color" options (R1, G2, B3) and "DBF containing the semivariogram of all parcels".

- In the same tab, press "Screenshots of objects/parcels" and select the directory in which all auxiliary outputs will be stored (eg, "c:\test_fetex").

Press "Start processing" at the bottom of the program interface. When you finish processing the 8 parcels you will see a table with information about global processing times and per plot.

Access the directory with the output data, examine and interpret the following results:

1. The auxiliary data found in each of the 8 folders identified with the ID of each plot.

2. Open the file "salida.dbf" in Excel that contains all the values of the generated characteristics (columns) for each plot (row).

3. Open in Excel the file "parcelas_semivariogramas", which contains the values of the semivariogram (semivariance) obtained for each plot, from the first value (D1) to the maximum value of the lag that we have set (D25). Selecting all the values you can generate a lines graph with the semivariograms of the 8 plots, similar to that of Figure 4. Identify each semivariogram obtained and interpret it comparing it with the object-image of each plot.
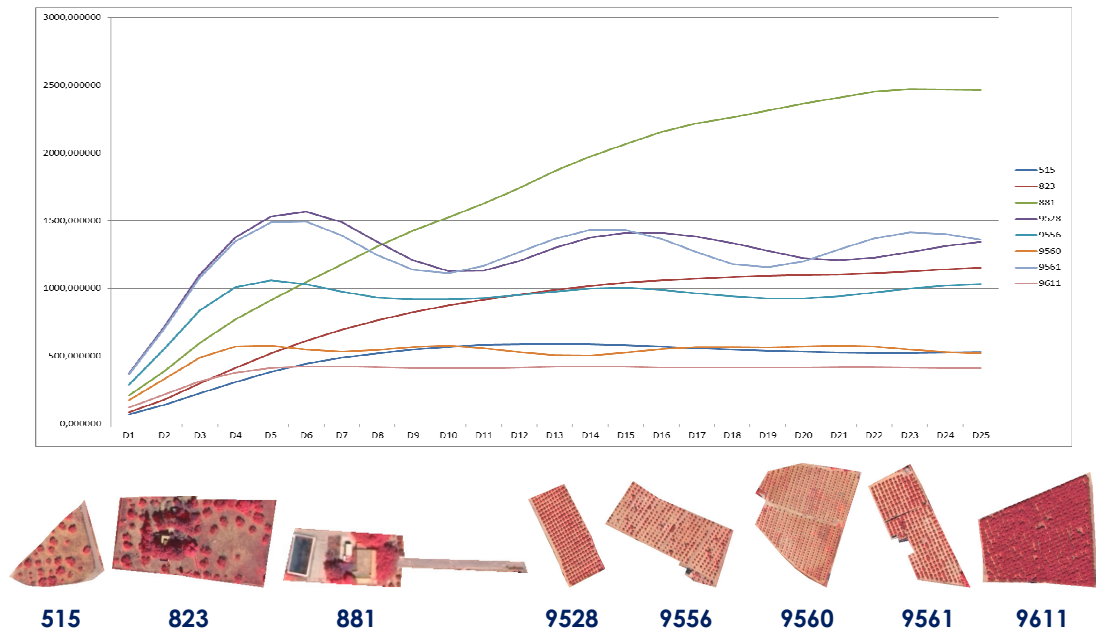
Figure 4.- Above: Semivariograms generated from the intensity values of the 1-NIR band of the 8 different plots used in the application example. Bottom: Object-images of the plots with their corresponding identifier.

8 Conclusions

Land use/land cover geodatabase updating often requires methods of object-based classification, one of whose crucial phases is the extraction of a set of features of each object that allows to adequately solve each classification problem. The *Fetex* program allows us to extract an exhaustive set of features per object-image by just entering the original image and a shapefile with the boundaries of the parcels or objects that we want to classify. In addition to the features or descriptive attributes, the graphic outputs obtained allow us for a more complete interpretation of results.

9 References

- Balaguer, A., Ruiz, L.A., Hermosilla, T., Recio, J.A., 2010. Definition of a comprehensive set of texture semivariogram features and their evaluation for object-oriented image classification. Computers and Geosciences, 36(2), pp. 231-240.

- Balaguer, A., Hermosilla, T., Recio, J.A., Ruiz, L.A., 2011. Semivariogram calculation optimization for object-oriented image classification. Modelling in Science Education and Learning, 4(7), pp. 91-104.

- Recio, J.A., Hermosilla, T., Ruiz, L.A., Fernández-Sarría, A., 2011. Historical land use as a feature for image classification. Photogrammetric Engineering and Remote Sensing, 77(4), pp. 377-387.

- Ruiz, L.A., Recio, J.A., Fernández-Sarría, A., Hermosilla, T., 2011. A feature extraction software tool for agricultural object-based image analysis. Computers and Electronics in Agriculture, 76(2), pp. 284-296.