Universitat Autònoma de Barcelona

Model-Based Segmentation of Images

Margarita Torre Alcoceba

UNIVERSITAT AUTÒNOMA DE BARCELONA

DOCTORAL THESIS

Model-Based Segmentation of Images

Author:
Margarita Torre Alcoceba

Supervisor:
Petia Ivanova Radeva
Fernando Martínez Sáez

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy in Computer Science

at the

Universitat Autònoma de Barcelona

April 26, 2020

Abstract

Photography freezes in an instant the data that can later be extracted, interpreted and transformed over time to communicate information in different formats. Making maps from photographs was a revolution in cartography. Advances in Computer Vision are helping to bring about the next revolution in this discipline, which aims at more and more detailed geographic information which is required in shorter periods of time. In this way, the process that goes from image to a map has become increasingly automatic. The images already captured with high-resolution digital cameras are automatically placed in the correct position of the terrain as if they were a sheet that covers it, thanks to the digital terrain models, thus obtaining orthophotomaps. In these circumstances, the only burden that remains to be lightened is the extraction of the topographic elements, without losing the precision and quality of interpretation that until now has been provided by human operators.

This research focuses on the development of new computerized methods that facilitate these tasks of extracting information from aerial images. We start with the development of a strategy to semi-automatically extract fields from the images. This approach uses the almost homogeneous response of the fields and how this response differs from that obtained from their neighbors. The process is carried out by means of the method in which adjacent regions compete to own a pixel. When the contrast lines of the images are also taken into account, it is possible to extend the previous methodology to extract roads. In both cases it is necessary to guide the entire process, not only by the points given by an operator, but by the model of the element to be extracted. The model helps to refine the results obtained. When Deep Learning burst onto the Computer Vision scene, all the processes of image classification were upended. So, we propose a joint venture between a deep network and an energy-minimization model-guided radiometric method that improves the benefits of each component. This approach reduces to a minimum the need for human interaction and obtains reliable results.

Acknowledgements

"This thesis has always been my destination, but the road to reach it has been long and winding."

As soon as I started to develop applications to generate maps from aerial images, I knew that the goal of my work was to include the knowledge of experts in tools that would facilitate the generation of precise geographic information. Now, I want to express my gratitude to all the people who I had the privilege of working with for so long in the Institut Cartogràfic de Catalunya. In particular, these people who were not responsible for training, but who passed on their knowledge and placed their trust in me as well: Xavier Alvaro, Mari Luz Ibarra and Toni Ramos in the field of photogrammetry, Isabel Ticó and Jordi Clua in cartography.

On my journey to this thesis, everything seemed easier because I saw that my co-workers had a passion for the same subjects that were guiding me. It was a pleasure to travel alongside you, Pere, Toni, Joan y Albert. You are the perfect team that most people can only dream of having. Miriam Amo and Margarita del Oso also, in their respective projects, shared our ups and downs. I thank you all.

The road has been long. It has taken twenty years to bring this cycle to a close, and I could make a lot of excuses. To tell the truth, I have always called this thesis "the broccoli thesis". This is because my moments to think and develop approaches to the thesis were mixed in with other tasks: while boiling the broccoli to make dinner, for example. But nothing happens just by chance. The idea was always in the back of my mind. So the inconvenience of not having the thesis as a first or second concern also has its advantages: you optimize all your steps and are very attentive to the opportunities. Twenty years have given me a chance to witness the birth of a discipline. Then there was a paradigm shift that, had I not had a specific need in mind, I might not have taken advantage of to the same extent. And in those moments when this long journey brought me to my knees, I have had people who asked me to get back up again. Who told me that it was worth it. Thank you, Petia.

As I have always found it difficult to compartmentalize my passions, I also owe my lifelong friendships to the Institut Cartogràfic. They supported me when any adversity plunged me into a sea of tears, and knew when to tell me that tears wash away all our worries. My eternal gratitude to Monica, Inma, Mariàngels, and especially to Carme, the sister I never had.

The fact that this thesis is written in comprehensible English is something I owe to the efforts of my friend Stuart. Although he is a linguist, it is possible that after this project, his knowledge of photogrammetry will be greater than my knowledge of English. Likewise, entering the field of neural networks, which was so new to me, would have been much more difficult if I had not had recourse to the support and knowledge of Bea Remeseiro, with whom I was also able to publish the article that culminates the research.

Traditionally, this would be the moment to thank the research programs which tend to facilitate access to media such as computers, software for developments, journals, and so on. In my case, all this as well as unconditional support through thick and thin has been provided to me by Fernando. Without him, this thesis would not exist.

If that had been the case, it wouldn't have been the end of the world. Other researchers would have come to similar conclusions. It is precisely for this reason that I am so grateful to you all. Thanks to you, I have had the pleasure of being the one to follow this long and winding road to the end.

Agradecimientos

"La tesis estaba clara, no así el camino para conseguirla."

Desde el primer momento en que tuve la oportunidad de desarrollar aplicaciones para generar mapas a partir de imágenes aéreas, tuve claro que el objetivo de mi trabajo era ayudar a incorporar el conocimiento de los expertos en herramientas que facilitaran la generación de información geográfica de precisión. Por lo tanto no puedo tener más que palabras de agradecimiento para todos con quien tuve el privilegio de compartir tantas horas de trabajo en el Institut Cartogràfic de Catalunya. En especial aquéllos, que no siendo técnicos de formación, depositaron en mí, no sólo sus conocimientos, sino su confianza: Xavier Àlvaro, Mari Luz Ibarra y Toni Ramos en el campo de la fotogrametría, Isabel Ticó y Jordi Clua en el de cartografía.

En mi viaje hacia la tesis todo fue más fácil viendo como mis compañeros de trabajo tenían pasión por los mismos temas que a mí me guiaban. Fue un placer contar con vosotros: Pere, Toni, Joan y Albert. Sois el *dream team* que cualquier persona pueda soñar tener. Miriam Amo y Margarita del Oso también, en sus respectivos proyectos, compartieron nuestras cuitas. Gracias a todos.

El camino ha sido largo. He necesitado veinte años para cerrar un ciclo. Podría poner muchas excusas. De hecho siempre he llamado a esta tesis, la del broquil. Porque mis momentos para pensar y desarrollar aproximaciones a la tesis eran intermedios entre otras tareas: mientras hervía el broquil para hacer la cena, por ejemplo. Pero nada es por casualidad. La idea siempre estaba en el fondo de todos mis pensamiento. Así pues el inconveniente de no tener la tesis como primera o segunda ocupación, también tiene sus ventajas: optimizas todos tus pasos y estás muy atenta a las oportunidades. Veinte años me han dado la oportunidad de ver nacer una disciplina y a su vez, de ver un cambio de paradigma que de no haber estado atenta con una necesidad, no lo hubiera disfrutado como lo he hecho. Y en los momentos en que el cansancio puede agotarte yo he tenido personas que me han pedido que volviera a recoger la toalla. Que merecía la pena. Gracias Petia.

Como siempre me ha resultado difícil separar mis pasiones, al Institut Cartogràfic también le debo mis amistades de largo recorrido. Que me apoyaron cuando cualquier adversidad me sumía en un mar de lágrimas, ellas supieron decirme que siempre aclara cuando llueve: Mónica, Inma, MariÀngels. En especial Carme, la hermana que nunca tuve.

El hecho de que esta tesis esté escrita en Inglés y se entienda, se lo debo en gran medida a los esfuerzos de mi amigo Stuart. Aunque conociéndolo, es posible que sus conocimientos de fotogrametria, él es lingüista, ya sean mejores que los míos de Inglés. Así mismo, entrar en un campo tan nuevo para mí, como eran las redes neurales, hubiera sido muchísimo más difícil si no hubiera contado con el apoyo y los conocimientos de Bea Remeseiro, con quién además he podido publicar el artículo que culmina la investigación.

Ahora vendrían los agradecimientos a los programas de investigación en que el tesinando ha participado y que le han facilitado el camino. Medios como ordenadores, software para desarrollos, acceso a revistas, etc. En mi caso, todo ello, así como apoyo incondicional en cualquier situación, me lo ha proporcionado Fernando. Sin él, esta tesis no existiría.

Y tampoco se acabaría el mundo, porque otros investigadores llegarían a conclusiones parecidas. Por lo tanto no os puedo estar más agradecida a todos, porque yo he sido quién ha disfrutado, gracias a vosotros, del camino.

To Marga

Contents

List of Figures

List of Tables

List of Abbreviations

CasNet	Cascaded end-to-end convolutional neural network
CNN	Convolutional Neural Network
DInSAR	Differential Interferometry Synthetic Aperture Radar
DL	Deep Learning
DSM	Digital Surface Model
DTM	Digital Terrain Model
FCN	Fully Convolutional Neural
GIS	Geographical Information Sysem
HDNN	Hybrid deep neural network
HRV	High-resolution Visible
IRS	Indian Remote-Sensing Satellite
ISODATA	Iterative Self-Organizing Data Analysis Technique
Lidar	Light detection and ranging
MDL	Minimum Description Language
SVM	Support Vector Machine

Chapter 1

Introduction and goals

The extraction of information from images is a discipline as old as the generation of information, and in many occasions it has evolved thanks to the advances that have taken place in other fields with which it has been interacting.

In the generation of geographic information, the most expensive part is the feature extraction from aerial images, in terms of time and human effort. This is traditionally called photogrammetric restitution.

For that reason our general purpose in this thesis is to alleviate these tasks taking advantage of the most innovative techniques in Computer Vision and Machine Learning, that are available today when we try to improve manually treated geographic information analysis.

1.1 Context

When extracting information from images, without a doubt, the most important turning point was the appearance of the scanners that transferred the entire photo library to digital format. With the availability of images in digital format, advances in Computer Vision and Machine Learning techniques can be transposed directly to make it easier to capture the data of topographic elements that appear in scenes. In Cartography, the field which is the object of our study in this thesis, the first techniques applied directly to automate certain processes were made in the classification field, using pixel-clustering methods guided by search techniques for patterns and similar characteristics. The techniques of focusing images have also opened a line of improvement that has led many automatism.

In addition, different sensors, not only those that work with the visible spectrum, open the door to capture different aspects of the same scene. The proliferation of engines using these sensors to observe the Earth's surface is

another challenge in terms of the number of images available and how often they are taken.

1.2 Motivation

In this scenario, to automate the vectorization of elements that appear in the images still remains one of the great challenges of Computer Vision, Machine Learning and Cartography. Since the 80's there have been many efforts in this direction. The elements on which more progress has been made are buildings, roads and fields. In the case of the first two (man-made objects) many congresses have been held. In particular, there are three specialized congresses of four-year periodicity which have been collecting the most important and decisive milestones. In the case of fields, the research has been much smaller. However, the case of the classification of land uses has been very different, since it was one of the first Computer Vision applications that was automated with relatively good results.

In addition, the time spent from when an image is taken until cartographic features are digitalized, more than 80% is dedicated to the manual feature extraction. This fact, together with the need for updated georeferenced information, makes the extraction of geographic information from images necessary in a shorter time.

Our research focuses on automating the extraction of fields and roads to the greatest extent possible, using the different advances in Computer Vision. These advances have a crucial role in how photogrammetry should go from analytic to digital. We will describe some milestones of this transition in this chapter and what have been our contributions to this field.

1.3 Main goals of the thesis

After having worked for a few years in building tools to support the delineation of topographic features from aerial images, we realized the importance of the automation of most of the entire photogrammetric process. Within the workflow that goes from a bunch of aerial images to the topographic map, the activity that requires most dedication by expert human operators is the geographic data features capture, extraction and analysis, to form a topographic map or to build a Geographical Information System (GIS).

Our research focuses on making the delineation tasks easier. In this thesis, we focus on developing robust and straightforward methods to assist operators to the maximum in digitizing aerial imagery features. These algorithms should make it possible to extract important geographic elements to be distinguished by their radiometric similarity, such as fields and roads.

The aforementioned elements, apart from their spectral response, share another important property: they must be recovered based on extracting specific contrast lines, regions and other geographic elements in the images. These elements are the ones that catch the eye when looking at an aerial image.

Given the previous general lines in which this thesis will be focused, our particular objectives are the following:

- To develop highly robust methods to extract fields from aerial images in a semi-automatic way. The fields extracted will be those that can be distinguished radiometrically from their environment.
- To extend the semi-automatic extraction methods to roads, taking into account not only their radiometry, but also the linear contrast elements that delimit them.
- To propose novel methods to delineate the field boundaries automatically. In addition, to expand the variety of fields extracted with the proposed semi-automatic method.

1.4 Contributions

In this context, we start our research relying on these radiometric questions and proposing answers in terms of methods and algorithms. For that reason we propose to face the problem of extraction regions by setting neighboring regions to compete for a pixel using a region competition methodology. The first approach to the problem of field segmentation is a semi-automatic development that helps human operators to automate the tedious time-consuming process of geographic image analysis. The development fulfills the first objective of the thesis. This proposal is described in the following papers:

- Margarita Torre and Petia Radeva. Agricultural field extraction on aerial images by region competition algorithm. In *Proceedings 15th International Conference on Pattern Recognition (ICPR)*, vol. 1, pp. 313-316, 2000.
- Timm Ohlhof, Eberhard Gulch, Hardo Muller, Christian Wiedemann and Margarita Torre. Semi-automatic extraction of line and area features from

aerial and satellite images. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, 2004.

Afterwards, and relying more on contrast linear elements, we propose to explore how to widen our field of work by applying an evolution of region competition to roads. In addition, we base our methods on knowledge about the features to extract following the model that leads the whole semi-automatic process. The second objective is described in a published paper:

- Miriam Amo, Fernando Martínez and Margarita Torre. Road extraction from aerial images using a region competition algorithm. *IEEE Transactions on Image Processing*, vol. 15(5), pp. 1192-1201, 2006.

Later, Deep Learning (DL) burst onto the Computer Vision scene, which upended all the processes of classification of images. So, we propose resuming our research to take DL into the field of geographic image analysis. In this step, we propose a new model that integrates a deep network, capable of learning hierarchical geographic features in a supervised way, into a model-guided energy-minimization framework.

The process not only improved, multiplying by three the success, but allows us to completely automate the image analysis without needing human help. The last objective of the thesis was achieved and described in the paper:

- Margarita Torre, Beatriz Remeseiro, Petia Radeva and Fernando Martínez. DeepNEM: Deep Network Energy-Minimization for Agricultural Field Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 726-737, 2020.

One additional contribution is proposed in the field of monitorization of the movements of a given area of land, called subsidence. It is also important in the pipeline to automatically generate GIS. This research is described in two papers:

- Oscar Mora, Roman Arbiol, Vicenç Pala, Albert Adell and Margarita Torre. Generation of accurate DEMs using DInSAR methodology (TopoDInSAR). *IEEE Geoscience and Remote Sensing Letters*, vol. 3 (4), pp. 551-554, 2006.

- Oscar Mora, Roman Arbiol, Vicenç Pala, Albert Adell and Margarita Torre. Medidas de deformación del terreno a vista de satélite. *Revista Catalana de Geografía*, vol. 12 (31), 2007.

1.5 Thesis organization

The following chapters of this thesis are organized as follows: Chapter 2 describes the background, the detailed problem definition and the state of the art. In Chapter 3 we show the semi-automatic approach to field extraction and how the way of working of human operators changed: from being the principal delineating "actor" to being a supervisor of the process. Chapter 4 describes how the semi-automatic approach alleviated the task of the delineation of roads. The human operator just has to provide some seeds to the system and latter to revise the results. We present in Chapter 5 the full automatic field extraction approach that extracts more than 90% of the fields that appear in aerial images. Chapter 6 describes the additional contribution that monitorizes subsidences. Conclusions and further research form Chapter 7.

Chapter 2

Background and State of the Art

In 2000, almost all the photographs used in the photogrammetric process were in digital format. Although digital cameras were not shipped –the economically viable solutions were still far away–, it was common to trigger the photogrammetric workflow when scanning the analog photographs. It was thanks to the fact that scanners, at an industrially affordable price, achieved resolutions of up to 7.5 microns. This resolution in raster images is more than enough to preserve all the details necessary to collect features, even if in the most demanding cartographic products.

FIGURE 2.1: Image with 5m resolution on the terrain. Contrast lines and details are lost.

Before the introduction of high resolution scanners, it was not possible to try to work directly with digital photographs when extracting topographic elements. Moving from analog to digital photography, if done at low resolution, contrast lines can be blurred. As shown in the Figure 2.1, choosing the appropriate resolution at which the photograph is scanned is crucial to keeping details. This image was taken at 5m resolution, for that reason contrast lines and details are lost.

Since in most of the approaches analyzed, a prior study of the work domain is carried out, we will dedicate the next section to give a brief review of the characteristics that some works have taken into account before starting the extraction. Furthermore, we will describe some preprocesses that leave the scope of action –the images– in better conditions to apply the extraction algorithms with greater guarantees of success.

2.1 Domain analysis

Given the variety of information available and its presentation, in most cases before starting the extraction process, the available data is analyzed and preprocessed. Thus, for example, when RGB images are available, they are transformed to the different color spaces to analyze in which of them we would have a greater contribution of the information to be extracted. Thus, in [1] an analysis of different algorithms based on Wavelets and their performance on different color representation systems is made. Some examples are:

- RGB normalized.
- xyY. Transformations from RGB to xyY, that in [2] is explained in detail, is done in two steps:
  1. It goes from RGB to XYZ by applying the transformation:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.49 & 0.31 & 0.2 \\ 0.17697 & 0.8124 & 0.001063 \\ 0 & 0.01 & 0.99 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

  2. Normalization of the X and Y bands, the signal is.

$$S = X + Y + Z, \qquad x = \frac{X}{S}, \qquad y = \frac{Y}{S}.$$

- HIQ where

$$H = \arctan\left(\frac{\sqrt{3}}{2}\frac{(G-2B)}{(R-G-B)}\right)$$

$$I = \frac{R+G+B}{3}$$

$$H_1 = 180\,H + 180$$

$$Q = \frac{H_1}{1406}$$

- HSV. Being MAX the maximum among (R,G,B) and MIN the minimum, so:

$$H = \begin{cases} 60\left(0 + \frac{G-B}{MAX-MIN}\right), & \text{si } R = MAX \\ 60\left(2 + \frac{B-R}{MAX-MIN}\right), & \text{si } G = MAX \\ 60\left(4 + \frac{R-G}{MAX-MIN}\right), & \text{si } B = MAX \end{cases}$$

$$S = \frac{MAX - MIN}{MAX}$$

$$V = MAX$$

When the location of elements is done using pixel grouping based on radiometric criteria, in most cases it works better in the HSV space, as can be seen in [3]. They aim to identify poplar areas, with three ranges of percentage of land occupation, using one of the commercial programs that provided better results at the end of 20th century: ERDAS IMAGINE Expert Classifier[1].

In this case, the color space that gives the best results is HSV, applying a 3×3 filter to the saturation channel to homogenize it.

2.2 Filtering

Once the workspace has been chosen, the next step is to highlight the elements to be extracted as much as possible, mainly by filtering the image information. One of the filters most commonly used is a convolution of 3×3 to homogenize the saturation channel, as we described above.

In the case that for the elements to be extracted not only the radiometric similarity is taken into account, but also the contrast zones, it is necessary to use filters that, at the same time, enhance these contrasts by keeping the homogeneity. Below we describe some of these filters:

- Anisotropic diffusion in the space scale. In [5] a non-linear anisotropic variation is introduced in classical isotropic diffusion, based on the magnitude and intensity of the gradient image:

$$I_t = \text{div}(c(x,y,t)\nabla I) = c(x,y,t)\triangle I + \nabla c \times \nabla I$$

  where I_t is the image at resolution t, that also represents the space-scale parameter. So instead of the uniformly blurry result provided by the

[1] ERDAS IMAGINE Expert Classifier [4] provides an approach to multispectral classification of images, based on knowledge rules, with post-process refinement and modeling of the result in GIS.

linear heat equation, the effect of a diffusion anisotropic softens the input image preserving the discontinuities (edges).

- *Mean shift filtering* [6]. Filter whose evolution takes place along the maximum gradients. A detailed description can be seen in [7].

In any case, it should be noted that as the sophistication of the filter increases, it becomes more difficult to select the appropriate parameters for each image. Moreover, frequently the time necessary to adjust them makes their application uneconomic. So, the advice is to choose simpler solutions.

2.3 Generation of Digital Surface and Digital Terrain models

With the photo library in digital format, the main objective was that many of the photogrammetry tasks move to a digital environment.

The first step was taken in 1995, when the first Digital Surface Model (DSM) were automatically generated.

2.3.1 Digital Terrain Models

A Digital Terrain Model (DTM) is a numerical data structure that represents the spatial distribution of a quantitative and continuous variable. The best known type of DTM is the DSM. It is a particular case in which the variable represented is the elevation of the surface of the earth and includes all the objects it contains, whereas DTM represents the surface of bare soil without any object, such as vegetation or buildings. The graphic difference between these concepts is shown in Figure 2.2, where can be seen that DTM recovers the bare Earth surface, meanwhile DSM points could be found on the top of buildings.

2.3.2 Photogrammetry matching algorithm

The attempts to automatize DTM generation started thirty years ago with the feature-based correspondence algorithm [9] that has been used and improved to obtain several off-the-shelf products, such as MATCH-T DSM [10]. These approaches generate regular grids or extremely dense point clouds from stereo imagery and guarantee reliable and accurate results.

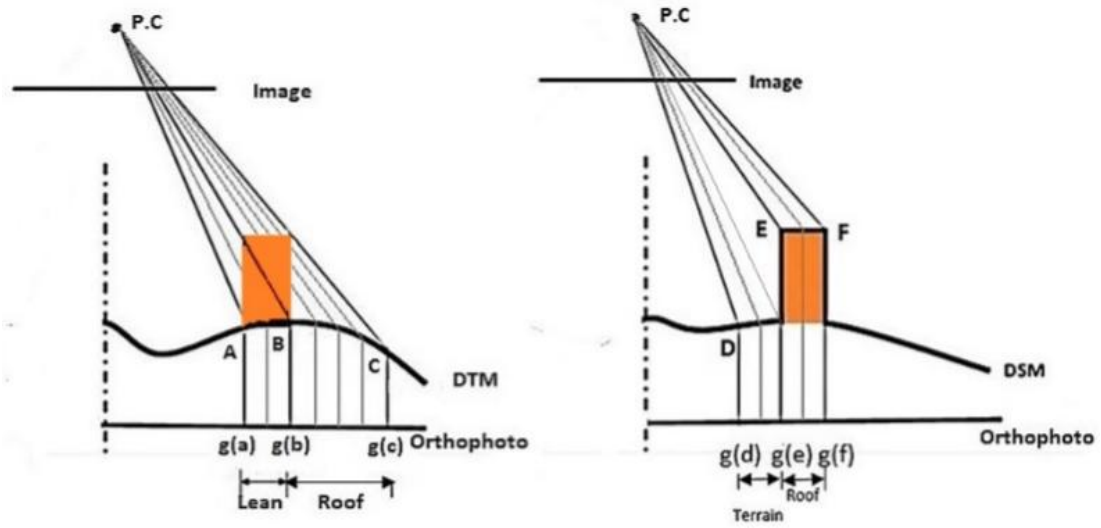


FIGURE 2.2: The conventional DTM on left hand side, and on the right is DSM with building represented. (Preprint [8])

From a pair of images with stereoscopic coating, using a three-stage algorithm:

1. Generation of points of interest for each image, using a search operation for high contrasts (edges) and their main characteristics using the Förstner operator [9].

2. Stereoscopic correspondence between points of interest following the epipolar lines. An example of these lines is shown in Figure 2.3. Epipolar lines are used in digital photogrammetry to solve the correspondence problem between two images. For a given 2D point in one image (x_1), the corresponding point in the other image will be located along the projected line –described by Q_1 (projection image point) and x_1– projected from Q_2 (projection point of the second image). With this approach, in a correspondence matching, the searching dimensions will be reduced from 2 to 1.

3. Generation of the surface that represents the elevation model (DSM), by interpolation of the previous points, using finite elements. The automatic system can even guide the human operators to those areas where it is already expected that the correspondence has not been successful.

With the need to have true orthophotos [2] or even with the necessity of having city modelling applications, the techniques of generating DTM have been

[2]Orthophotos in which occlusions have been eliminated.

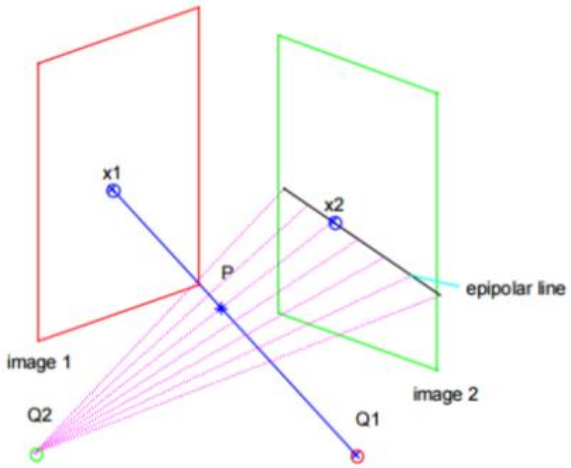


FIGURE 2.3: Corresponding epipolar line of a point, over two overlapping images. Q_1 and Q_2 are the optical centers of the images. x_1 the selected point in the first image, x_2 one tentative point in the second one.

reinforced with the non-ground objects removal to achieve bare earth DTM using robust filter methods.

2.3.3 Lidar

Lidar, used as an acronym of light detection and ranging, is a surveying method that measures distance to a target by illuminating the target with laser light and measuring the reflected light with a sensor. Differences in laser return times and wavelengths can then be used to make digital 3-D representations of scenes. The cloud of points obtained by Lidar is another way to obtain DSM.

The DSM generation by using photogrammetry techniques or Lidar point clouds both have their advantages. In terms of accuracy, Lidar is hard to beat, but the time necessary to capture all the area and to filter the information is also a setback of the Lidar approach. Moreover, it is impossible to map the whole area homogeneously and without gaps and overlaps, for that reason it requires a further interpolation to generate a regular mesh.

2.3.4 DInSAR

DTM not only is necessary to build orthophotos, but also to measure surface deformations due to groundwater extraction, excavation of tunnels, slow processes of dissolution and lixiviation of materials, consolidation of soft soils, organic soils, etc.

Subsidences are defined as slow and gradual movements of the terrain or built surface. These may affect all types of terrains, and are caused by tension-induced changes for many reasons, such as the aforementioned.

The measurement and monitoring of land subsidences are the major components of the auscultation of infrastructures in their construction and monitoring phases. It is important to recognise and evaluate subsidences before they cause damage, and their measurement relies on very specific instrumentation.

The detection of small altitude variation by using Differential Interferometry Synthetic Aperture Radar (DInSAR) opened a broad range of possibilities. The altitude information delivered by DInSar is computed from the phase difference calculation between pairs of radar images, as is shown in Figure 2.4.

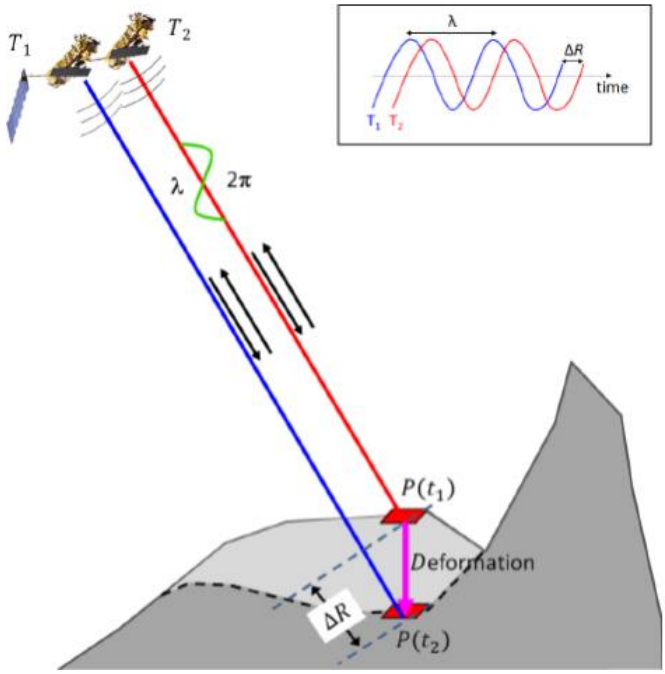


FIGURE 2.4: How DInSAR gets DSM.

2.4 Orthoimages

One of DTM/DSM most extended uses is the generation of orthoimages. For many years, the generation of DTM/DSM was the bottle-neck of the process. For that reason the generation of an accurate elevation model was a candidate to be automated as soon as possible.

An orthophoto is an aerial photograph geometrically corrected such that the scale is uniform and it follows a given map projection. For that reason an

orthophoto can be used to measure true distances, because it is an accurate representation of the surface of the Earth. It has been adjusted for topographic relief, thanks to a DTM, lens distortion and camera tilt.

When intending to map a wide area, it is necessary to have a set of images that cover it. Adjacent images should have an overlapping space to ensure the information recovery in three dimensions, thanks to the stereoscopy effect. The processes of linking overlapping images by relevant points identified in them was automated more than a decade ago. At present, orthoimage generation from aerial images is a simple and straightforward process. The three steps necessary to generate orthophotos from the images of a flight are the following:

- Single orientation of each frame, and block aerotriangulation by adjustment of beams with autocalibration from the points of support and the observations made in aerial images. This aerotriangulation results in the external orientation of all the frames.

- DTM generation, which is necessary to have a digital model that is sufficiently correct to orthorectify the flight within geometric prescriptions. The detection of break-lines is even more important than the raw interpolated mesh obtained from homologous point correlation techniques. These lines represent a wire model that triangulates the area. In each area the correspondence between homologues takes place.

- From the correctly oriented frames and the DTM, the process of orthorectification of the frames and the generation of the orthophoto mosaic is carried out with the corresponding radiometric adjustment. This last step standardizes the visual appearance of the whole product. It is necessary to find the smoothest path along the overlapping area (Figure 2.5), to see the orthophoto as a seamless output.

Once we have the images of a flight, the generation of orthoimages is much faster than obtaining a topographic map, since it requires less human interaction. Orthoimages offer a quickly available look at the land uses. So, it is a product that is in great demand to update geographic information systems. Currently, the orthophoto acts as a precursor to topographic maps.

2.5 Linear man-made objects: roads

The road network has always been a main element in maps due to its structuring character in a Geographical Information System (GIS). Moreover, these

FIGURE 2.5: Overlapping area between two orthophotos to create an orthoimage

elements are crucial to analyze the mobility of people and goods in urban planning. Nowadays, this role has increased because of the addition of automatic road navigation and unmanned vehicles.

2.5.1 Road appearance in aerial images

To characterize the mobility network in term of its appearance in aerial images, it is worth bearing in mind that this network guides the way of observing a map.

Its geometrically almost continuous appearance and its clear radiometric response mean that it has been one of the man-made features that has been approached first for automatic extraction. Given their extension, roads are elements that appear in different images, and for this reason their extraction can be seen from different perspectives and can also be an element to give continuity to the different orthoimages that form an orthophoto.

On the other hand, the local structure of roads is complex and, very often, the road segments are irregular. Along a road it is possible to find shadows of trees or occlusions caused by buildings. So, the larger image resolution, the more difficult it is to accurately extract the whole infrastructure. This phenomena is shown in examples of Figures 2.6 and 2.7.

In Figure 2.6 it seems easy to locate and extract main roads. On the other hand, when getting closer to the terrain (Figure 2.7), some occlusions and shadows will interrupt the road continuity.

FIGURE 2.6: Image of road extraction at small scale.

FIGURE 2.7: Image of road extraction at large scale, where red circles point out occlusions and radiommetric discontinuities

To sum up, although road image appearance can be described in terms of nearly homogeneous geometric and radiometric elements, there are image characteristics that affect their appearance such as spectral and spatial resolution, weather, light variation and surrounding elements.

2.5.2 Extraction methods

Given the challenging characteristics mentioned above, the first attempts at automatic feature extraction were made with roads, and very promising results were obtained. The classification of the algorithms that have been applied in road extraction throughout history can be made from different perspectives.

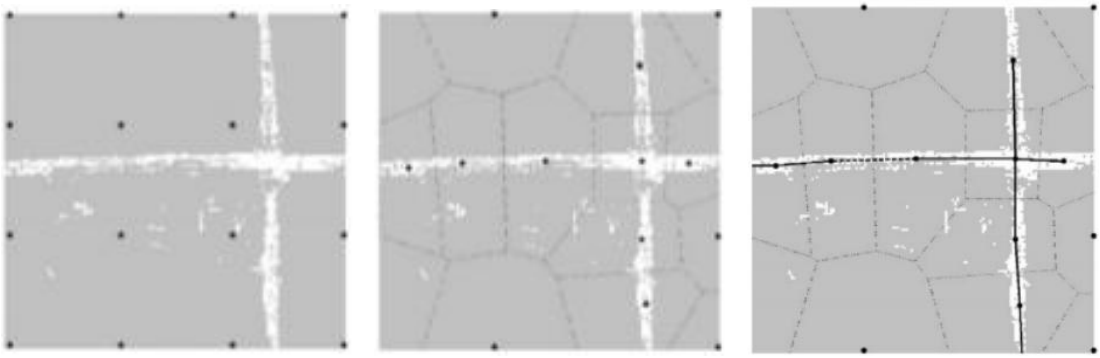
One possibility is to classify the extraction methods depending on the information that is taken into account and go from feature to the knowledge. It was also a trend to classify these methods taking into account whether or not operator assistance is requested: semi-automatic or fully automatic methods. But all these classifications fade from the moment the concept of Deep Learning appears on the scene, in 2006 [11].

Its particular evolution has already gone from trying to classify the information of the image as it is, to embedding high-level and multi-scale information. For that reason we will describe some methods by dividing the history into before DL and after DL.

Image Classification before DL

With respect to radiometry, the roads to be segmented are based on their characteristic smooth variations in homogeneity and outlined by other elements with different homogeneity values. Classification-based methods mainly use photometric and texture features of a road, and their accuracy can be affected by the misclassification between road and other spectrally similar objects.

Unsupervised classification methods do not need training samples, and the most common algorithms are based on clustering processes, which include K-means ([12], [13], [14]) and mean shift ([6], [15]), which are non-parametric iterative methods for locating the maximum of a kernel density function. One example can be seen in the Figure 2.8, where the three steps in the process of road extraction by using K-means are shown.



(a) (b) (c)

FIGURE 2.8: Three steps in the process of road extraction by using K-means: (a) initialization, (b) clustering and (c) linking. (Preprint [16])

When the classification is supervised by training with labelled samples, the reliability depends on the significance of the labelled samples. In the last decade, these methods have taken off because they have the support of neural networks. But, the origin of this approach was Support Vector Machine (SVM) [17], which was firstly proposed for classification and regression analysis. The first attempt to exploit SVM classifier using edge-based features (gradient and intensity) was done at [18]. The selection delivered by SVM is shown in Figure 2.9 where the input image is composed of edges from Canny operator [19], and the results are either road centerlines and parallel sides.

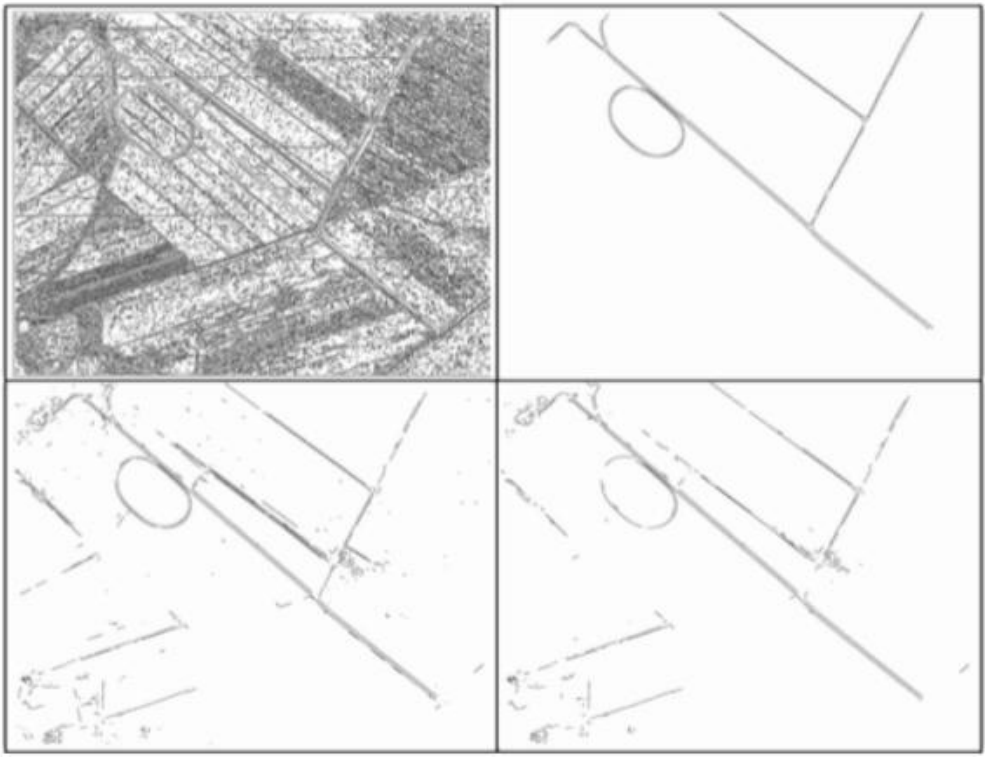


FIGURE 2.9: Example of SVM to select edges delivered by Canny operator [19] (top-left). Below, left image shows road edges and right edge-pair. (Preprint [18])

Maximum likelihood classifiers [20] are one of the greatest exponents in the Bayes classifier. They use the Gaussian probability density function and calculate the attribution probability for each pixel, then put the pixels into the maximum probability categories. This approach is applied in [21] to extract roads.

Knowledge-based methods

Due to their design constraints, the majority of the roads can be split into parts where the curvature is subject to small variations. So, it is difficult to extract roads from aerial images only using the radiommetric information (signal and texture). For that reason, knowledge about the element to extract has been included in the process ([22], [23], [24]):

- Clues on feature level. Most of them were given by human operators, for that reason the first approaches started in a semi-automatic way. The template matching method [25] extracts the roads from the image according to extracted seed pixels or a specific template to form an initial road network. The edge and parallel lines approach [26] uses the fact that

the road edges are usually parallel lines. The completition of the road that starts with seeds of high relevance is done by two approaches: path finder or path optimizer. Both try to cope with the lack of information or misinformation found when building the linear element. This approach opens the door to model methods, such as the snake model [27]. When using parameter models, energy functions can be used to operate on, and the results are the ones that reach the maximum value of this function. The most common parameter models usually extract some structural elements according to the relationship among them. One example is shown in [28], which started by extracting lines at coarse scale, these lines initialize ribbon snakes at coarse scales, where roads often appear as bright, more or less homogeneous elongated areas. As these roads are disturbed, the evidence for the road in the image can only be exploited with additional constraints. These constraints are low curvature and constant width of roads as well as the connectivity of the road network. The strategy is to optimize the center of the ribbon snake. The steps of this approach are shown in Figure 2.10. They go from evidence of road, to the extraction of optimal path, followed by the verification by optimization of width and end with the acceptation of a hypothesis with low variation of width.

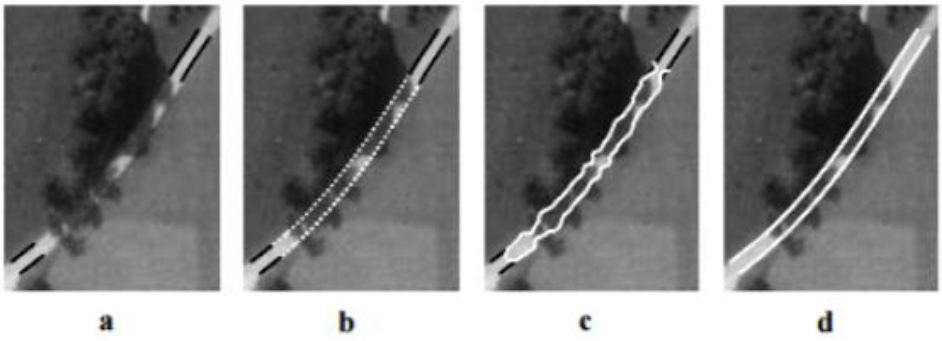


FIGURE 2.10: From evidence of road (a), extraction of optimal path (b), verification by optimization of width (c) and acceptation of a hypothesis with low variation of width (d). (Preprint [28])

All the image evidence used can be reinforced by using filters, so the road network is extracted based on enhanced road pixels. Although the abuse of these filters could reduce the accuracy of the results.

When defining a model, it is necessary to look for evidence that will follow a predefined model. [29] proposed a method for road and bridge

detection from Synthetic Aperture Radar (SAR) images by using geometric features to extract general objects. When the extraction is completed, mathematical morphology tools and Hough transformation [30] were adopted to extract the small regions and to connect the discontinuous segments.

- On the object level, which includes multiresolution and regional statistical analysis. The multiresolution analysis method improves the precision of road extraction by combining different resolutions of the same image, as if the image were a pyramid of different scales. For example in [31] the road centerlines are extracted from high resolution imagery based on multiscale structural features and SVM.

- On the knowledge level, such as the multi-source data fusion method, the extraction is assisted by road networks in existing road databases or other data, such as vector maps and documents. The road characteristics and other knowledge-related theories extract roads based on their own characteristics, such as spectrum and context features. These methods also have made advances in complex scenes, but the design of such algorithms is more complicated and the operating efficiency is not ideal.

Deep Learning

The early DL approaches, used to extract information, can be understood as classification algorithms, because an input image results in another one where the pixels have been labelled depending on their membership in a class or not. But, unlike traditional algorithms that just use low-level information for road extraction, DL can reduce false detection by embedding a lot of high-level and multi-scale information. In [11] the concept of DL was proposed, which provides the basis for techniques that later had been widely used in classification and object detection, mainly applying two kind of networks: the Convolutional Neural Network (CNN) and the Fully Convolutional Neural (FCN).

The CNN [32] is composed of layers focused on extracting local features in an image. At the end of the convolutional layers, the fully connected layers integrate those high-level local feature maps into a n-D vector by the inner product operation to predict the label to be assigned to the image. The CNN architecture is shown in Figure 2.11.

Since the architecture of this network does not predict the label for each pixel, it is not suitable for image classification as it is. In 2015, [34] propose

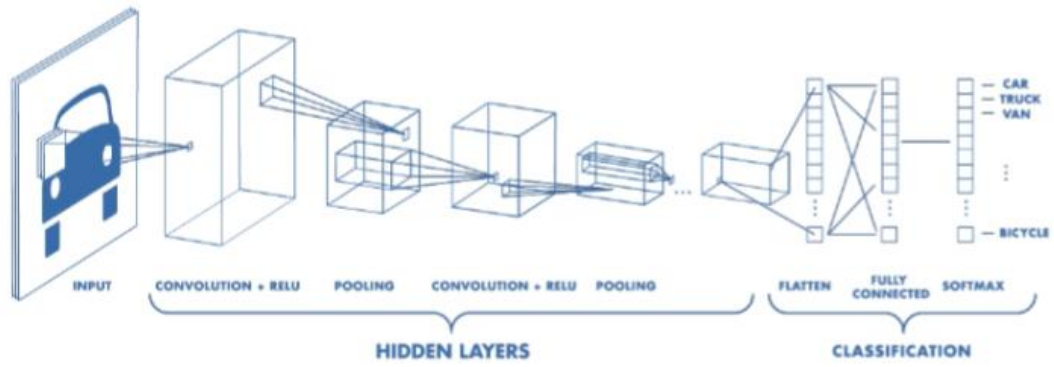


FIGURE 2.11: CNN structure. (Preprint [33])

FCN that replaces all the fully connected layers with convolutional layers to produce an arbitrary-size output. The structure is shown in Figure 2.12.

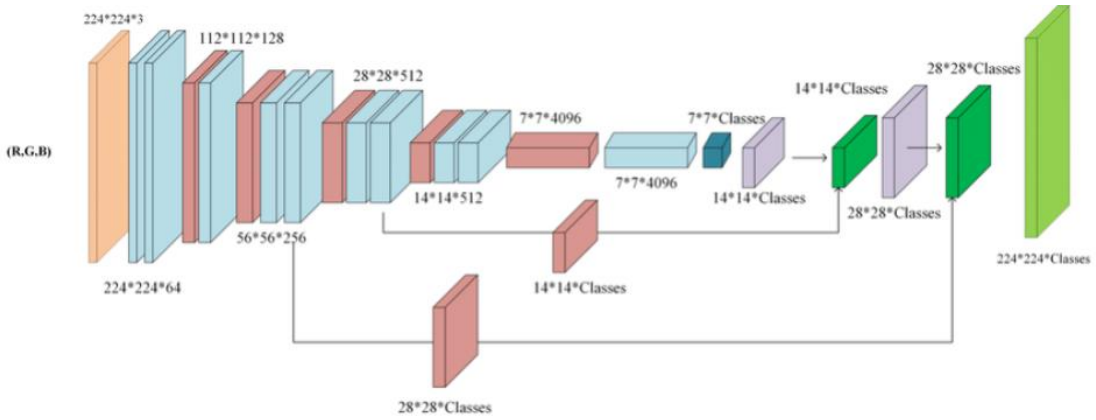


FIGURE 2.12: FCN structure. (Preprint [35])

Since these methods just look into the information contained in the image, they could be put into the category of image classification that just uses low-level information. But, unlike these algorithms, CNN reduces false detections by embedding a lot of high-level and multiscale information. An example is described in [36], where CNN is used to extract buildings and roads directly from raw remote sensing images.

In case of road extraction, although CNN has achieved quite good results, the local field of work of convolutions yields some missing patches, which results in incomplete roads. In that case some improvements have been made by FCN, due to the fact that their output are high-level features that are completed

as if they used context information. [37] proposed a cascaded end-to-end convolutional neural network (CasNet). It contains two convolutional neural networks: a road detection network and a centerline extraction one.

In terms of road and paths, more important than centerlines, are edges. Edges in an image are sharp variations of the intensity function. In grayscale images this applies to the intensity or brightness of pixels. In color images it can also refer to sharp variations of color. An edge is distinguished from noise by its long range structure. And the importance of edges in road extraction is because some road boundaries fall over edges.

All the previous approaches described when working with DL are only for end-to-end and pixel-to-pixel training, which can not fully acquire the abstract information on each convolution layer, for example the edges. The first attempt was to use a DenseNet [38] to extract a feature vector for each pixel, and classify them by using a SVM in terms of edge non-edge. Moreover, edges can be recovered by Deep Supervision approach, which helps side-output layers to produce multi-scale density predictions and a fusion output, which can make full use of the complementary information between different convolution layers. It was done in [39] with the introduction of the Holistically-nested Edge Detection (HED).

2.6 Agricultural fields

Among the variety of elements that form a topographic map, the most prominent, for the amount in which they appear, are buildings, closely followed by agricultural fields. Therefore, their delineation takes a great amount of time, if it is done from the bare aerial photographs.

The fact that they are so abundant means that, on many occasions, their exhaustive delineation is uneconomical, and they are only collected selectively. The exhaustive field segmentation only takes place when compiling a cadastral map. Moreover, the segmentation could also be applied for GIS updating. Since the cycles to cover the surface of the earth with photographs are getting shorter and shorter, the need to update the GIS is more urgent.

2.6.1 Fields' appearance in aerial images

As a counterpoint to the segmentation of man-made objects, fields have more natural variability. Therefore, many of the approaches that have been made

try to classify a reduced set of crop area types, for example by identifying their spectral responses to a given sensor and trying to locate them in the images.

Some of the fields have, as their main quality, their radiometric homogeneity, that is a certain similarity in the pixel values that integrate them. Another set of fields have some of their boundaries well defined. For example, they are delimited by other fields with strong levels of homogeneity but with different statistical parameters. Moreover, they can be limited by roads or forested masses. On the other hand, there are other cultivation areas that can only be recognized by contextual interpretation. This is because the rest of the information in their environment leads us to label them as fields. Some examples ares shown in Figure 2.13. The examples go from some fields surrounded by wood, on the top, to fields with different levels of homogeneity, some of them surrounded by paths, the image in the middle, and assorted textured fields, on the bottom.

2.6.2 Extraction methods

Here, we will describe the different methods chronologically as they have appeared.

Merge: from pixel to region

Merge or grouping techniques put together all the pixels that share the desired properties. Among the most relevant works in this field, the first ones use the spectral signatures and the distances to them. Two examples are shown in Figure 2.14: Iterative Self-Organizing Data Analysis Technique (ISODATA) [40] and the K-means clustering. The four images show the different classification degree when dealing with different road materials, surrounded by fields.

Advanced clustering techniques, such as fuzzy segmentation, allow these methods to better retain the radiometric variation of the agricultural fields. Fuzzy C-Means clustering approaches are found in [42], used together with spatial constraints.

Unsupervised clustering methods have several limitations: the number of clusters are often unknown, and the predefined parameters often deliver over/under-segmented results, requiring further split and merge procedures in combination with interpretation of the segmented images. Another problem related with this approach is that the more detailed the resolution, the greater the number of spectral misclassifications.

(a)

(b)

(c)

FIGURE 2.13: Different field appearances: (a) some fields surrounded by wood, (b) fields with different levels of homogeneity, some of them surrounded by paths and (c) assorted textured fields.

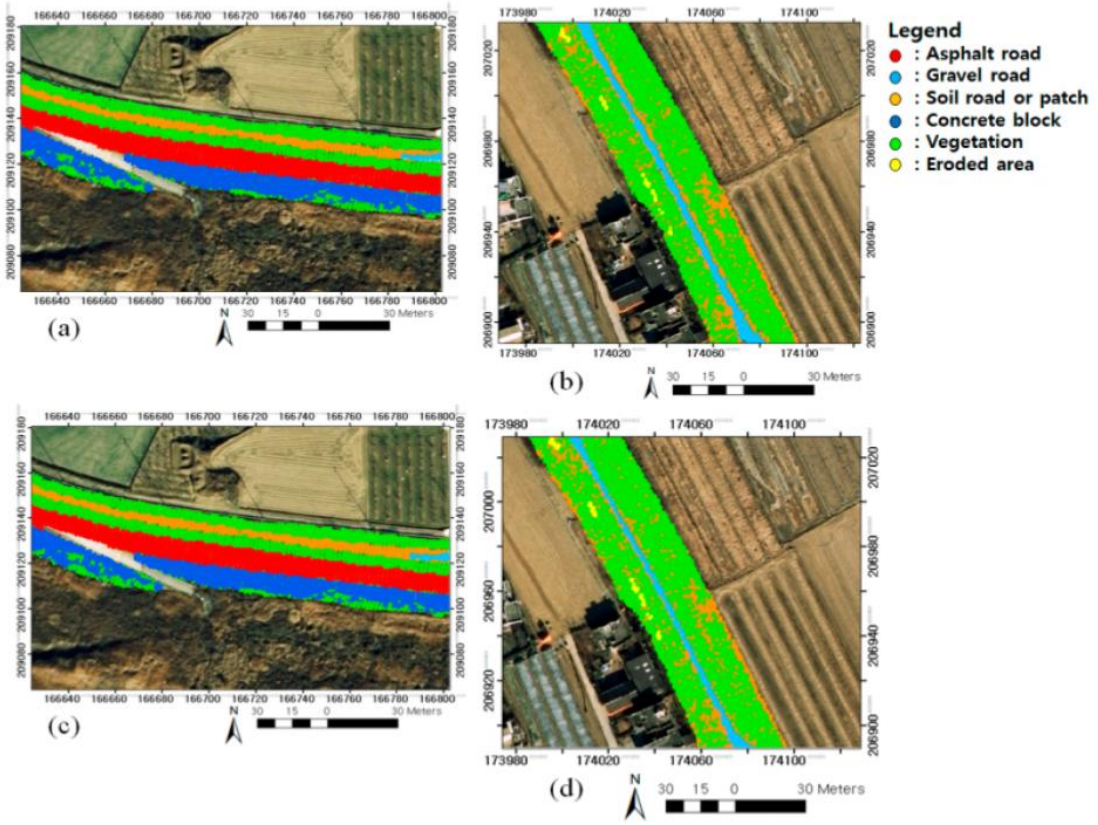


FIGURE 2.14: Multiple level components identified by ISODATA clustering (a,c) and K-means clustering (b,d). (Preprint [41])

Sometimes, to solve these limitations the process becomes semi-automatic by delivering clustering centers. Hierarchical decisions, inspired by the principle of photo-interpretation, are a way of reducing human interaction.

In the case of regions with texture, the algorithms started to extract this information by applying Gabor Wavelets [43]. One example is shown in [44] that provides invariant results to the rotation and to the scale, since each filter to be applied is a symmetric quadratic function. The results presented there establishes that the classification produces good results when acting again upon the training set. On the other hand, the results for data not previously seen by the system are not good. Both arguments are related to the development of DL, since rotated filters are introduced as convolutional layers, and training on purpose is the best way to improve DL performance.

Split: from region to pixel

Another strategy is the division of the image in different areas, and this process is mainly led by edges. They are used in automatic and semi-automatic techniques for boundary delineation. Broadly speaking, edge detectors extract

edges by calculating gradients of local brightness. Among the different methods found in the literature, the Canny detector [19] is accepted as a gold standard. One example described by [45] is the Line Segment Detection algorithm, aimed at detecting straight contours in images [46].

These methods perform reasonably well in regularly-shaped agricultural fields, but fail when dealing with heterogeneous datasets. Additionally, they are generally sensitive to intra-class variability which leads to over-segmentation. This phenomenon is shown in Figure 2.15, where the image on the right has a lot of edges inside textured field. The image on the left has few disturbing elements inside fields, due to their lower intra-class variability.

In any case some aid could improve the results. For example, the selection of the examples of each area to be classified is feasible if some previous GIS of the area is available. GIS data are also used for the analysis of declassifications that occur at the borders, generating a buffer around them. For this reason, these elements do not enter into the determination of the classes to which they are assigned, which will be the one with the highest percentage of assigned pixels.

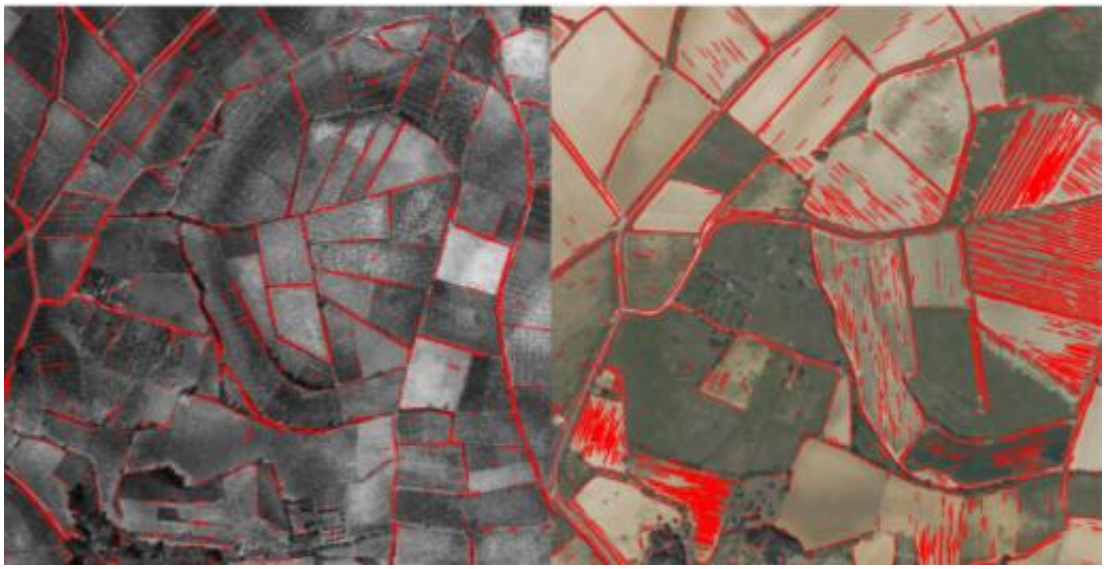


FIGURE 2.15: Lines are detected with Line Segment Detector algorithm [46]. (Preprint [47])

Models to split and merge

The extraction of information from the images aims to establish a correspondence between the characteristics of the image and a series of possible models. There are two types of models: rigid and deformable. In rigid models, the shape and geometry of the object to be extracted are known previously, and

with the image data the few parameters that describe them are determined. Deformable models are specified with much more general restrictions, such as tolerances in curvatures and degree of required continuity, characteristics to be located in the images. These are models that facilitate the integration of the two restrictions that we must have in our work: geometric and radiometric.

Snakes is a special case of the more general technique of matching a set of deformable models with an image through the minimization of an objective function that in this case takes the form of energy. It will be described in depth in Section 4.5.

Energy-minimization theory [48] delivers a common framework to unify different model-based approaches, such as graph cuts [49], random walker [50] and shortest path [51].

The simulation of hierarchical decisions as taken by a human operator is introduced with the application of these algorithms to the different levels of resolution of an image. This process involves different scales and reinforcing the importance of the extracted elements as resolution dissolves. A multi-scale segmentation strategy may be useful in determining which segmentation is "correct", simulating the human operators' decision.

The arrival of DL

With the appearance of DL and the networks that we have already described in Section 2.5.2, the merge process occurs with the use of networks for classification. Likewise, the division is also done using networks that can extract edges or take into account the intermediate results of each layer.

The pixel classification by CNN can be done either by applying SVM to the output, and it delivers the more probable label, or by using logistic regression, which can simultaneously fine-tune the whole pre-trained network and predict the class label in the way probability distribution does.

Classification is not enough to handle spatial features. Fields need to be drawn and distinguishes from their neighbors. So, the use of spatial features not only improves classification, but also delivers image feature representations. For that reason some modifications in CNN have been added, such as in [52] that proposed a hybrid deep neural network (HDNN) by dividing the maps of the final convolutional and the max-pooling layers into multiple blocks of variable receptive field sizes or max-pooling field sizes. This approach enables the HDNN to extract variable-scale features for detecting features at different scales.

Chapter 3

Region competition. Semi-automatic field extraction

3.1 Introduction

Moreover, some processes to get the aerial images ready to delineate the cartographic features were automated. For example, the necessity of human interaction in the frame orientation tasks was considerably reduced.

At this point, many of the photogrammetric projects were prepared from the flight to the data capture in a few days. Therefore, the obvious questions were:

- How to speed up the next process bottleneck: the feature digitalization?
- And, if that were possible, what were the most tempting features to automate from the point of view of their abundance and relevance to mapping purposes?

In 2000, **the automatic extraction of buildings** was the most widely treated topic in Central Europe. Almost all the research was focused on defining models according to the taxonomy of the buildings to be extracted.

Given the constructive characteristics of Central Europe, these models differ considerably from the type of housing found in the domain that we will work on, the Mediterranean area.

In the case of **fields** it is difficult to define models that recover their different appearance. Fields have more natural variability than man-made objects. On the other hand, their shape can be adapted to linear models.

Another reason to try to automate this extraction to the greatest extent possible is the proliferation of images of the same zone, at higher resolution, dynamic and spectral range, etc. In addition there is increasing demand for detailed land use information.

All these considerations make the task of capturing and interpreting complicated, since it often requires the handling and mixing of various sources of information. Therefore, it was mandatory to think about aids to speed up this delineation task.

Since the 80's there have been many efforts to automate the extraction tasks. In the case of land uses, the particular investigation of fields has been much smaller than the analysis of land covers to obtain land uses. Some of this research would be partially applicable to this case, especially when it is from a discipline with an extensive accumulated body of work done.

3.2 Our approach to field segmentation

We consider the problem of segmenting agricultural fields in digital aerial images by using a generalization of region growing techniques [53] combined with deformable models [27]. This mixed approach is called region competition [54]. The goal of this approach is to alleviate the tasks of digitizing the field boundaries, to obtain the vector representation of the boundaries that appear in an aerial photo.

Our aim was to segment areas that are homogeneous enough to be represented by a Gaussian distribution, and different from the neighbor regions or delimited by lineal features like roads or rivers. Due to these characteristics, regions can be segmented by a combination of region growing and deformable models.

Deformable models, such as snakes [27], are defined as elastic curves that dynamically adapt a vector contour to a region of interest by applying energy minimization techniques.

At the same time, given the problem of field segmentation, we need a region-growing approach to divide the raster image into homogeneous regions. Region competition combines the best features of snakes/balloon models and region growing techniques. In operation time, these techniques are applied to the case of having only two regions: the field to be segmented and its complementary.

3.3 State of the art

Usually, the area segmentation techniques are focused on pixel grouping approaches and their further classification [55]. These techniques do not have information about the region number and location, neither do they control the

boundary shape. The snake contribution consists of recovering the boundary information refining a coarse initial curve. Other techniques for region segmentation that preserve the information details are presented in [56]; however, often over-segmented results are delivered.

Region growing [57] controlled by the snake constraint generates region boundaries in a similar way to the manual operation. The region competition algorithm is based on an energy minimization approach that actively optimizes the region contours and updates the probabilistic distribution parameters of the region to be segmented.

The existing techniques of snakes/balloon models, region growing and Minimum Description Language (MDL) [58] can address different views of the segmentation problem. The proposal unifies them within a common statistical framework to take advantage of all of them [54]. Using this strategy, the preservation of topological features of the agricultural fields guides and makes more robust the pixel aggregation process of nearly homogeneous regions.

3.4 Our semi-automatic proposal for field extraction

Our experiences in a cartography productive environment have shown us how difficult full automation is when extracting geographic information, like the elevation terrain model [59]. For this reason and to start in an unexplored field in terms of automatic feature extraction, we implemented a semi-automatic tool for fields segmentation. This choice is also reinforced by [60] and [61] that explain that due to the lack of maturity of automated extraction tools, the best way of increasing the productivity consists in designing semi-automatic tools for image processing. Often the user is required to provide an initial position of the feature to be extracted and to perform an adjustment with the information delivered by the image. When the automatic process delivers some imprecise results, the operators feel more comfortable when controlling rather than looking for incorrect results.

We apply the general region competition algorithm [54] to the agricultural field segmentation. We make it the most operative possible by studying its dependency on the parameterization, as well as implementing convergence criteria and validating it in a practical environment.

In the next two sections we will introduce the region competition approach and we consider its applicability to the fields segmentation.

3.5 Unified frame for snakes and region growing

The type of fields that will be extracted with our proposal are:

- near-homogeneous fields, whose statistical parameters are different from the ones corresponding to their neighbors,
- fields with quite well defined boundaries, i.e., other near-homogeneous fields or linear elements.

There are some other kinds of fields that can only be identified by contextual reasoning, where the interpretation knowledge leads us to identify them. Those fields, however, can not be extracted by our development.

Our approach is a combination of deformable models and region growing techniques in a statistical framework under the MDL environment.

In general, the automatic feature extraction from images establishes a correspondence between some image characteristics and models. In our particular case, deformable models are used, which describe general geometric and radiometric characteristics: degree of continuity and kind of radiometry to be recovered. Snakes is a special case of a general technique that matches a set of deformable models with an image by minimizing a cost function. That function represents, in our case, an energy function, composed by a weighted combination of internal (model) and external energies (image).

$$E_{image} = \omega_{edge} E_{edge} + \omega_{region} E_{region} \tag{3.1}$$

where E_{edge} is based on the image contrast and will attract boundaries to contours of high image gradients, whereas E_{region} pushes the snake to enclose quite homogeneous areas.

The first attempt to recover fields from aerial images did not take into consideration the E_{edge} component. This addition really improved the first results obtained in [62], where the model was only based in radiommetric homogeneity, that we define as follows:

A region R is considered homogeneous when its intensity values are consistent with the ones that will be obtained in case of being generated by a family of pre-established probability distributions $P(I|\alpha)$, where α are the distribution parameters.

Our aim is to represent a continuous grayscale image by a vector set representing the fields boundaries by MDL. For that reason, we represent the image broken down into region entities. So, our aim is to represent the image as a

degraded version from an ideal image that is assumed to be piece-wise continuous [54]. Let $\Gamma = \bigcup_{i=1}^{M} \Gamma_i$ be the segmentation boundaries of the entire image, where $\Gamma_i = \delta R_i$:

$$E[\Gamma, \{\alpha_i\}] = \sum_{i=1}^{M} \left\{ \frac{\mu}{2} \int_{\delta R_i} ds - \int \int_{R_i} \log P(I_{(x,y)}|\alpha_i) dxdy + \gamma \right\}, \tag{3.2}$$

where I is the input image, the first term in (3.2) describes the curve length of ∂R_i that is the boundary of R_i, and μ is the code length codification. Since in general all the segments are shared by two adjacent regions, this term is divided by 2.

The second term is the addition of the cost of codifying the intensity of each pixel (x, y) inside the region R_i with probability $P(I|\alpha_i)$. γ stands for the codification length to describe R_i, assuming that the coding cost is the same for all regions.

The energy in the previous equation depends on two groups of variables: the segmentation of Γ and the parameters α_i. So, the steepest descendent is the way to obtain the functional minimization. Thus, for any point $\vec{v} = (x, y)$ on the contour Γ, its position in each time step will be given by:

$$\frac{d\vec{v}}{dt} = -\frac{\delta E[\Gamma, \{\alpha_i\}]}{\delta \vec{v}}$$

applying it to the formula (3.2) would be the equation of motion of $\vec{v}$:

$$\frac{d\vec{v}}{dt} = -\sum_{k \in Q(\vec{v})} \left\{ -\frac{\mu}{2} \kappa_{k(\vec{v})} \vec{n}_{k(\vec{v})} + \log P(I_{\vec{v}}|\alpha_k) \vec{n}_{k(\vec{v})} \right\}$$

where $Q(\vec{v}) = \{k \,|\, \vec{v} \in \Gamma_k\}$ and $\kappa_{k(\vec{v})}$ is the curvature in Γ_k. The sum is made over all regions R_k, if $\vec{v}$ is in Γ_k.

In the case of a point in common between two regions R_i and R_j, as the curves Γ_i and Γ_j they have opposing normal vectors: $n_i = -n_j$, so $\kappa_i n_i = \kappa_j n_j$, the movement equation results for $\vec{v}$ is:

$$\frac{d\vec{v}}{dt} = -\mu \kappa_{i(\vec{v})} \vec{n}_{i(\vec{v})} + \Big(\log P(I_{\vec{v}}|\alpha_i) - \log P(I_{\vec{v}}|\alpha_j) \Big) \vec{n}_{i(\vec{v})}. \tag{3.3}$$

According to this formula, regardless of the smoothing term, the movement of $\vec{v}$ is determined by the proportion of the similarity test. Therefore, the border will move along the normal, with the magnitude and the sign determined by the previous proportion.

As shown in Figure 3.1 adjacent regions compete for the property of pixels along their borders, taking into account the softness constraint. Depending on the likelihood test result, a movement along the normal at the point $\vec{v}$ produced.

Therefore, the algorithm is called competition between regions, which results from the unification of the algorithms of growth of regions and snakes with MDL, within a statistical framework.

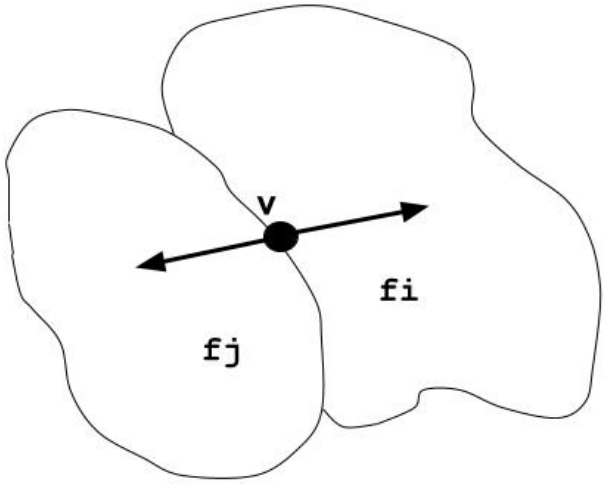


FIGURE 3.1: Depending on the likelihood test result, a movement along the normal at the point $\vec{v}$ is produced.

If no additional information is given, the statistical model chosen to describe the homogeneity of the region is the Gaussian distribution. So, for each region it is necessary to initialize:

$$\begin{aligned} \alpha &= (\mu, \sigma), \\ P(I_{(x,y)}|(\mu,\sigma)) &= \frac{1}{\sqrt{2\pi\sigma}} \exp\left\{ -\frac{(I_{(x,y)} - \mu)^2}{2\sigma^2} \right\}. \end{aligned} \tag{3.4}$$

This formulation leads us to the consideration that each point of the border depends on the $P(I_{(x,y)}|(\mu,\sigma))$; in other words, it leads us to evaluate the probability that $I(x,y)$ at the border can be generated by a Gaussian distribution $N(\mu,\sigma^2)$.

In practice, the range of action is extended to analyze a point, and therefore a small circular sample of the image is taken around each point of the border that is analyzed, denoting for S the variance of the radiometric values contained in this circular area.

Substituting the previous formula of distribution in the Equation (3.3) we will obtain the movement of a point $\vec{v}$ on the boundary $\Gamma_i \cap \Gamma_j$:

$$\begin{aligned}\frac{d\vec{v}}{dt} = &- \mu\kappa_{i(\vec{v})}\vec{n}_{i(\vec{v})} \\ &- \frac{1}{2}\left\{\log\frac{\sigma_i^2}{\sigma_j^2} + \left(\frac{(I-\mu_i)^2}{\sigma_i^2} - \frac{(I-\mu_j)^2}{\sigma_j^2}\right) + \left(\frac{S^2}{\sigma_i^2} - \frac{S^2}{\sigma_j^2}\right)\right\}\vec{n}_i.\end{aligned} \tag{3.5}$$

This equation governs the movement of the curve. When this movement stabilizes, the curve follows the border shared by two adjacent regions. The Fisher test is the similarity measure that is used.

3.6 From algorithm to application

We compared the polygonal representation considered in [54] to a B-spline representation of the snake. Given the structure of the polygonal, the discontinuity of curve derivatives is not introduced. Some results can be seen in Figure 3.2. The irregularity of the curve shape demands the replacement of the snake parameterization by B-splines as follows

$$Q(u) = \sum_{i=0}^{m} V_i B_i(u) \tag{3.6}$$

where V_i are the m control points and B_i are the blending functions [63]. Using this representation it is possible to introduce curvature constraint into the model to achieve smoother boundaries. In Figure 3.2 the boundaries are represented by polygonals and B-splines, and the difference between both representations can be appreciated. Boundaries recovered by B-splines, on the right, are smother than their polygonal representation, on the left.

One of the reasons to select the B-spline representation is its easy and compact way to represent the regularity of the shapes of the different fields. Another B-spline advantage is the fast computation of the spline derivatives, hence the internal forces, controlling curve curvature, can be introduced at a very low cost. The contours represented by B-splines are smoother and closer to the ones that can be drawn by a human operator, although oversmoothing can occur. To cope with it, we introduce a refinement step only used when the delivered result, at the first approach, does not follow all the boundary details. And it consists in increasing the number of spline control points.

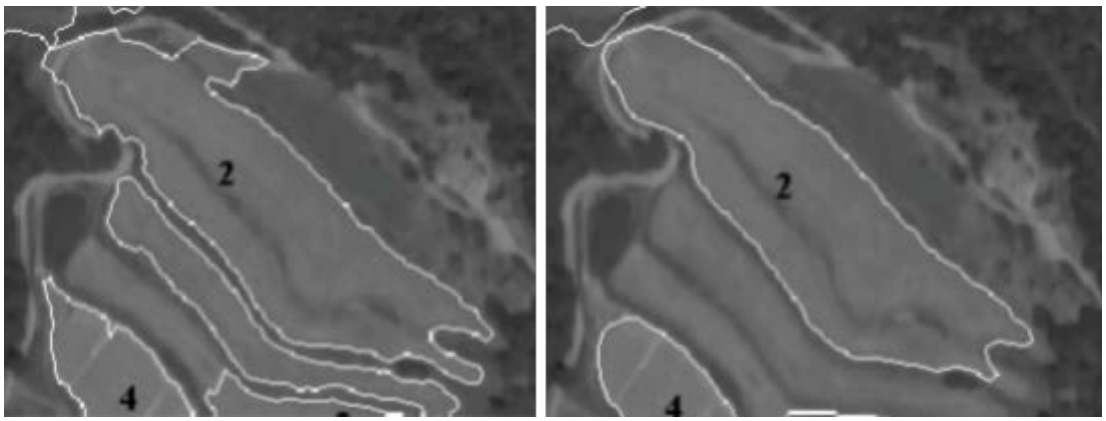


FIGURE 3.2: On the left the representation by a polygonal snake, on the right, B-spline representing a snake.

3.6.1 Convergence criteria

By studying the convergence of the process in the different cases of parameterization, we analize different strategies to stop the process when a solution is reached. The first approach computes the number of pixels inside the growing polygon. The deformation is stopped when this number becomes constant. A second approach is based on shape correlation between two followed iterations. It is applied to the case of representing the contour by a B-spline. This strategy is recommendable due to the fact that the shape correlation can be computed directly from the analytical B-spline representation. Shape correlation not only gives information to stop the process, but also detects balanced contour oscillations.

3.6.2 User interaction

Since the operator interaction is to give a point or a seed region from which the initial approximations are computed and to validate the result, his/her knowledge and experience assure "better" initial conditions. A seed, given by the operator, determines the initial snake as a small circular area that is also the first approximation of the field statistical parameters.

We study the dependency of the initial conditions and detect that in most of the cases it is very difficult to obtain a reliable homogeneity description of the area only by giving a point as proposed in [54]. Figure 3.3 shows how the mean varies considerably and the deviation is large enough to consider a small circular area a good initial approach.

In these cases it is necessary to deliver a region to take the sample of the probabilistic distribution used to describe the region. The operator defines an

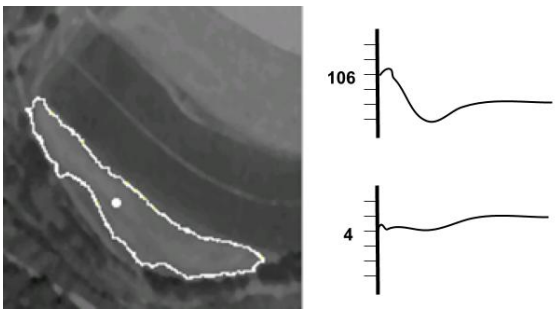


FIGURE 3.3: Point seed result and mean and deviation evolution.

area that stands for the region homogeneity. Also this region is the first approximation to the field boundary. With this approach a better process performance is achieved in terms of reducing the time needed to deliver a solution, because the stability of the distribution parameters comes faster and the system does not need to compute them any more.

The same field of Figure 3.3 with a seed area can be seen, labelled with the number 9, in Figure 3.5. There, the delivered seed is a larger area than just a point. So the extraction recovers the field boundary successfully. On the other hand, in Figure 3.3 some parts of the limits are not reached, due to the radiometric difference with the seed point.

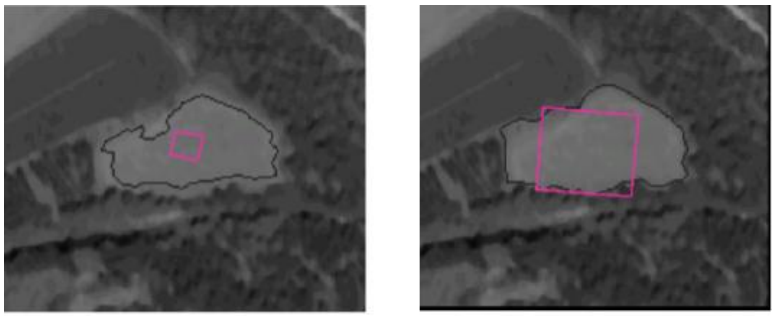


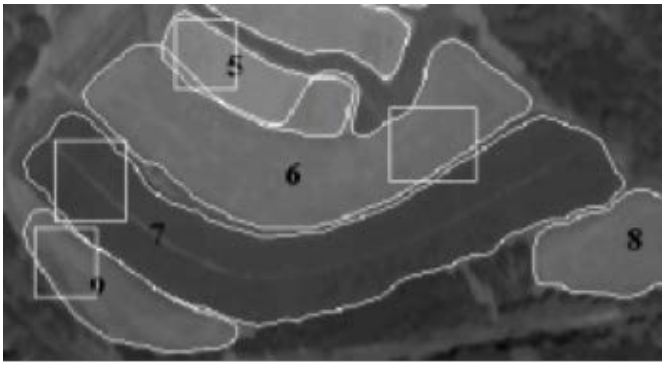
FIGURE 3.4: Seed selection effect.

The first seed in Figure 3.4 is small and can not reflect appropriately the grayscale gradation. The time needed to obtain the semi-correct solution is twice as much as in the right image, where a bigger sample gives a better approximation of the statistical parameters.

3.6.3 Validation

The algorithm has been tested on twenty different aerial images, with 30 fields on average in each. We have detected that 70% of the cases are fully recovered

and only a few of them need small edition tasks. In these cases the time reduction compared to the manual design is 60%. This fact justifies the use of the algorithm in a productive environment. Some results are shown in Figure 3.5.



(a)

(b)

FIGURE 3.5: Seeds and boundaries obtained 60% faster than manually drawn and with a sub-pixel accuracy.

3.7 When introducing edges

In [64] we introduced the information associated with edges, in the Equation (3.1) by defining:

$$E_{edge} = -|\nabla I(x,y)|^2$$

being I the input image. So, now the Equation (3.2) becomes:

$$E[\Gamma,\{\alpha_i\}] = \sum_{i=1}^{M}\left\{\frac{\mu}{2}\int_{\delta R_i} ds - \int\int_{R_i}\frac{1}{m}\int\int_{W(x,y)}\log P(I_{(u,v)}|\alpha_i)\,du\,dv\,dx\,dy + \gamma\right\} - \lambda\int\int_{R-\Gamma}\parallel \nabla I \parallel^2 dxdy. \quad (3.7)$$

Again and for any point $\vec{v} = (x, y)$ on the contour Γ, its position in each time step will be given by:

$$\frac{d\vec{v}}{dt} = -\frac{\delta E[\Gamma,\{\alpha_i\}]}{\delta\vec{v}}$$

that applied to the Equation (3.7) will define the movement equation for $\vec{v}$:

$$\frac{d\vec{v}}{dt} = -\sum_{k\in Q(\vec{v})}\left\{-\frac{\mu}{2}\kappa_{k(\vec{v})}\vec{n}_{k(\vec{v})} + \log P(I_{(\vec{v})}|\alpha_k)\vec{n}_{k(\vec{v})}\right\} + \lambda\vec{\nabla}|\nabla I\nabla I| \quad (3.8)$$

where $Q(\vec{v}) = \{k \,|\, \vec{v} \in \Gamma_k\}$ and $\kappa_{k(\vec{v})}$ is the curvature in Γ_k. In (3.8) the threshold to decide where to stop the movement of a contour point will be determined by local measurement of the contrast line (last term).

The movement is completed in the case of reaching an edge, or when the region growing process stops. So, Equation (3.8) weighs two possible non-excluding situations: the contours move to high gradients or enclose pixels with similar statistical parameters. Again, we choose the Gaussian distribution to describe the homogeneous appearance of fields. In our case these parameters are obtained from the clues delivered by human operators when they draw a starting polygon composed of at least 3 vertexes. It is the rough approximation of the field to be extracted. From this starting polygon the system will initialize the statistical parameters and the initial model, which will be deformed until the result is reached. Figure 3.6 shows an example of this approach.

In the final process step, the field contours are smoothed using the smoothing tool, delivered to the human operator in the toolbox of the application. The smoothing tool simplifies the contours and eliminates their oscillations, as is shown in the same Figure 3.6 on the right. On the left, field extraction example with starting polygon (black) and extraction result without smoothing (red) is shown.

FIGURE 3.6: On the left, field extraction example with starting polygon (black) and extraction result without smoothing (red). On the right, extracted field after smoothing.

3.8 Editing tools

The semi-automatic tool has been integrated into an editing menu, useful in case of necessity of small editing actions, like point modification or addition. With this utility it is also possible to draw from scratch the field boundaries when it is impossible to obtain them by automatic means (e.g., in case of lack of homogeneity, or fields detected by context information). The different actions that can be done with the editing menu are grouped into three sets.

- The first group includes the actions applied to one element, like moving, simplifying, and deleting.
- The second one contains the partial actions affecting one element such as modifying, inserting or eliminating vertices.
- The last set has commands that group several elements at once, for example union of boundaries. They are useful since the semi-automatic algorithm can over-split some areas that the user interpretation could have put together.

3.9 InJECT

The whole approach was included in a semi-automatic software package, inJECT [64], sharing other approaches to extract buildings and roads.

The complete system was delivered to the Geo-Information Office of the German Federal Armed Forces and it has been in practical use since May 2003

for the update of VMap Level 1 vector data and the generation of the military basic vector database. The extraction tools for area features are integrated into Inpho's software platform inJECT, which was originally designed for the measurement of 3D building models in digital imagery.

The semi-automatic extraction is preferably done in digital orthophotos for the capture of 2D GIS vector data. In addition, the software is available for the capture of 3D features using oriented aerial imagery. The algorithms and workflows have been extensively tested with IKONOS 2 and IRS satellite imagery as well as orthophotos with 50 cm pixel size.

Some examples are shown in Figure 3.7. As far as fields are concerned, it is remarkable to note the textured ones that have been extracted successfully. Moreover, there is another example with isolated trees inside, that have not stopped the growing process, and the linear B-spline has reached the field boundary. As well as the small hut inside has failed to mislead the growing process.

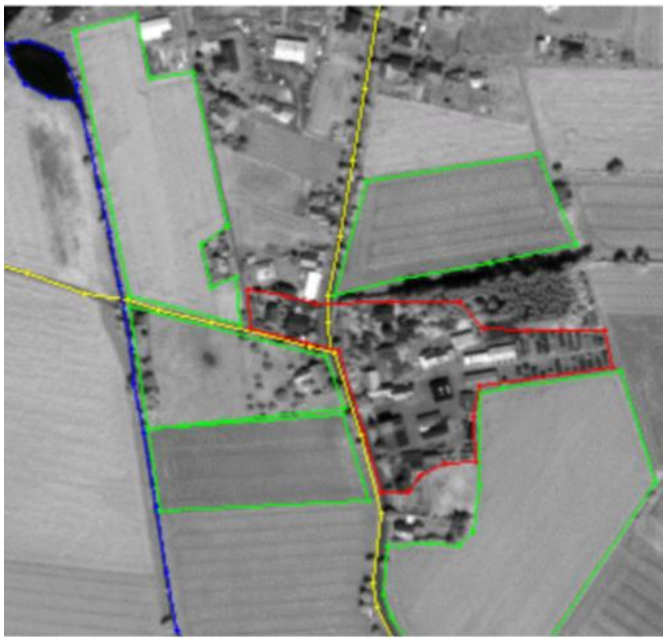


FIGURE 3.7: Extracted line and area features from an orthophoto. InJECT software

3.10 Results

In order to see if pre-processing the images affects the extraction of information or not, we consider two types of data sources: orthophotos and aerial photographs.

3.10.1 Field extraction in orthophotos

As has been described in Section 2.4, to obtain an orthophotomap several processes are carried out to radiometrically compensate adjacent photographs. Histograms are also equalized, so the resulting images have more pixels spread across the different spectrum values.

In addition, filters are passed to highlight linear elements and contrasts while maintaining the homogeneous character of similar areas. This is how it is achieved:

- to increase the differences between adjacent fields,
- in textured fields formed by trees and/or shrubs, to highlight these elements with respect to the background.

The former, in our algorithm causes a positive effect, but in the latter case, it can determine, and sometimes stop the growth of the seed. As for the scale of the input images, it has no greater relevance while distinguishing the elements that provide the contour. The extraction is only done in a channel, since it has been proved that the results are very similar to doing them in some other color space. In the Figure 3.8 some results are shown.

FIGURE 3.8: Color orthophotomap with the results of some fields obtained by semi-automatic extraction.

- Regions 2 and 4 although they are textured, have a certain radiometric homogeneity, therefore the algorithm picks them up well. In these cases, the algorithm abstracts from the edges that appear inside.

- We have tried to extract two much more textured regions and we see that although it provides a good approximation, there are some points where the growth process stops. In field number 3 some radiometric similarity is presented in a neighboring region, which causes the process to cross the border and include a small part of the adjacent region.

3.10.2 Field extraction in aerial photos

In this case we have not run any filters to improve the field appearance. However, the preprocess undergone by the frame is that it has been necessary to pass it from analog to digital: a cubic convolution. The result is a slightly less contrasted image than if we had a direct digitalization at the scanner resolution.

Some results comparing the semi-automatic approach with the boundaries obtained by manual digitalization can be appreciated in Figure 3.9.

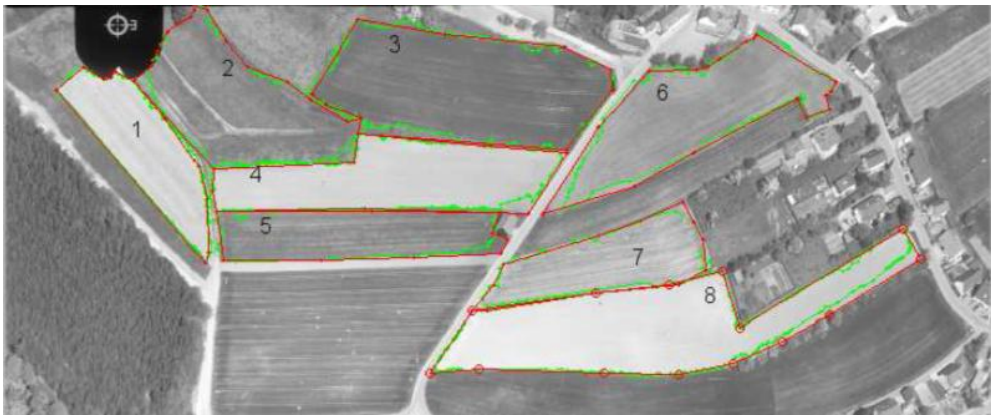


FIGURE 3.9: BW photography with semi-automatic extraction results (green) and their corresponding manual digitalization (red).

In Table 3.1 we show objective measures of comparison between results. So for the same field, the resulting areas are compared between semi-automatic and manual approach.

- The better results area are obtained by low-textured fields. On the other hand, fields number 4 and 8, although they are homogeneous, they have an illumination change near their boundaries. So, there the growing process stops.

- Textured regions, such as 3 and 6, reach successful segmentation in terms of areas, but the growing process does not reach some parts of their boundary, and in others the process crosses it. The most successful case

Field	Extracted area	Delineated area	Area ratio
1	6302	6311	0.9984
2	9650	10270	0.9396
3	14260	14628	0.9748
4	10137	10553	0.9606
5	7054	7094	0.9944
6	10046	10708	0.9382
7	6218	6706	0.9272
8	14474	15470	0.9356

TABLE 3.1: Statistical information associated with Figure 3.9.

is region number 2, where the internal edges have been absorbed by the growing process.

In Figure 3.10 we show areas that disturb the competition procedure between regions and that only the interpretation of the scene by an operator could undo the indefinition. The associated numerical values are shown in Table 3.2.

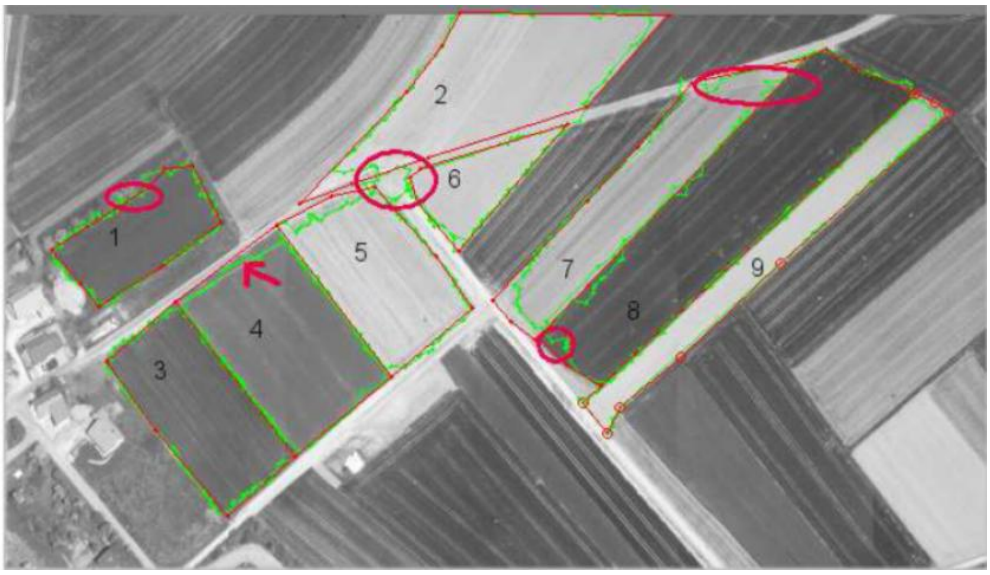


FIGURE 3.10: BW photography with the results of the semi-automatic extraction of some of its regions (green), their manual delineation (red) and some conflict zones (surrounded in red ellipses).

- Fields number 2, 5 and 6 have similar radiometric characteristics and they are separated by a road with similar radiometric values too. The conflict zone is defined by an ellipse. We see how the region number 2 grows through this area. These incorrect areas of growth remain reflected

Field	Extracted area	Delineated area	Area ratio
1	5143	4949	0.9623
2	15515	11819	0.7618
3	7935	7857	0.9902
4	9562	10503	0.9104
5	8161	8061	0.9877
6	3365	3436	0.9793
7	7980	9040	0.8827
8	16019	16308	0.9823
9	6307	6838	0.9223

TABLE 3.2: Statistical information associated with Figure 3.10.

by substantial differences between the values of area extracted and digitized.

The same happens in field number 7, which is delimited on the north by the same previous road, and expands across it in its growing process. It is a phenomenon of difficult solution if the problem is addressed from the radiometric point of view. Without the intervention of a human operator who edits the results, this problem could only be solved by adding context analysis.

- On the contrary, in regions number 5 and 6 the comparison is better, but it is only due to a problem of magnitude. They are small regions, in which the conflict zone would have a contribution smaller than the contour of the growing region. The growing process would produce a extreme curvature and, therefore, the growth that would generate this phenomenon is forbidden.

- In the case of region number 1, a human operator would not have delineated the trees that form the border, however the process does not detect a variation even in radiometry or in the presence of edges. We have circled the area that produces this irregular growth.

- Region number 4 has detected in its growth, a line –edge– which stops the process as if it were the border of the field. To a lesser extent this phenomenon has occurred in field number 5.

- Although textured, given its homogeneity, regions 3 and 8 provide a fairly real contour.

Field	Areas ratio	μ,σ seed	μ,σ complementary
1	0.995	230.38, 6.77	158.41, 34.47
2	0.960	137.66, 7.51	162.37, 38.27
3	0.996	196.44, 10.19	166.40, 39.76
4	0.948	125.23, 10.93	161.03, 40.87
5	0.956	210.08, 10.37	150.23, 34.87
6	0.869	155.27, 19.87	154.69, 41.29
7	0.987	213.04, 9.63	152.71, 39.78
8	0.940	203.90, 17.72	155.55, 41.29
9	0.941	206.96, 9.58	153.97, 40.50

TABLE 3.3: Statistical information associated with Figure 3.11.

Another example is shown in Figure 3.11, and its associated Table 3.3 where mean and deviation values appear, both of the field and its complementary region, which compete for the pixels' ownership.

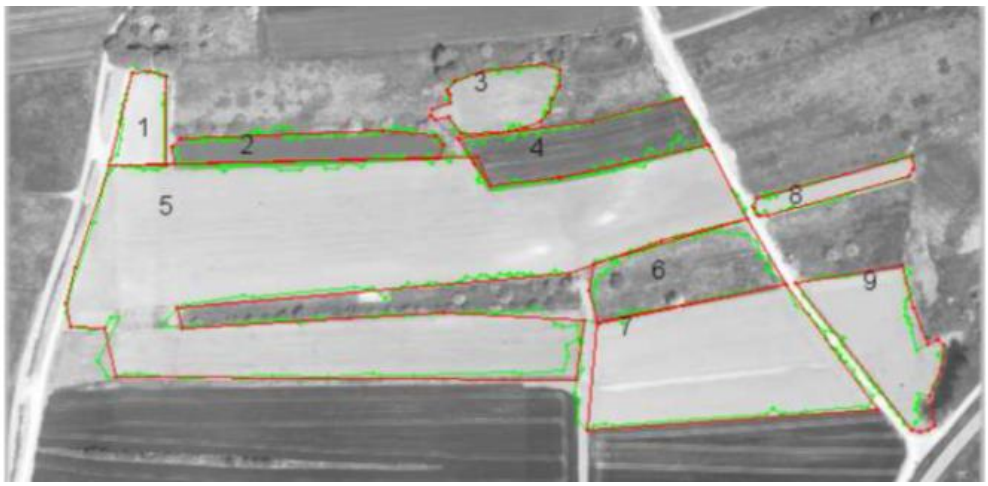


FIGURE 3.11: BW photography with semi-automatic extraction results and their corresponding manual digitalization.

- The fields whose extraction results are the worst are those whose deviation is larger. In particular fields number 6 and 8, whose deviation is larger than 15. Therefore when we look at the image, this numerical difference is increased by their small size. In fact, few are the contours points where the process has failed.
- Previous comments related to the textured field appearance can be applied to those fields with the maximum difference from the human delineation: 2, 4 y 6.
- Again we have examples in which lighting conditions cause the growing process to stop, and the contour can not span the region to its last pixels.

It happens in regions 5 and 9, although the difference in area is very small.

- Fields with irregular contours, but that can be distinguished from their environment, are successfully recovered. As is shown in fields number 1, 3 and 7, where their deviation is under 15.

In the last comparative example presented in Figure 3.12, apart from evaluating the same aspects as in the previous cases, we show there the associated edge-image to give an idea of how these elements affect the process.

FIGURE 3.12: BW photography with semi-automatic extraction results (green) and their corresponding manual digitalization (red).

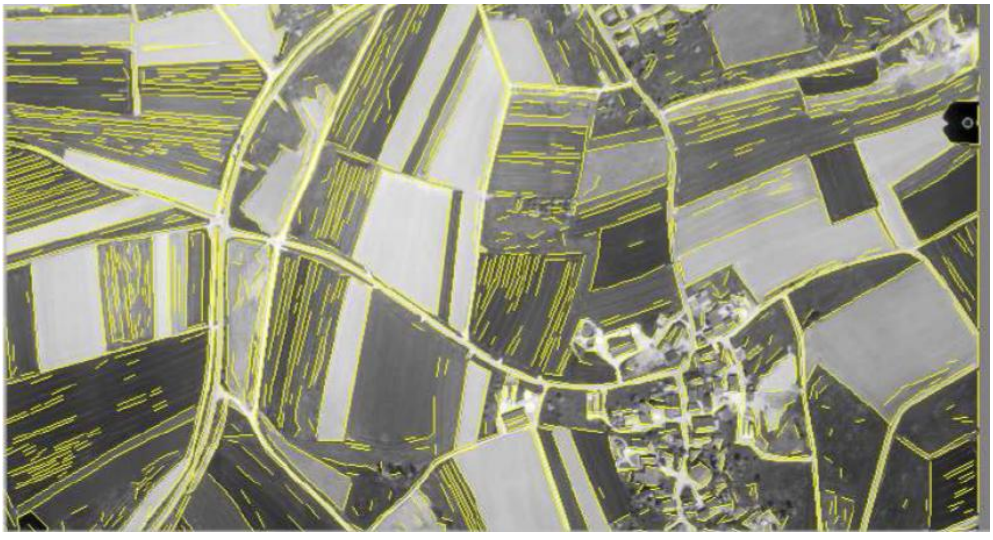


FIGURE 3.13: Edges associated with image of Figure 3.12.

Field	Ratio between areas	μ,σ seed	μ,σ complementary.
1	0.969	215.55, 7.20	143.84, 50.19
2	0.992	196.53, 12.64	142.86, 49.91
3	0.924	203.82, 19.45	147.10, 46.58
4	0.983	210.01, 15.98	141.76, 47.62
5	0.889	217.49, 7.99	140.74, 48.82
6	0.923	202.77, 7.44	148.22, 44.49
7	0.955	153.50, 7.48	141.86, 46.04
8	0.977	73.54, 8.10	138.87, 47.12
9	0.971	83.92, 11.92	130.93, 38.97
10	0.915	188.17, 19.74	128.25, 36.33
11	0.949	186.12, 9.77	132.02, 42.10
12	0.990	160.67, 8.66	126.28, 36.46
13	0.882	175.67, 17.95	136.03, 54.45
14	0.966	201.01, 21.05	138.06, 52.26
15	0.941	108.21, 11.69	137.59, 52.31
16	0.927	196.49, 21.03	137.50, 49.86

TABLE 3.4: Statistical information associated with Figure 3.12.

- Fields with sudden changes in radiometry in some area, as a result of the different state of growth of the crop or due to lighting conditions, are those with worse results: as happens with the field number 5. Its variability near the boundary is not recovered by the radiometry associated with the seed, which is around 8. The same effect happens to field number 13, where it is even worse due to the large internal edges, as can be seen in Figure 3.13.

  This argument is even reinforced in the fields number 11 and 12. There the edges are elongated enough to force the growth process to stop.

- On the other hand, in fields number 7, 15 and the left side of field 13, edges are not important enough to disturb the growing. What is more, the discontinuities between the edges are large enough to leave room for the growing process to escape through them.

- High seed variability in fields 10 and 14, together with the high variability of the complementary region make oscillations at the border.

- The strong edges of the border contribute very positively in fields of appearance like 3, 6, 8 and 11. As for field 3, they prevent the high variability of the region from affecting the stabilization of the growing process. On the contrary, the part of the boundary of field number 9 where the

growing process escapes is due to the edge discontinuity as can be seen in Figure 3.13 .

- In general, 80% of fields have good enough results. It will be only necessary to apply a simplification to obtain an acceptable result that can be similar to that obtained by human digitalization. This simplification process can be seen in Figure 3.6.

3.11 Conclusions

With this research we show the applicability of the region competition technique in a teledetection environment to obtain field boundaries. It changes the role of the operator from being the principal delineating "mean" to being a supervisor of the agricultural field segmentation.

It is very important that the operator is the one who controls the process and has a clear guide of what to do to obtain the best possible results. In addition, and since it is a semi-automatic tool, it was necessary to integrate it into an environment of editing tools.

Taking into account the edges as attractors of growth led to a faster stabilization of the growth process. On the other hand, this incorporation, when some contrast lines appear within the field, leads to obtaining over-segmentation.

This research has been published in two papers:

- Margarita Torre and Petia Radeva. Agricultural-field extraction on aerial images by region competition algorithm. In *Proceedings 15th International Conference on Pattern Recognition (ICPR)*, vol. 1, pp. 313-316, 2000.

- Timm Ohlhof, Eberhard Gulch, Hardo Muller, Christian Wiedemann and Margarita Torre. *Semi-automatic extraction of line and area features from aerial and satellite images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, 2004.

Chapter 4

Region competition. Semi-automatic road extraction

4.1 Introduction

In 2005, some of the previous attempts to automate photogrammetric steps were introduced to the orthophotomap generation. DSM to orthorectify aerial images was built by using automatic approaches delivered as commercial products, which also provide for the areas where control by an operator will be necessary and therefore possible editions.

Also in these years the blooming of GIS for the treatment of geographic information starts to replace the handling of vector files. In GIS environment, primary data capture, although important, is being replaced by the need to update existing information. One way is to locate those areas where changes are detected, so the work can be focused in some areas where sometimes the previous information will trigger or guide semi-automatic approaches.

At this time, roads and buildings were the topics that had been most investigated. And from the point of view of performance that new research can give, there are roads that seem to almost be trying to avoid human interaction. By this time, some approaches to assist in the delineation of roads had been studied. Most of them used seed points delivered by operators to ascertain the road centerline.

4.2 Our approach to road segmentation

Our purpose is to extract as many roads as possible independently of how wide and sinuous they are, from any kind of aerial image, without the flight scale constraint.

In view of the difficulties inherent in a fully automatic process when extracting geographic information [59], and following the good results that we had achieved with the semi-automatic approach used in field extraction [62], we have implemented a semi-automatic tool for the road segmentation.

In addition, several experiences described in ([65], [60], [66]) show that, due to the lack of robustness of automatic extraction tools, the best way to increase the productivity lies in designing semi-automatic tools. If an automatic process is prone to fail, the best approach is to let the human operator gain full control of the process, rather than make him/her search for the incorrect results on the whole image.

Moreover, the knowledge needed to make the system fully automatic for such a general purpose –several image scales, broad and narrow roads– could have been so complicated that it may make the system inefficient. For these reasons, we decided to wait for some information provided by an operator that guides the system at the beginning.

Since the margins are very important in order to recover the real visible extent of the road, we extract the complete road model by fitting two curves to the road margins.

The user provides the initial position of the feature to be extracted by giving some seed points along the road and editing the results interactively, whenever this is required.

4.3 State of the art

The semi-automatic methods may be grouped into two categories: path finders and path optimizers.

The authors in [67] introduced one of the most common path finder methods, which uses three initial points to define advancing direction and road width. Each new point P is searched for along the parabola which comes closest to the points last obtained, at the position that maximizes the homogeneity of the pixels in a mask whose middle point is P. The position of all the points is refined by means of edge analysis. With this approach the gaps –*cars and shadows*– cannot be ignored, and the process is obstructed by these elements.

The authors in [68] describe one of the successful path finder approaches, which differs from the previous one in the selection of the new advancing point. As is shown in Figure 4.1, the tentative next position and the road width are proposed. A position estimation is performed by profile matching with a

reference profile. Several profiles around the tentative point are extracted and weighted depending on the degree of correlation with the reference profile.

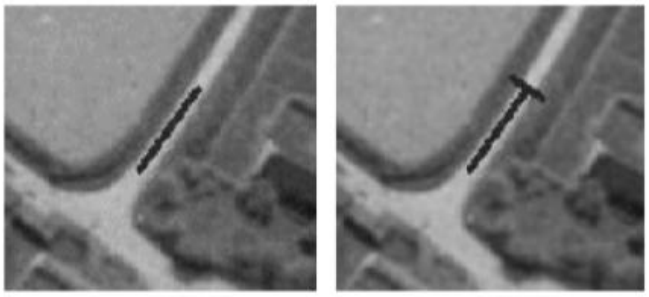


FIGURE 4.1: Initialization of a path finder. (Preprint [68])

The Kalman filter is used as a predictor-corrector method to combine the previous results obtained with the estimated positions. This method projects a prediction based on the previous results to the next position and corrects this prediction by using the weighted extractions referred to earlier. The process ends, either when evidence appears that the new point is not on the road centerline, or when the image boundaries are reached. It can be restarted by marking a new point on the road centerline. As may be observed, these approaches are digitization aids, due to the fact that they require user attention and interaction throughout the process.

The interaction is reduced when using a path optimizer method, which only requires user attention at the beginning –when giving seed points– and at the end –when reviewing the results. Snakes and dynamic programming are the most successful path optimizer methods. Both methods are based on solving a system of equations and inequations, but their difference lies in how the system is built. [65] outlines how the system is composed of equations that describe the suitable geometric and radiometric characteristics of the road. The equations involve the points that describe the output curve.

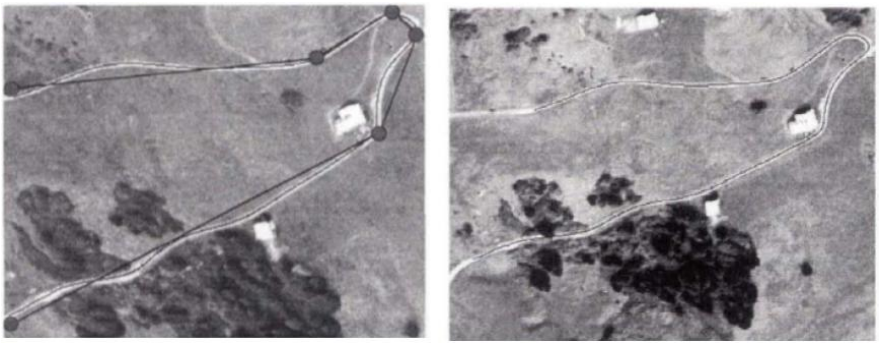


FIGURE 4.2: Initialization and results of road segmentation by a path optimizer. (Preprint [65])

Since the most convenient position of a point only depends on its neighbors, the system can be solved by dynamic programming. Although this approach generally produces correct results, some problems still remain related with occluded parts, and with the assumption that the radiometry of the road is nearly constant all along the road, because the seed points delivered by the user are strongly involved in the geometric constraints. One example is shown in Figure 4.2 where the initialization of the algorithm and its results, as an example of path optimizer, can be seen.

Some automatic algorithms give good results when a rough road delineation exists, and its purpose is to update the information. For example, [69] generates knowledge rules by using information about the kind of roads to be handled in order to update the road information in the Swiss official 1:25,000 database. In other cases making the process automatic comes from the restriction of having prior knowledge of the appearance of the roads to be extracted from the specific images used. For example, the definition of road width in the images involved, as can be seen in [70]. Whereas with this approach, the centerlines of roads in open rural areas can be extracted. Problems may arise in urban regions, where the system may be hindered by shadows and occluded parts.

On the left image of Figure 4.3, selected edges candidates for the roadsides are shown. With these lines, road segments are generated and represented by the points of the medial axes and attributed by the road width (on the right).

FIGURE 4.3: On the left, input data for the fusion. On the right, hypotheses for road segments. (Preprint [70])

4.4 Semi-automatic proposal for extracting roads

Our aim is to represent a portion of a continuous grayscale image, on which roads appear, by a set of vectors that follow the roadsides. We will start from a set of generic models that initially approach the road boundaries using region growing techniques, and later these approximations will be deformed by region competition to refine the extraction of roadsides.

With respect to radiometry, we use this approach to segment the roads that are characterized by smooth variations in homogeneity and outlined by other elements with different homogeneity values. On the other hand, due to their design constraints, the majority of the roads can be split into parts where the curvature is subject to small variations.

Since in our case these radiometric and geometric characteristics appear together, it is feasible to apply a combination of deformable models guided by region competition criteria. This method is used to refine a first road approximation, obtained from a very simple model that is deformed by parts according to the information delivered by the image, by following region growing techniques. Since the region competition is applied over points located on image boundaries, the selection of the regions that compete for roadside points is easier, and their statistical parameters can be distinguished. Region growing makes the first steps faster, and region competition delivers more accurate results. Our proposal is a generalization of region growing techniques [53] combined with deformable models [27]. This mixed approach is called region competition [54] and has yielded good results when segmenting agricultural fields [62], especially when the regions handled are quite homogeneous and their homogeneity is sufficiently different from that of their surroundings. The aim of this approach is to alleviate the tasks of digitizing the centerlines and roadsides. It is also possible to recover roadsides in the event that the road changes its width in the image.

The application of the combination (deformable models and region competition) is appropriate, given that the majority of the roads have two important characteristics: in a **small** portion of road the radiometry and curvature changes are **small**.

Deformable models, such as snakes, are defined as elastic curves that dynamically adapt a vector contour to a region of interest, by applying energy minimization techniques. For the road extraction, we use region growing techniques to adjust an initial model to the sides of the roads that can be extracted

from the image. The initial model, hereinafter referred to as a *tube*, is a centerline with a parallel copy on each side.

From the seed points, the centerline is built by drawing a B-spline that comes closest to these, adding more points between the seeds when it is necessary. To obtain the roadsides, it is necessary to make *a specialized parallel copy* applied to the centerline. We have tested two approaches, both of which are based on region growing algorithms. The growing process, from an initial starting point on the centerline, determines the statistical parameters of a small region around this point, and adds or refuses pixels on the growing path, depending on whether or not they fit the previously determined statistical model.

The tube is modelled by two B-splines, one for each roadside, and through its construction it becomes easier to preserve smoothness, the main geometric characteristic of the road. This result can be considered as a road symbolization, because in some parts the model does not exactly follow the information that appears on the image. At the model refinement stage, we introduce the region competition algorithm. It is applied to the points used to draw *the parallel copies* and, where necessary, to the points added to fulfill the output requirements.

The deformed model obtained may have undesirable effects, such as loops and extreme concavities. These problems are corrected at the final stage, when a minimum output description is built to draw the resulting element.

4.5 Adaptive contour models

In the correspondence between image features and linear representation, there are two types of models: rigid and deformable.

In rigid models the shape and geometry of the object to be extracted is known, and with the image data its parameters are determined.

Deformable models are specified with much more general restrictions, such as tolerances in curvatures, degree of required continuity, which are characteristics located in the images. They are models that facilitate the integration of the two restrictions that we must take into account in our research: geometric and radiometric. Snakes must be understood as a special case of the more general technique of matching a set of deformable models with an image through minimization of an objective function that in this case takes the form of energy:

$$E_{snake} = \int_0^1 E_{snake}(\mathbf{v}(s))ds.$$

The energy function that is minimized is a weighted combination of internal and external forces. As is shown in formula (3.1). The internal force emanates from the shape of the snake, while the external one comes from the image.

An example of internal forces are those which are imposed on the curve that you want to obtain, for example restrictions of continuity and softness. External forces deform the snake to follow features of interest of the image or structuring characteristics of it.

The traditional example of internal energy is represented as:

$$E_{int} = \alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2$$

where the first term of the sum is called elasticity parameter and causes the snake to act as a elastic spring between the different points that define the surface. The second term is known as parameter of rigidity and causes the snake to behave like a sheet.

An example of the external forces that deform the snake towards characteristics of interest of the image would be the monitoring of elements of high contrast or linear elements. Therefore, the analysis of the gradient image would be considered.

The solution is found when snake minimizes the objective function, although it is possible that this function may have more than a local minimum. As the process stops when a minimum is found, it may not find the optimal solution. To improve stability and convergence, as well as to be able to incorporate rapid calculations of softness and curvature, the implementation is based on a parametric approach B-spline of the curve. The resulting model is named B-Snake.

The snakes model needs to be given an initial approximation of where the contours searched for are located. For example, if the initial curve is not close enough to contour, it will not attract it, since the curve will not be subject to an external force that counteracts internal energy and, as in most cases the internal energy premiums the simplicity, the curve will shrink to a point.

Therefore an alternative would be the introduction of a **balloon** that inflates the initial curve until it finds some clues of contour. With the introduction of

this **globe** model we provide snake with a good initial approach. An example would be the proposed by [71]:

$$F = k_1\vec{\mathbf{n}}(s) - k\frac{\nabla I(\mathbf{v}(s))}{|\nabla I(\mathbf{v}(s))|}$$

The coefficients k and k_1 are chosen so that they are the same order, but k is slightly greater than k_1, so that a point of contour can stop the force of the **globe**. As we see, it is a force that pushes the initial curve along the normal, until it finds contour evidence. So that the influence of addition does not cause loss of precision in the location of the contour, when evidence of it has already been reached, this energy component lowers the growth rate and the contour curve is better adjusted.

There are some variations of the snakes model that allow broadening the spectrum of action. They are described below:

- **Ribbon snakes**

  When instead of the axis of the road we want to extract its margins, the natural extension of the model is to provide the width curve at each point. This is called **ribbon snake**, leaving the representation of the curve as $v(s,t) = (x(s,t), y(s,t), w(s,t))$, see [28] for more details.

  In this approach, the only change in the **internal energy** is that the same tension and stiffness restrictions are also applied to the width. In contrast, and unlike the original **external energy** concept, image forces are applied to the sides of the snake. Therefore, on a road that has contrast to the background, the functional that represents the external energy must be redefined as the sum of the magnitudes of the gradient image along the curves that would define both margins. The way this alteration is proposed is by projecting the gradient image over the normal to the ribbon snake and restricting the projection to be positive on the left side and negative on the right, leaving the formula:

$$E_{ext}(v(s,t)) = (\vec{\nabla} I(v_L(s,t)) - \vec{\nabla} I(v_R(s,t))) \cdot \vec{n}(s,t).$$

- **Zipplock snakes**

  Even with the **ribbon snakes** approach, one of the well-known limitations of deformable models is evident: their sensitivity to initialization. Thus, obstacles between the initial position and the desired one attract

the extraction and make it wrong. To try to solve this problem [72] proposes an alternative that consists of dividing the snake, or the ribbon snake, into active and passive parts.

During the optimization process, the external forces of the image propagate from the ends towards the center of the snake. During this optimization the external forces are only applied to the active points, while the passive part is only optimized with respect to the internal energy. This is the only the geometric constraint that the curves have to follow.

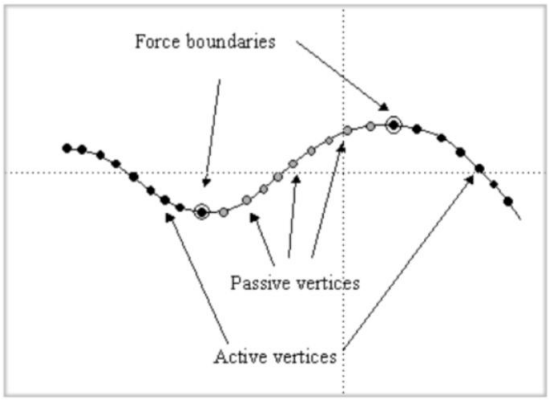


FIGURE 4.4: Relevant points in Zipplock snakes

.

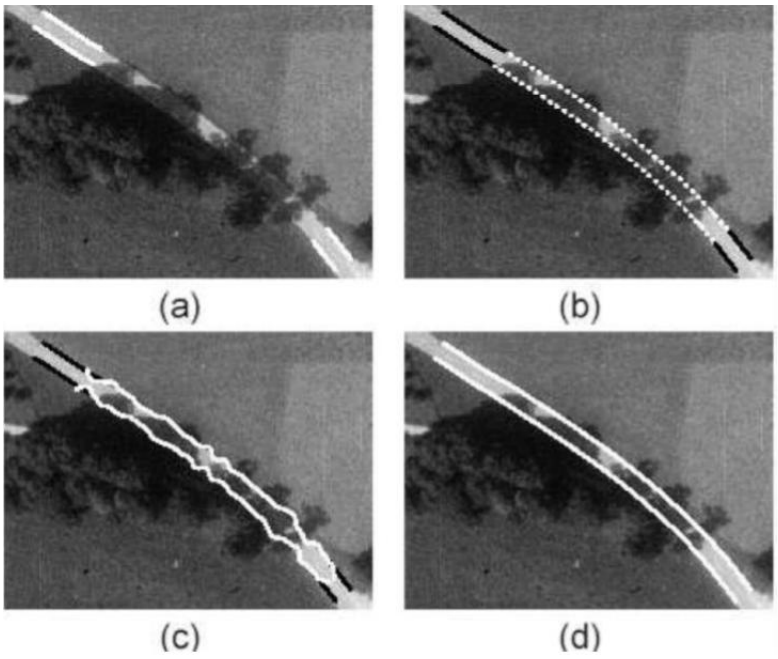


FIGURE 4.5: Four steps in the extraction of a road with few verification zones.

Therefore, given a correct initialization of the extreme points, this strategy ensures that the active parts are always close to their desired positions, while irrelevant image structures around the passive parts have no influence on the extraction.

Thus, in each step we try to bring the extreme points closer to each other, so that in each iteration the optimal solution is sought, only incorporating the external energy to the extreme points. Once these points are stabilized, because in the new step there is no movement of the treated points, then the application of external forces to each of the extreme points is extended. The types of points and the results after the three-step process are shown in Figures 4.4 and 4.5.

- **Eigensnakes**

  In the traditional approach of snakes, they are attracted to image features. In many cases it is very difficult to parameterize in an analytical model the characteristics of the elements to be extracted. In the case of the Eigensnakes the external energy of the snake is defined as a function of the Mahalanobis distance of the characteristics of the images analyzed at the center of the cluster of learning, as can be seen in [73]. With this approach, the model of the elements to be extracted is given more and more variability.

4.6 Deformable models for roadsides

In this case we have not reduced the type of roads to be extracted. We try to extract as many roads as they appear in the image. The considerations about how to represent the knowledge that we have about the image appearance of the roads are similar to the ones explained in Section 3.5 when talking about fields.

A region R is considered homogeneous if the intensity values are consistent with its generation by a distribution family of a pre-specified probability $\mathcal{P}(I : \alpha)$, where α are the distribution parameters.

We continue with a function to represent (by MDL) the portion of the image where the road is located. We will note this as Γ.

$$E[\Gamma, \{\alpha\}] = \frac{\mu}{2} \int_{\partial R} ds - \sum_{i=1}^{M} \int \int_{R_i} \log \mathcal{P}(I_{(x,y)} : \alpha_i)\, dx\, dy. \tag{4.1}$$

This expression reflects the energy associated with the snake curve that will represent the roadsides. The first term is the interior energy associated with the curve, and forces it to be the shortest possible. The parameter μ is the code length per unit arc length and ∂R is the boundary of the region R. This region is enclosed between the roadsides and represents the road. The second term is the exterior energy along the curve, due to the image radiometry that inversely decreases with the degree of similarity of the intensity value at (x, y) $(I(x, y))$ to a homogeneous region described by $\mathcal{P}(I : \alpha)$. The regions considered in this term are the ones that surround the road and the road itself (M is the number of all the regions involved in the process).

The additional information that we have about the radiometry of the road can be included in the statistical model. On the aerial photographic images the roads appear to be fairly homogeneous and bright. This means that if a point on a road is considered in its immediate environment, there will not be significant radiometry changes. This leads us to assume that its radiometry can be described by a Gaussian distribution, parameterized by α, which expresses the mean and the variance.

The brighter appearance of the center of the road, which decreases when moving to its sides, could help the growing process from the centerline to the roadsides, but will not be used at the region competition stage.

The minimization of this energy gives the contour positions at each time step. The contour is parameterized and denoted by v. The solution is reached by the steepest descendent method:

$$\frac{d\vec{v}}{dt} = \frac{\mu}{2}\kappa(\vec{v})\vec{\mathbf{n}}(\vec{v}) - \sum_{k \in \mathcal{Q}(\vec{v})} \log \mathcal{P}(I_{(\vec{v})} : \alpha_k)\, \vec{\mathbf{n}}(\vec{v}), \tag{4.2}$$

where κ is the curvature of the contour, $\mathcal{Q}(\vec{v})$ is the set of regions that for a given point at $\vec{v}$ will compete for it and $\vec{\mathbf{n}}(\vec{v})$ is the normal unit vector to the contour v at a given point $\vec{v}$.

In our case, for each point P at the deforming contour, we handle two regions: the road region R and the one outside it R_j. For each P, a small circle, centered on it, is drawn. The outside region is a circle containing a portion of the image that does not intersect the road deforming region, the center of which lies on the normal unit vector to ∂R at P. The location and extent of all these elements that take part in the region competition algorithm can be seen in Figure 4.6. As observed before, since region competition is applied to the first model that reached the roadsides at some points, the selection of these circles can be made over regions with statistical parameters that can be distinguished

from one another.

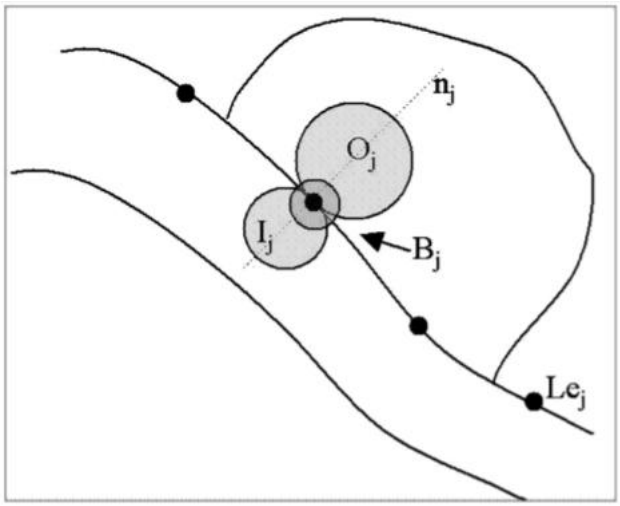


FIGURE 4.6: Position of the different circles considered in the region competition algorithm for a point at the roadside.

Therefore, for the portion of the contour headed by P we have the equation:

$$\frac{d\vec{v}}{dt} = -\Big[\mu\kappa(\vec{v}) - \Big(\log \mathcal{P}(I_{(\vec{v})} : \alpha) - \log \mathcal{P}(I_{(\vec{v})} : \alpha_j)\Big)\Big]\vec{\mathbf{n}}(\vec{v}). \tag{4.3}$$

Formula (4.3) represents the evolution of a road boundary driven by keeping the curvature low, and modifies the contour depending on the similarity to the intensity distribution of one region or the other. When replacing the probability by Gaussian, we obtain the following region competition formula:

$$\frac{d\vec{v}}{dt} = -\left[\mu\kappa(\vec{v}) + \frac{1}{2}\left\{\log\frac{\sigma^2}{\sigma_j^2} + \left(\frac{(I-\mu)^2}{\sigma^2} - \frac{(I-\mu_j)^2}{\sigma_j^2}\right) + \left(\frac{S^2}{\sigma^2} - \frac{S^2}{\sigma_j^2}\right)\right\}\right]\vec{\mathbf{n}}(\vec{v}). \tag{4.4}$$

The pixels are then tested in order to decide which of the two regions they belong to: for each point P of the contour a small circle is taken and its statistical parameters I, S (mean and standard deviation) are computed. The parameters of the road region are μ, σ (mean and standard deviation), and μ_j, σ_j corresponding to the competing region at point P. The location and the extent of the regions involved can be seen in detail in Section 4.7.3.

The likelihood ratio is then computed to decide whether the region parameters associated with P fit better into the distribution describing region j or the road. Thus two adjacent regions are competing for pixel ownership

(region competition). The new position of the contour for P will be obtained by moving along its normal unit vector at P, determined by formula 4.4, in order to take P into R or to leave it to the outside region R_j, depending on which one is more similar under the curvature constraints.

4.7 From algorithm to application

In short, the algorithm can be divided into the following successive steps:

1. Seed points are manually placed at significant radiometric and geometric road sections.
2. Model building:
   (a) Seed densification, by taking radiometry of closer seeds and geometric constraints into consideration.
   (b) Centerline generation by B-splines, using all the aforementioned points as control points.
   (c) Roadside generation by specialized parallel copy, taking account of radiometry characteristics, or building the convex hull of growing circles around the control points.
3. Refinement:
   (a) Roadside densification and adjustment by region competition.
   (b) Model refinement and, if requested, output simplification.

4.7.1 Initialization

The starting model is a centerline and its parallel copies to each roadside. The parameterization chosen for each one of these copies has been the approximation by B-splines, which is represented by $\sum_{i=1}^{m} V_i \mathcal{B}_i(u)$,
where V_i are the control points and $\mathcal{B}_i$ are the blending functions [63].

One of the reasons for selecting the B-spline representation is the easy and compact way in which it represents the regularity of the road shapes. Another advantage of the B-spline is the fast computation of its derivatives, and thus the internal forces controlling curvature can be introduced at a very low cost. The contours represented by B-splines are smoother and closer to the ones

drawn by an operator, although over-smoothing may be obtained; this undesired effect will be solved at the end of the iterative process that delivers the solution.

The user interaction starts by giving some initial points seeds along the road to be extracted. These points must be placed at:

- locations where the radiometry is representative of a portion of the road,
- sites with large curvature changes.

The first condition helps to initialize the statistical parameters. The second one reduces the number of points to be added between seeds, and improves the results. The statistical parameters that will drive the growing process will be computed from a circle around each seed, and the first model will be drawn as a tube that contains these circles.

From these seeds, further points are added in order to make the initialization more robust. This process is based on two conditions:

- The maximum distance allowed between initial points. This is a parameter δ delivered by the user and it depends on the output resolution desired.
- The curvature changes between the position of the seeds. For every triplet of points delivered by the user, the system computes its angle. If it is smaller than the parameter θ delivered by the user, the process will add points.

When the distance between two seed points P and Q is larger than δ or the angle that is formed with the following seed point is smaller than θ, the process places as many initial points S_i between P and Q as is necessary to avoid this situation.

$C(\lambda)$ will represent the parameterization of the curve that is under construction, $\lambda^{'}$ and $\lambda^{''}$ the parameters for P and Q, so $P = C(\lambda^{'})$ and $Q = C(\lambda^{''})$ and finally λ_i, $\lambda^{'} < \lambda_i < \lambda^{''}$, represents the parameter value for which the process needs an initial point. The search will be restricted to a region around the B-spline guide[1] and along its normal unit vector. The point selected S_i must minimize the equation:

$$E(i) = dR(S_{i-1}, S_i) + PdR(P, Q, S_i) + Curv(P, S_{i-1}, S_i, Q) \tag{4.5}$$

[1] B-spline guide is the curve that approaches the delivered seed points and the initial ones that have been found so far.

where:

- $dR(A, B)$ is the radiometric distance between the statistical parameters associated with A and B. This means that for each point a small circle centered on it will be used to compute the median. The radiometric distance will be the absolute value of the difference between the medians.
- $PdR(A, B, P)$ also means a radiometric distance from A to P and from B to P, but in this case weighted by the Euclidean distance.
- $Curv(A, B, P, D)$ is the curvature value at the point P of the B-spline whose control points are A, B, P and D. By differential geometry it is known that the vector $\frac{C'(\lambda) \times C''(\lambda)}{|C'(\lambda)|^3}$ has the same magnitude as the curvature. The expressions of the curve derivatives when the curve is parameterized by B-splines are really compact and fast in computation, as can be seen in [63], page 396.

This equation takes account of the previous initial point found S_{i-1}. Note that when $i = 1$ the first term disappears and at the third term only three points are considered.

4.7.2 Building the model

When the initial centerline is obtained, it is time to look for roadsides. Two approaches have been used, and in both cases the results are practically the same.

- Parallel copy of the B-spline that represents the centerline.
- Convex hull of the circles that grow around each initial point.

Parallel copy of the centerline

From the centerline $C(\lambda)$ drawn from the seed and the initial points S_i, $i = 1, ..., n$, its specialized parallel copy is built using the tangent t_i and the normal n_i lines to $C(\lambda)$ at each S_i. It is also necessary to introduce a small circle B_i around these points.

An iterative process moves two copies of t_i along n_i –one in each direction. We will refer to the corresponding intersection between the moving copies and n_i as $\mathcal{L}_i^j$ and $\mathcal{R}_i^j$, where j indicates the iteration number. The process ends when the radiometry of a small area around the point $\mathcal{L}_i^j$ is not inside the statistical model defined at B_i. The same reasoning is applied to $\mathcal{R}_i^j$.

Convex hull of growing circles

With the second approach, for each S_i the algorithm makes the circle B_i grow. This ends when the radiometry of B_i^j is different from that obtained at B_i^{j-1}. The tolerance level in this case is small and we use the characteristic of descending lightness when moving from the center to the roadsides.

We will refer to the corresponding intersections of the resulting B_i with n_i as $\mathcal{L}_i$ and $\mathcal{R}_i$.

In both cases the parallel copies will be generated by B-splines, using as knots the points $\mathcal{L}_i$, $i = 1, ..., n$ for the left side and $\mathcal{R}_i$, $i = 1, ..., n$ for the other side.

4.7.3 Refining the model

Once the model has been generated there is a post-process to analyze its correctness and completeness. During its construction there are some defects that may appear, as confirmed by our experience:

- missing parts of the roadsides,
- parts in the model that are not allowed: loops and extreme curvatures, and
- redundant points.

We will proceed to describe how we deal with each of these failures.

Missing parts of the roadsides

Once the B-spline representing the roadside has been generated, the process analyzes its points to check its reliability. The aforementioned parameter δ – the maximum distance allowed between consecutive seed points– will also be used at this point.

We will present the reasoning for $\mathcal{L}_i$ points, and the same will be applied for $\mathcal{R}_i$ points, simply by considering the opposite direction. So for each couple of points at the roadside, $\mathcal{L}_i$ and $\mathcal{L}_{i+i}$, the system will generate the additional initial points Q_k, $k = 1, ..., m_i$, as described in Section 4.7.1. Later, the system will check for the reliability of each of the margin points $\mathcal{L}_j$ and Q_j.

This test of reliability and, when necessary, the refinement of the location of points, is performed by region competition, adapted to our purposes, as described in Section 4.6. To do so, a small circle, B_j, is drawn centered on the

point to be tested. Two disjoint circles, I_j and O_j, are generated from B_j. The three circles have their center on the normal unit vector n_j and their position is as described in Figure 4.6. Both I_j and O_j overlap B_j, and both compete to own B_j. The region competition method will define the refined position of the margin point.

Shapes not allowed

In the model that we have chosen to describe the majority of the roads, there are some effects that cannot be allowed, either when generating or refining the model: loops and extreme curvatures. A loop in a contour, which is parameterized by λ and denoted by C, is a closed portion of it. In other words, we detect the presence of a loop at the point P on the contour v, because it can be reached at two different stages of parameterization: $P = C(\lambda_1) = C(\lambda_2)$, where $\lambda_1 \neq \lambda_2$. We will call P the origin of the loop.

When detecting a loop, by locating its origin, this point splits the contour into two parts. At least one of these is closed. By constructing the roadsides, it is possible to find nested loops, as can be seen in Figure 4.7. Therefore, the loop analysis is based on moving along the curve by increasing its parameter λ to determine whether the loop is correct or not, once all the segments have been visited. A loop is valid when all its segments are far enough from the centerline. For example, in Figure 4.7 the loop Rb_1 is invalid and Rb_2 is valid. In this case, the loop analysis order will be Rb_4, Rb_3, Rb_5, Rb_2, ending with Rb_1. Since this analysis is done at each iteration, the loop is eliminated as soon as it is detected, without altering the rest of the main contour.

If it is found that a couple of points, P and Q, which are the closest in location, are not consecutive in parameterization, we compute curvature at these points. If it is above the tolerance level, we face the problem of extreme shapes which are not allowed. Being λ_1 and λ_2, the parameterization of P and Q, the contour $C(\lambda)$, is split into two parts:

$$\begin{aligned} C_1(\lambda) &= \{C(\lambda), 0 \leq \lambda \leq \lambda_1,\ \lambda_2 \leq \lambda \leq 1\}, \\ C_2(\lambda) &= \{C(\lambda),\ \lambda_1 < \lambda < \lambda_2\}. \end{aligned} \tag{4.6}$$

The system eliminates $C_2(\lambda)$ and reparameterizes $C_1(\lambda)$ so that it becomes a continuous curve, P being the previous point to Q. The resulting curve is the input for the next iteration. Again, since the analysis is done constantly, the elimination acts locally at the problematic part as soon as it appears.

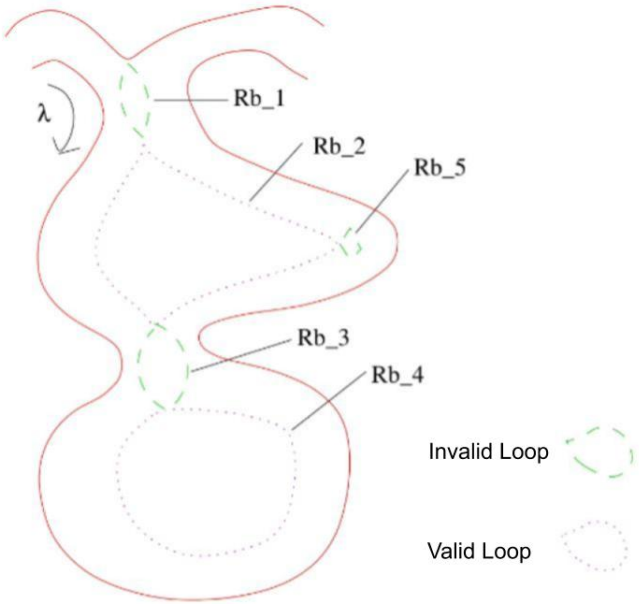


FIGURE 4.7: Valid and invalid nested loops.

Redundant points

The process analyzes the resulting B-spline using pieces of a given length named Λ. We will refer to the ones used to generate the B-spline as original points. The global curvature is computed for each piece limited by two points –initial and last, which we call starting points– taking the angle defined by the initial, the virtual middle point and the last one. An analysis is then made of how far the original points are from the curve defined by these starting points. Then, if necessary, the most important original point –the one whose absence causes most discrepancy– is added to the starting points. This is the kernel of the iterative process, which finishes when non-significant original points are ignored.

4.7.4 User interaction

On the basis of our past experience, a semi-automatic tool was chosen because:

- it provides a better approach to the initial position of the starting curve;
- it increases the reliability of the initial statistical parameters, describing the road by radiometric parts where differences can be found;
- it makes it possible to detect the problems related to automation by validating the results.

Apart from accepting/refusing the results, since the operator interaction is to give seed points from which the initial approximations are computed, the knowledge as well as the experience of the operators assure "better" initial conditions.

The first B-spline following the centerline is drawn by using the points provided by the user. Later, a small circle is defined around each seed point. The pixels that this circle encloses are used to compute the statistical parameters that will describe the practically homogeneous radiometric values region by region.

Therefore, it is important that the seed points are delivered in regions where changes in radiometry and in curvature occur.

As mentioned before, past experiences with other photogrammetric tasks have led us to design and implement a semi-automatic approach. Any automatic tool that does not assure the correctness of the results in almost all cases and whose errors are not automatically labelled, obliges the operator to make an exhaustive check of all the information. This makes automatic tools hard to use, and the time spent on quality control could cancel out the advantage of automation which is aimed at.

There is also another type of interaction: the input parameters. There are three necessary parameters:

- the maximum distance allowed between initial points: δ,
- the minimum angle defined by two adjacent road segments: θ, and
- the length to analyze redundant points: Λ.

The first and the second parameters depend on the output resolution desired. Λ is also related to output resolution, but since the redundance analysis is done without considering the image data, its effects are more important with regard to the vector appearance.

4.7.5 Specific guide points in branches

In cases where the system has to make a choice, such as crossroads, the seeds and the geometric constraints will deliver the solution that follows the given points by keeping the curvature under the threshold permitted. One example is shown in Figure 4.8, where there is a sequence of interactions and results in the case of one crossroad. The first image shows the seeds that will guide the extraction. The second shows the solution obtained, as well as the seed

points for the other branch. The final image shows the segmentation of both branches, without there being a need for any edition.

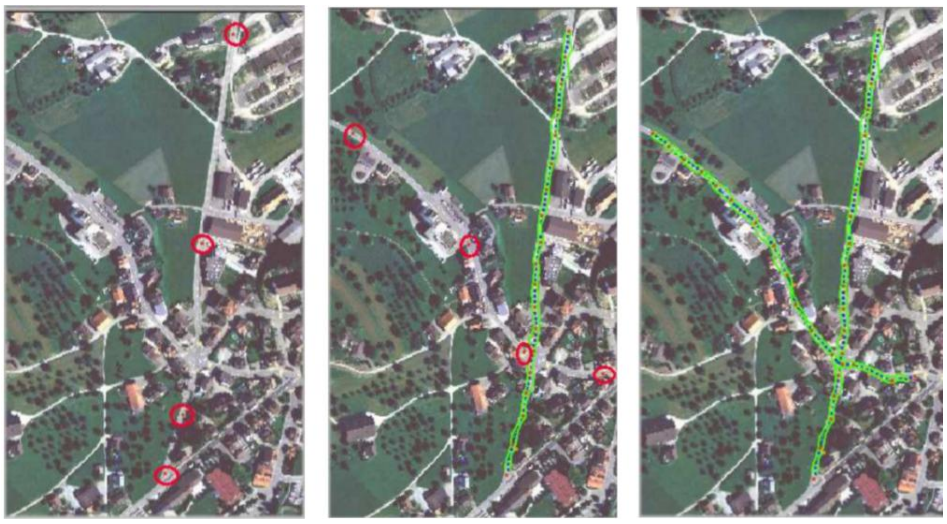


FIGURE 4.8: Image sequence of actions for dealing with a crossroad.

4.8 Results

We have chosen three images to show in detail the behavior of the proposed approach, and how it handles some particular road appearances.

At the end of this section we will present some more general results and quantitative evaluations over a set of assorted images. We will finish with a comparison with one of the most relevant semi-automatic approaches.

In the first image a sinuous but very homogeneous road is shown. Therefore, the seed starting points are placed at sites where there is high curvature. In Figure 4.9 the seed points appear in red and they are bigger in size than the rest of the points. The output roadsides are shown in green.

As can be seen in the image, the road is not constant in width, so the region competition algorithm, applied to fit the road-model to the image information, has extracted the road appearance from the image. This effect can be clearly seen at the road portions labelled A and B. This is possible due to the fact that the specific parallel copy is done in segments, and the result is adjusted to the margins using region competition.

In the second example presented in Figure 4.10, the roads may be confused with their surrounding fields. In some parts the appearance of the road is not lighter than its environment, and the radiometric difference between the road and its neighboring areas is insignificant. Under these conditions the

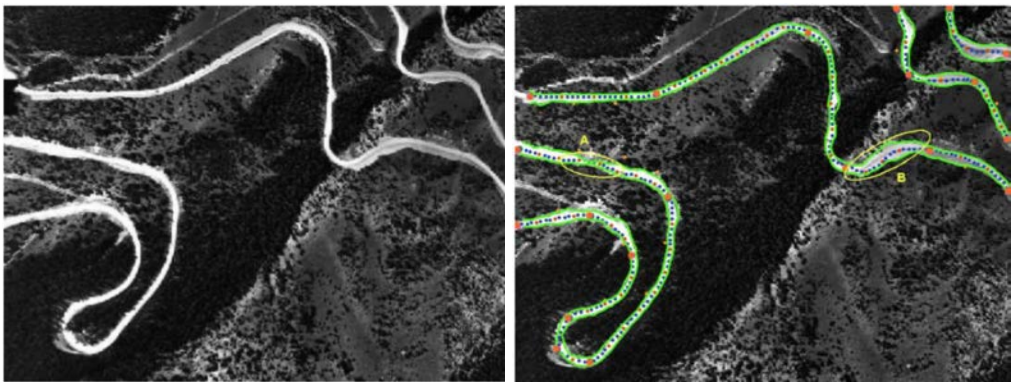


FIGURE 4.9: Original image and semi-automatic road extraction results obtained from the algorithm in the case of a sinuous and homogeneous road.

geometry constraints will be of more importance than the radiometry ones. The results are good, except for the point named A, where the roadsides have invaded some portions of neighboring fields, because the statistical parameters are practically the same for the road and for the invaded fields. In this case it will be necessary to modify the location of one resulting point, while the rest are all correct roadsides.

Seed points provided for this image are mainly placed at junctions, where they can help the algorithm to give the correct solution.

In the final example the roads are light and with few curvature changes, as can be seen in Figure 4.11, but some parts of the roads are surrounded by elements that are even lighter than the road to be extracted. Therefore, at these points the algorithm, which adds initial points between seeds, may give an incorrect answer, for example at the position labelled A, as may be observed in Figure 4.12.

The initial parameters chosen to obtain the result that appears in Figure 4.12 were not very restrictive. The solution at point A could either be to decrease the range to ascertain densification points, or to lower the curvature output tolerance, so the system cannot be diverted to those points that delineate such extreme curvatures. The complete result with more appropriate parameters is shown in Figure 4.11. In all the examples shown in this section the parameters have been set once for all the roads on the same image, and the variation among images is only dependent on the context: rural or urban, apart from the output resolution needs.

The same effect described at point A of the previous example may occur in the case of car parks, where some cars will divert the process and generate extreme curvatures, without any seed clue that makes it feasible. The solution

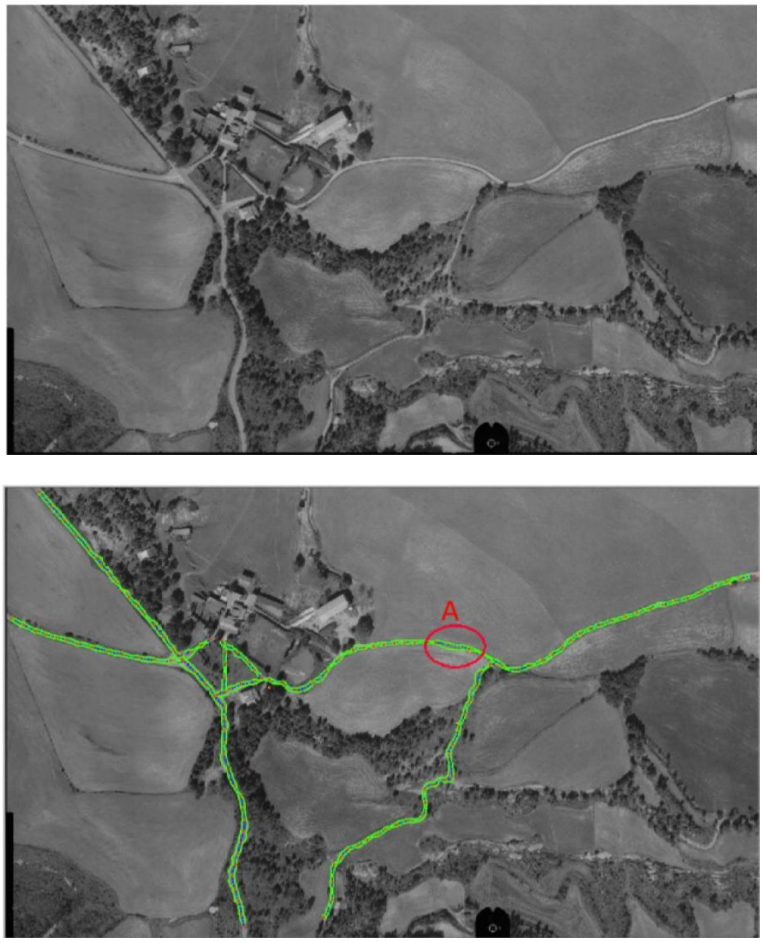


FIGURE 4.10: Original image and semi-automatic road extraction results obtained in the case of roads with similar homogeneity to their surrounding areas.

is analogous to the previous one. Due to the fact that car parks are very regular structures, it is appropriate to lower the tolerances and searching ranges.

We have tested the approaches in several images, some of which are presented in Figures 4.13, 4.14 and 4.15. Our aim is to show the algorithm behavior in different kinds of images: rural, urban, with high and low contrast, and roads surrounded by elements with similar radiometry. The numbers shown in table 4.1 compare the proposed approach with manual digitization in terms of the work required by the operator.

Figure 4.13 shows two kinds of results: those obtained by the approach presented here and those obtained by digitizing all the road components. As can be seen, the results are very similar; if they were to be superimposed, the only difference would appear at two points, where roads are partially occluded. This is because the automatic approach cannot find the roadsides radiometrically and only takes account of the geometric constraints; this will be slightly

FIGURE 4.11: Original image and semi-automatic road extraction results obtained in the case of roads with light surrounding elements that could lead to the failure of the automatic process.

different from the approach with regards to interpretation, in which the operator hazards a guess. On the other hand, as can be seen in Table 4.1 identified as Figure 4.13, the number of points that are needed to obtain the semi-automatic results is 10 times lower than the number of points needed when digitizing, and the number of resulting points by semi-automatic extraction is over twenty times greater than when digitizing. Compared with the results outlined manually, the operator cannot achieve the sub-pixel accuracy, although the outline can be drawn with fewer points. For this reason we have endeavored to reduce the number of points that represent the final result and some smoothing functions have been implemented. The function parameters depend on the accuracy requirements. The results presented in this paper do not have any simplification.

Figure 4.14 shows several images in which the road extraction approach has been performed, followed by the results. In Table 4.1 it may be observed

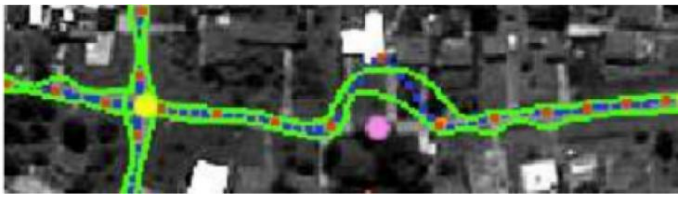


FIGURE 4.12: The semi-automatic solution, delivered by increasing the searching range parameter, which defines the buffer to look for seed densification.

Image Id	Seed Points	Editions	Resulting Points	Delineated Points
Figure 4.13	35	4	8003	822
Figure 4.14(a)	50	6	5923	540
Figure 4.14(c)	28	2	7011	939
Figure 4.14(e)	30	3	7588	1405
Figure 4.15	28	2	7252	410

TABLE 4.1: Comparison of the operator's work in the case of a semi-automatic approach and fully manual digitization. Each row has numbers referring to ImageId. The second column shows the seed points provided, and the third the editions needed to finally obtain the result, which has the number of points that appears in the fourth column. The final column shows the number of points that remain after the elimination of redundant points.

that in the worst case, if the number of seed and edition points are added, the work required by the operator has been reduced up to nine-fold. Moreover, the interaction is at the beginning and at the end of the process, in contrast to the continuous attention required by the fully manual or any *path finder* method. If the manual and semi-automatic outlines are overlapped, only slight interpretation differences will appear, which can hardly be noticed; most of these stem from the different amount of output points that appear in both cases. Since Figure 4.14(a) is an urban area, it needs more seed and edition points. This is also due to the weak curvature restriction and to the broad searching range required to obtain elements with different width. In Figure 4.14(c) these constraints can be stronger, producing better automatic results, with few edition requirements. The last image, Figure 4.14(e), has many occluded or missing road portions, but the geometric constraints deliver feasible results, which need few modifications.

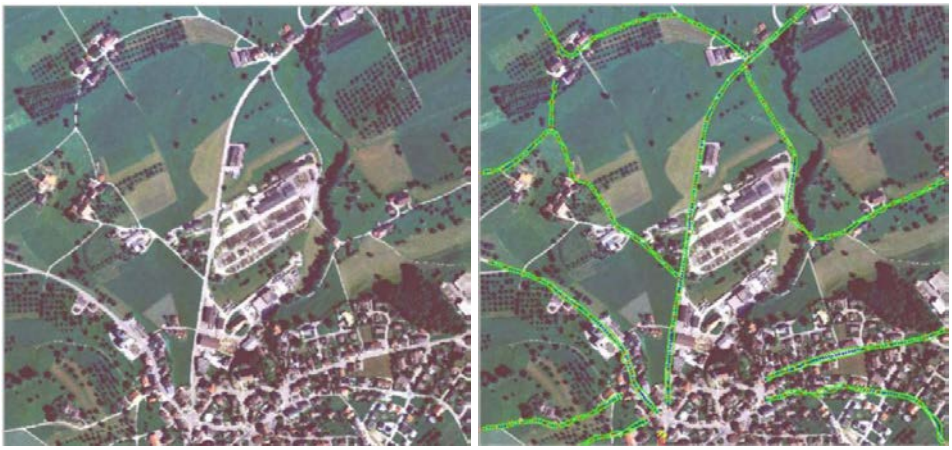
We have compared the results obtained by this algorithm with those obtained with one of the most successful general road extractor methods that exists, described and tested in [64]. In all cases, the number of points needed to obtain similar results is smaller using our approach, due to the fact that the occluded parts cannot be ascertained by the aforementioned method and the

(a) Original image (b) Semi-automatic road extraction results

(c) Results obtained by delineation

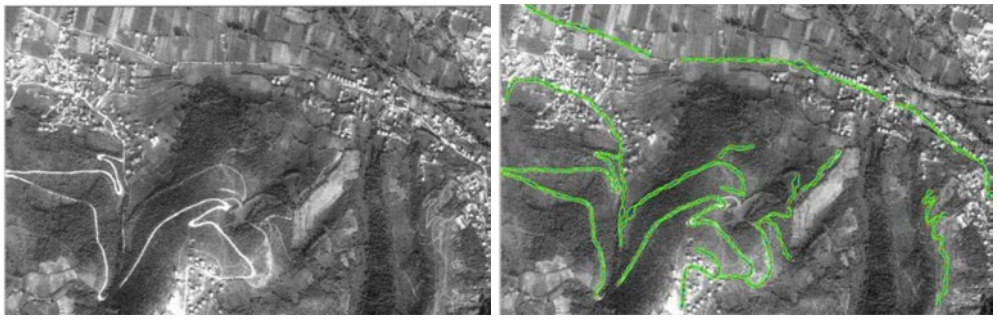
FIGURE 4.13: Comparison between the semi-automatic road extraction results obtained from the algorithm, and the results obtained by delineation.

system stops until the operator offers a new clue for reference profile matching. For example, in Figure 4.15 two points can be seen –circled in red– where the system failed and required two more interactions, apart from the three initial points delivered by the user to start the process. With the region competition approach, the two roads, where the problem appears, have been fully digitized by providing each one with three seed points. In the case of all the roads in Figure 4.15 the average number of points needed to obtain the results shown is three, and only in the case of two roads has some edition been required (deleting two points). If our method is compared with another path optimizer method, for example, the one presented in [65], the results obtained for the centerline are similar, but we need fewer seeds, due to the radiometric considerations taken into account to densify the centerline. Apart from this, the aforementioned system does not deliver the roadsides, whereas our method is robust with respect to finding margins accurately.



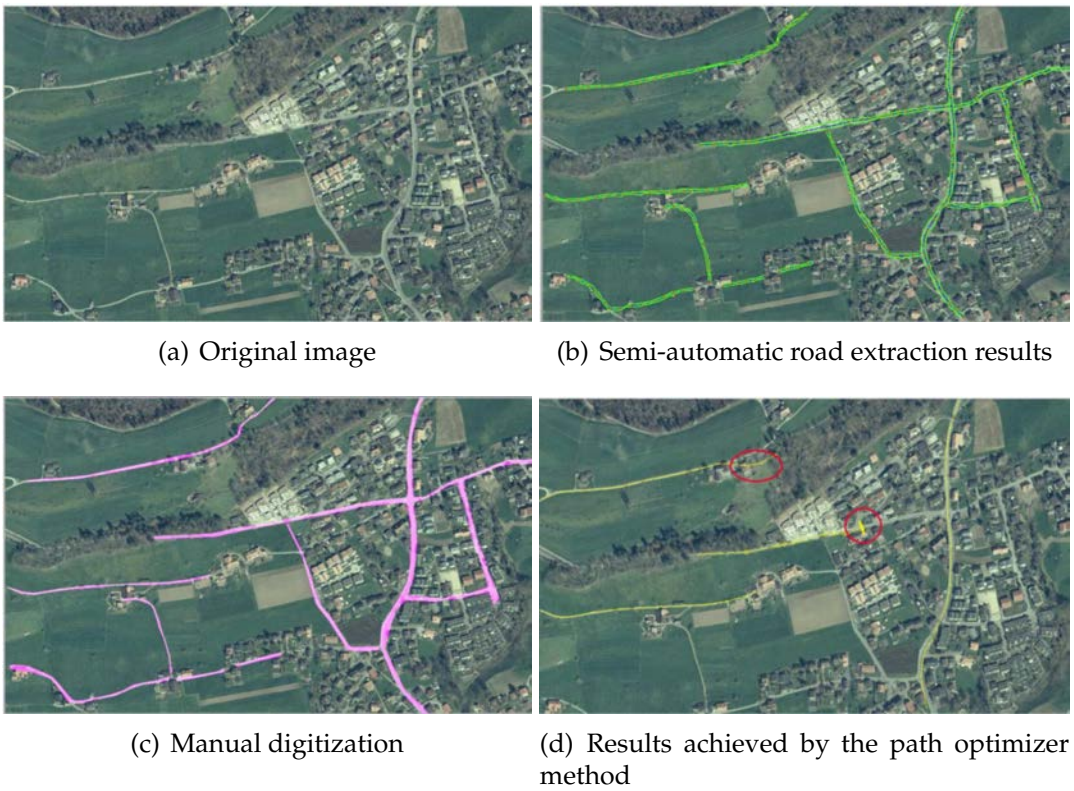
(a) Original image (b) Semi-automatic results

(c) Original image (d) Semi-automatic road extraction results



(e) Original image (f) Semi-automatic road extraction results

FIGURE 4.14: Several images in which the road extraction approach has been used.



(a) Original image (b) Semi-automatic road extraction results

(c) Manual digitization (d) Results achieved by the path optimizer method

FIGURE 4.15: The region competition road extraction method is compared with another semi-automatic approach.

4.9 Conclusions

With this approach we have shown the applicability of a combination consisting of region growing and region competition to extract roads: their centerlines and sides. The proposal is semi-automatic, so the role of the operator is focused on the provision of significant seeds and supervision, rather than delineation.

The majority of the roads that appear on aerial images can be extracted, because the algorithm is mainly based on their light appearance and on two characteristics: small changes in radiometry and in curvature. The region competition algorithm refines a first approach delivered by region growing techniques, and ensures that the road details that appear on the image will be recovered. Therefore, it will be a useful technique when recovering elements for large-scale purposes. The parameters to be delivered by the operator can be set up at the very beginning, once the image scale is known and the output vector scale desired is defined.

The tests have shown that a high degree of reliability can be obtained, and in some cases minor edition tasks can redress the undesired effects.

This research has been described in the published paper:

- Miriam Amo, Fernando Martínez and Margarita Torre. Road extraction from aerial images using a region competition algorithm. *IEEE Transactions on Image Processing*, vol. 15 (5), pp. 1192-1201, 2006.

Chapter 5

Deep Network Energy-Minimization

5.1 Introduction

In November 2008, to give a known example, the Institut Cartogràfic i Geològic de Catalunya made the last analogue flight. An important step was thus taken towards the complete digitalization of the photogrammetric workflow. This stops scanners from being used to introduce analogical images into the digital productive pipeline. Two limits were overcome in this step:

- The effective resolution of the digital images was good enough to ensure the recovery of all the feature details.
- The speed to save sensor information into a file on a disk in time to take the next photograph.

Little by little the size of the images has been approaching the analogue one. But some adjustments in time and distance between snapshots needed to be made. In addition, the number of images that covers an area also has been modified. A detailed analysis of a three-year experience in digital cameras is described in [74].

Now, from the flight to the digital photogrammetric station, everything is in digital format and with automated processes. We have partially overcome the challenge of providing support and automatic tools to the process of recovering geographic information that appears in aerial images.

The automation tools have to be reliable enough, because if after doing a process it is necessary that the human operators review it in depth, it is not worth it. We have carried out some tests with different approaches to try to improve the results obtained and described in the previous chapters. Once again, it was not until the appearance of the DL that a definitive step was taken in the reliability of automatic feature extraction.

5.1.1 State of the art

Despite hundreds of algorithms having been proposed for segmentation, only a few of them have been implemented and are available as a production tool [75]. Among them, eCognition [76] is the most popular and widely used segmentation software, converted into a productive software gold standard tool [77]. It is based on fuzzy segmentation allowing better retention of the radiometric variation of the agricultural regions. eCognition provides region contours by aggregation, which sometimes leads to over-segmentation and thus leaves room for improvement, mainly in post-processing editing tasks.

Edges have been another approach used in automatic and semi-automatic techniques for region boundary delineation [78]. However, these methods suffer the problems of detecting false edges, locating poor edges and missing edges, which limit its applicability. A more recent agricultural boundary detection technique used by Alemu [45] is the line segment detection algorithm, aimed at detecting straight contours in images [46]. But, the advantages and setbacks described for these methods in Section 3.3 still remain.

Energy-minimization theory delivers a common framework to unify different model-based approaches, such as graph cuts [49], random walker [50] and shortest path [48]. One limitation of these approaches is that, in practice, the edge-stopping in the minimizing function is never exactly zero in the edges, and so the curve may eventually pass through object boundaries.

In recent years, DL approaches have been achieving popularity and impressive performance in detection, segmentation and recognition of objects and regions in images [79]. DL techniques, and more specifically CNN, have also been applied in an important number of research works focused on edge detection. Among them, it is worth noted the Holistically-Nested Edge Detector (HED) ([39], [80]), an efficient and accurate network that performs image-to-image training and prediction. The proposed architecture connects its side output layers to the last convolution layer of each stage in a VGGNet [81]. Further works found in the literature focused their attention on HED, thus highlighting its reliability for edge detection. [82] used relaxed labels generated by bottom-up edges to guide the training process of HED. [83] proposed an edge detector that uses different image scales and aspect ratios to learn rich hierarchical representations, with an architecture that only adds 1×1 convolutional layers to HED. The term nested is due to the inherited and progressively refined edge maps produced as side outputs, thus making successive edge maps more concise. The term holistic, despite not explicitly modeling the structured output,

is because the network aims at training and predicting edges in an image-to-image fashion. More recently, [84] introduced a bi-directional cascade structure to enforce each layer (BDCN), which aims at focusing on a specific scale. However, deep learning techniques applied to agricultural field segmentation is a highly under-explored field of research.

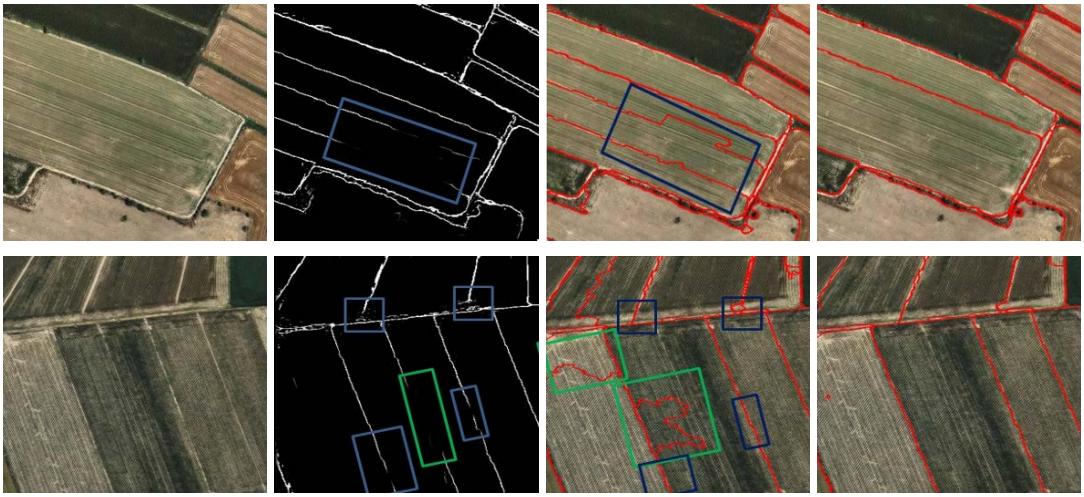
In terms of segmenting land coverage elements, [75] showed a complete review of segmentation methods widely used in this context, whose application is not limited to urban coverage. To enlighten some combined approaches, [85] presented two proposals based on a multi-paradigm collaborative framework: the first one is inspired by cascading techniques in machine learning, whilst the second one applies many collaborating one-vs-all class extractors in parallel. [86] proposed to use a partition delivered by the simple linear iterative clustering superpixel algorithm as a starting segmentation point. [87] presented another approach composed of a minimum spanning tree algorithm for the initial segmentation, and the minimum heterogeneity rule algorithm for object merging in a fractal net evolution approach.

5.2 Our approach for automatic field segmentation

All these methods show that the segmentation of aerial images should be based on edges instead of regions, since different agricultural fields often can be composed of similar or even the same crop (see Figure 5.1). For this reason, recent and powerful generic methods for semantic image segmentation based on neural networks [88] are unsuitable for aerial image segmentation.

Given the problem of locating and extracting agricultural fields in aerial images, the most relevant clues that assist users in detecting their boundaries are their regular shape and the fact that a field can be distinguished from its surroundings –since it is limited by linear elements or by other fields that do not follow its homogeneity or its texture pattern. As far as clues are concerned, edge extraction procedures play an important role in the problem at hand. For this reason, we propose taking advantage of the HED network [39], based on nested multi-scale feature learning and deeply-supervised networks [89]. HED is able to automatically learn the type of rich hierarchical features that are crucial in the process of disambiguation of natural aerial images and field boundary detection.

Deep networks such as HED are good candidates to fully extract an important portion of aerial field boundaries. However, some of them may not be detected, whilst other elements may be wrongly detected as boundaries. For



(a) (b) (c) (d)

FIGURE 5.1: (a) Two original images, (b) the edges delivered by HED (with gaps and isolated elements highlighted in squares), (c) the regions obtained with the energy-minimization process, and (d) our final solution after the model fitting step.

this reason, [83] proposed refining the output provided by HED, using only the pixels with the greatest amount of annotators labeled as positive samples, due to their high consistence and their ease of training. When working with aerial images, gaps or isolated elements may also appear when applying edge detectors such as HED, as illustrated in Figure 5.1(b). In this case, all the edge pixels obtained with HED must be considered, and properly selected depending on their reliability in model extent, thus making the approach proposed by [83] not entirely adequate. Although HED provides excellent results for image edge detection, it cannot ensure obtaining closed boundaries of image regions. To this aim, we propose to integrate the HED detector into an energy-minimization framework. The main idea is that the edges obtained with HED must be processed, categorized and completed by the energy-minimization process to assure straightforward extraction of agricultural fields. Moreover, the boundaries added must be finally accepted or not depending on whether they follow sufficiently the tracks given by the edges delivered by HED. Figure 5.1 (c,d) shows how the energy-minimization step followed by a model fitting process are able to solve the problems detected when applying HED (i.e., gaps and isolated elements).

Our proposal, called DeepNEM, is an automatic global segmentation approach that integrates HED as a Deep Network (DN) for edge detection with an Energy-Minimization (EM) procedure. DeepNEM relies on boundary clues, which must follow a predefined model and are also used to complete the

boundaries by an energy-minimization global process.

5.3 Methodology

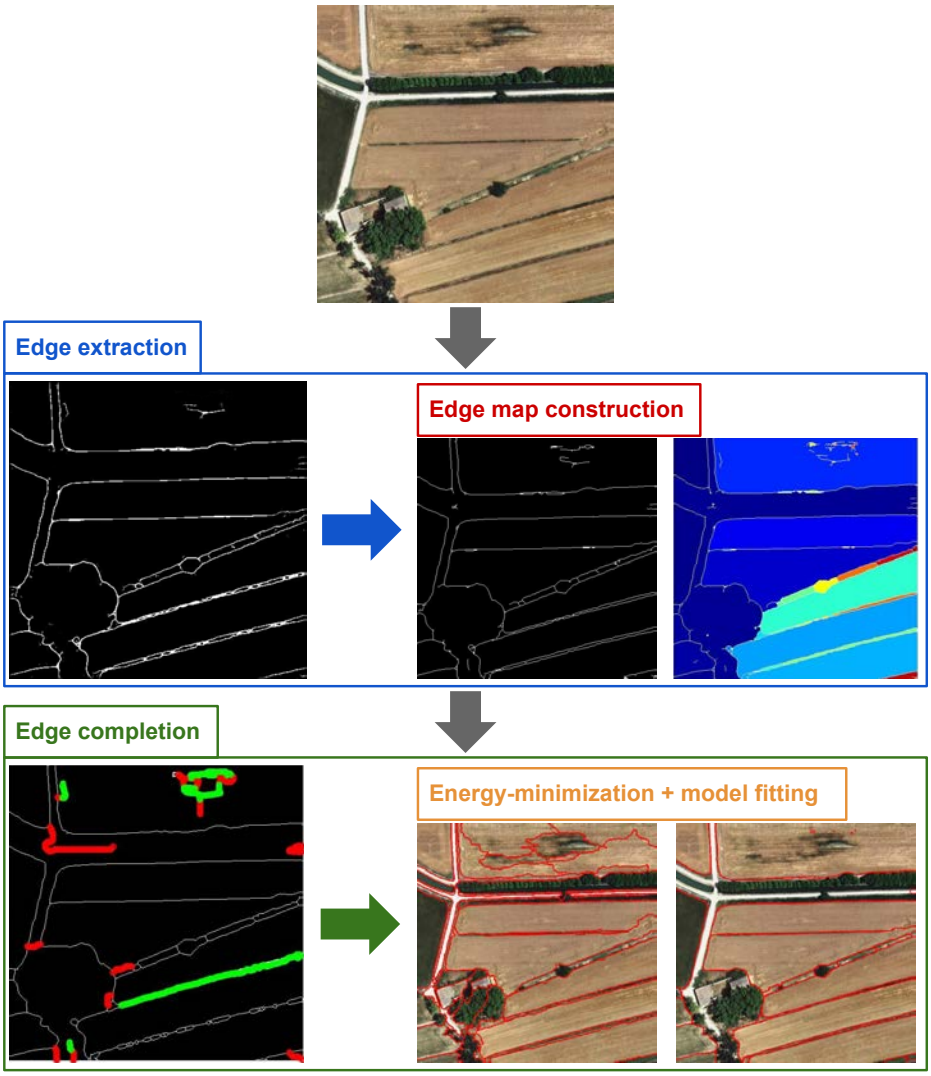


FIGURE 5.2: Main steps of our DeepNEM approach for region segmentation of agricultural fields. The edge extraction step (second row) goes from the edges delivered by HED (left) to the selection of relevant ones (center), and the main regions formed by them (right). The edge completion (third row) is composed of the energy minimization step applied to the edges previously selected (left) –red lines are edges allowed to be modified–, whose results are expanded under radiometric constraints (center), and the final output with the fields that fulfill the model (right).

Discontinuity in terms of radiometry or texture homogeneity is the characteristic that catches the eye when operators delineate the field boundaries in aerial images. Since these interruptions are the common evidence among

neighboring fields, we strongly rely on edges to drive and define the mainstream of the process. Our method is divided into two main steps, as depicted in Figure 5.2: edge extraction and edge completion. These main steps contain side refinements to reinforce some evidence and to complete or dismiss some clues. In particular, the edge extraction is completed with morphological operations to deliver an edge map, whilst the edge completion delivers enough cues to analyze if they are worth including in the final solution. Below, both steps are described in depth.

5.3.1 Edge extraction: from image to edges

To recover discontinuities in radiometric or texture homogeneity of aerial images, we rely on HED [39] in order to locate regions of contrast changes, usually corresponding to different fields, and to maintain their singularities such as wooded areas, trees and high contrast linear elements. The HED architecture, illustrated in Figure 5.3, is composed of a single-stream DN with multiple side outputs, and uses a deep architecture to simulate the perceptual multilevel human approach.

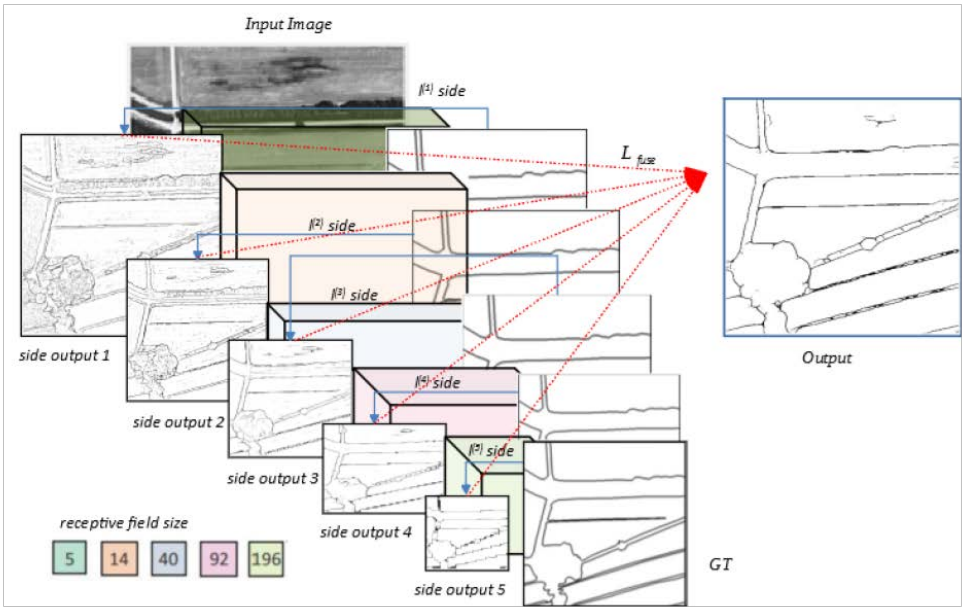


FIGURE 5.3: Illustration of the HED architecture for edge detection over an aerial input image, highlighting the error back propagation paths. Note that side-output layers are inserted after convolutional layers, which are shown as colored boxes, with the side output plane size becoming smaller and the receptive field size becoming larger.

Notice that the structure in multiple stages with different strides is useful to capture the inherent scales of edge maps. There is also a weighted-fusion layer-error to help with the update of the output-layer parameters by back-propagation and to learn the fusion weight during training.

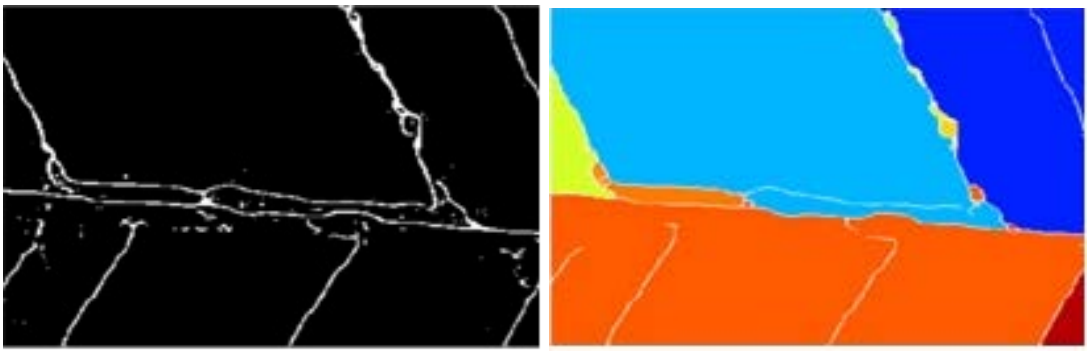
Figure 5.3 depicts the error back-propagation paths as well as the importance of inserting the output convolutional layers. For each side-output layer, deep supervision is imposed, guiding the side-outputs towards edge predictions with the desirable characteristics. Figure 5.3 also shows that the outputs of HED are both multi-scale and multi-level, as well as how the side-output-plane size becomes smaller while receptive field size becomes larger. Note that the entire network is trained with multiple error propagation paths (dashed lines).

Despite the high quality edge detection by HED, it does not assure closed regions. For this reason, we rely on an energy-minimization model to guide the process that extracts from the edges as many clues as possible, moreover to complete them when it is necessary. Some linear elements detected by HED may reflect confusing clues, because they lie inside some regions (if kept, they will over-segment them) or near to hard edges, when undoubtedly reliable boundaries are extracted. For this reason, we propose a side refinement of this stage, guided by an energy-minimization model that keeps the knowledge of what real agricultural fields are like.

Edge map construction

Among all the edges detected by HED, it is necessary to select the ones that better suit an agricultural field segmentation, since they define a preliminary image division into the main regions. If we process all the edges in terms of linear elements, we will find that some of the selected segments will not deliver relevant clues (e.g., close parallel elements that cause closed elongated areas, or small gaps between long segments that the model tends to close). Since we want to detect areas, instead of acting directly on the linear evidence, we will always take into consideration the regions that edges form. This approach will handle boundary information such as bushes, regardless of whether they divide regions or appear within closed regions, as well as other important elements (see Figure 5.4).

In order to reinforce strong clues or to eliminate these potentially misleading clues, a morphological process is applied, which is composed of several serialized morphological operations to keep and merge adjacent small areas



(a) (b)

FIGURE 5.4: (a) Output edges obtained with HED, and (b) result in terms of structuring edges.

and to narrow boundaries. These operations are a sequence of opening, closing and thinning morphological functions to filter profiles, without losing relevant evidence. This process is addressed by a predefined threshold, named A_{min}, which represents the minimum size of the allowed artifacts.

5.3.2 Edge completion: from edges to regions

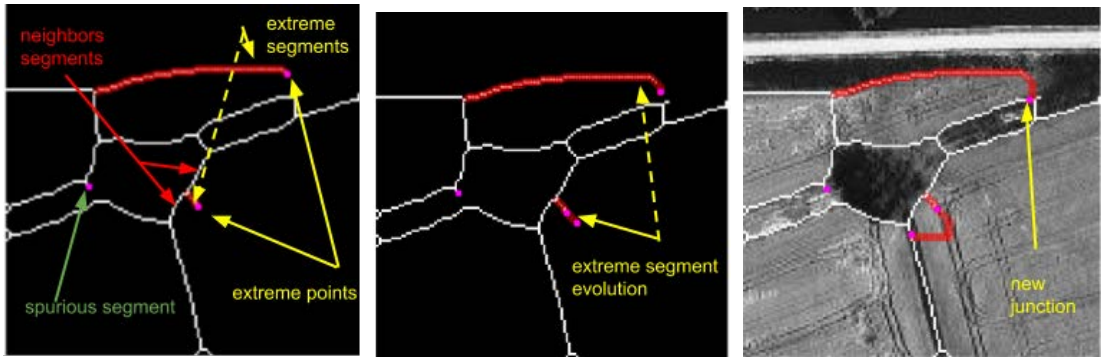
At this point, the image is divided into regions whose boundaries are completely closed by edges, named tiles. For each one, the algorithm analyzes its content, when it has edges with gaps or isolated edges. The last ones are prone to becoming part of a boundary, while the others are analyzed to be completed. Despite the high quality of the edges provided by HED, the algorithm does not guarantee that edges would form closed contours. An additional step is necessary to complete clues and avoid under-segmentation problems[1]. Figure 5.4 shows an example of boundaries with gaps and others with isolated clues.

Energy-minimization

In order to complete the boundaries, we integrate the HED output into an energy-minimization process that obtains a complete first approximation of the boundary segmentation inside each tile. This process finds the shortest sequence of pixels between segments that have an extreme. The sequence is obtained by forcing the total energy of the edges to be minimum. Not only does it take into consideration the nearest environment of each gap, but it also considers each edge within a broad area.

[1] A key problem in segmentation is that of dividing a region into too few (under-segmentation) or too many areas (over-segmentation).

The tiles obtained in the previous stage, formed by a closed chain of long segments, will be the domain of the edge completion step. This process is driven by *relevant points*, namely X-connected points (those with X edge pixels among their eight closest neighbors) with $X \neq 2$, and divided into: *extreme points*, when $X = 1$; and *junctions*, when $X \geq 3$. For each *extreme point*, its associated chain of 2-connected points is considered a *segment*. Note that segments are divided depending on their length, defined by the parameter T_{min}, into long and short segments. The different classes of long and short elements are illustrated in Figure 5.5, and described below.



(a) (b) (c)

FIGURE 5.5: (a) The different elements involved in the energy-minimization step, (b) their evolution throughout the process, and (c) the completed edges obtained at the end.

Long segments are classified into:

- *Extreme segments*: segments limited by at least one extreme, whilst the other end can be an extreme or a junction.
- *Arcs*: segments limited by two junctions.

Short segments are considered in the minimization process, but are forbidden from growing. They are classified into:

- *Isles*: segments limited by two *extreme* points. They will be taken into account in the filling process to attract the edges.
- *Spurious*: segments limited by an extreme point and a junction. In the junction, they join a single long segment, alongside other spurious ones that may also share the junction.

For each *extreme* segment, the energy-minimization will deliver a new end location at each step. This sequential iterative process will stop when, in the

elongation process, another segment is reached. Depending on the type of element reached, the process will end by elongating both segments to a common point, or by creating a junction. Next, these elements whose length is under a threshold (T_{min}) are deleted (short segments), whilst the other ones are kept (long segments).

The minimization result follows the desired model, which is a combination of smoothness and minimum length segments, where the relevant points are the *junctions* and *extremes*. For each *extreme* segment, a potential is defined by a force that rejects the *extreme* point –the n_i pixels that form the segment (x_i, y_i) try to expel the *extreme*–, and a force that attracts this *extreme,* due to all other surrounding segments. So, for each *extreme* segment, its *extreme* point (x_{n_i}, y_{n_i}) will be newly located at the position that tries to minimize the functional:

$$V_{n_i} = \sum_{i \in edge} \frac{1}{r_{i,n_i}} - \sum_{i \notin edge} \frac{1}{r_{i,n_i}},$$

where $r_{i,n_i} = \sqrt{(x_i - x_{n_i})^2 + (y_i - y_{n_i})^2}$. Minimizing V_{n_i} is equivalent to solving[2]:

$$\vec{F} = -\vec{\nabla} V = 0, \qquad \vec{\nabla} V = \left(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y} \right).$$

We propose a greedy method that at each step finds a new *extreme*, stopping when this *extreme* reaches another segment. If the segment reached is an *extreme* segment, it can be analyzed to become a corresponding element. Otherwise, the process will stop by reaching an *arc* segment and the growing process is frozen. Note that the *extreme* segment will be awakened if another *extreme* segment connects to it, during its growing process. In such a case, the segments will become corresponding segments. If the process is completed and no other *extreme* segment has been reached, a *junction* point in the *arc* segment is created. Figure 5.5 illustrates an example of these elements.

The notation used in the detailed formulation is as follows:

- For each *extreme* segment, e^i, we describe its coordinates obtained from the edge extraction as $\vec{e}^i_j \{j = 1 : n_i\}$, where n_i is its length.
- Associated with each element, e^i, we code the *spurious* segments, such as s^{ij}, where $j \in \{1 : n\_esp_i\}$. The identification of segments is esp^i_l, where $\{l = 1 : n\_esp_i\}$. Figure 5.5 includes an example of *extreme* and their spurious segments.

[2]If a multi-variable function $V(\mathbf{x})$ is differentiable in a neighborhood of a point $\mathbf{x_0}$, then $V(\mathbf{x})$ decreases fastest if one goes from $\mathbf{x_0}$ in the direction of the negative gradient of V at $\mathbf{x_0}$.

- The *arcs* that connect directly with *extreme* segments (e^i) are called neighbors; being ngh_i the number of segments associated with e^i. The components of these segments are *prohibited* locations for the growing process in the first iterations, in order to avoid loops.

- During the growing process, the *extreme* segment is allowed to reach another segment that is not from its neighbor. In this case, we consider that the segment touches an edge at the position $\vec{p}$ (bit on). $I(\vec{p})$ represents the edge image value (1 or 0) at $\vec{p}$.

In the energy formulation, for each *extreme* segment e^i, we consider that the rejecting force over the point $\vec{x}$, $f_R(e^i, \vec{x}) = \sum_{k=1}^{n_i} \frac{\vec{e}_k^i - \vec{x}}{\|\vec{e}_k^i - \vec{x}\|^3}$, comes from the *extreme* segment components: both the original and the added points, as well as the expelling force due to the associated spurious segments $\vec{s}_k^{ij}$ and the *neighbor* segments. Each one is weighted in different ways, depending on the confidence in the data. For example, the weight associated with the edge points, ω_e, is greater than the one associated with the added coordinates, ω_{add}. The *spurious* segments linked to an *extreme* segment will produce the same effects as the edge points of the *extreme* segment. For that reason, the parameters ω_{esp} and ω_{ngh} normally have the same value as ω_e. As we can see, the point distance will reduce the magnitude of the force.

The subtractive contribution, $f_A(e^i, \vec{x}) = \sum_{k=n_i+1}^{m_i} \frac{\vec{x} - \vec{e}_k^i}{\|\vec{x} - \vec{e}_k^i\|^3}$, carried by the attracting components, comes from the rest of the *arcs* not directly connected to the *extreme* segment, e^i. We apply the same reasoning to the rejecting force, for the different elements: the original edge points are stronger than the added ones. We use ω_{biton} to weight the elements that come from the edge image map; and we take into account especially the points in a round environment $B(\cdot, r)$ –in the examples we have chosen $r = 2 \cdot A_{min}$, due to the fact that larger circular areas take more computational time and the results do not improve. From the final *extreme* position $\vec{e}_{m_i}^i$, the force delivers a new segment end position by generating a new component $\vec{e}_{m_i+1}^i$:

$$\begin{aligned}\vec{e}^{\,i}_{m_i+1} =& \vec{e}^{\,i}_{m_i} + \omega_e\, f_R(e^i, \vec{e}^{\,i}_{m_i}) + \omega_{esp} \sum_{j=1}^{n\_esp_i} f_R(s^{ij}, \vec{e}^{\,i}_{m_i}) \\ &+ \omega_{ngh} \sum_{j=1}^{n\_ngh_i} f_R(e^j, \vec{e}^{\,i}_{m_i}) \\ &+ \omega_{biton} \sum_{\substack{p \in B(e^i_{m_i,r}) \\ I(\vec{p})=1}} f_A(p, \vec{e}^{\,i}_{m_i}) \\ &+ \omega_{add} \sum_{\substack{j=1 \\ j \neq i}}^{n_e} f_A(e^j, \vec{e}^{\,i}_{m_i}).\end{aligned}$$

The first phase of the process finishes when all the *extreme* segments stop because they have reached a corresponding *extreme* segment or created a junction with an *arc* segment. The second phase starts with only one difference: the attracting contribution is reduced to the corresponding *extreme* segment l, providing it exists. The evolution equation is:

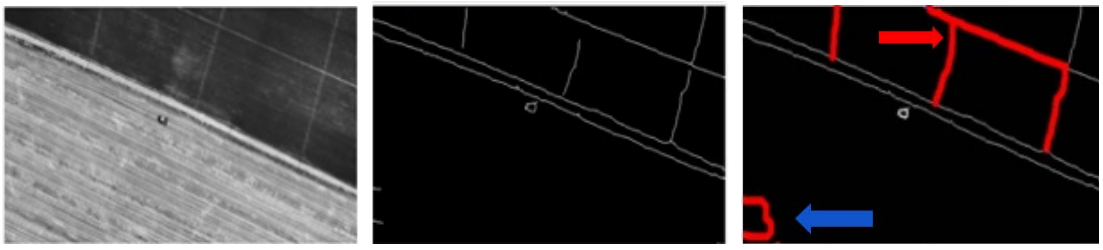
$$\begin{aligned}\vec{e}^{\,i}_{m_i+1} =& \vec{e}^{\,i}_{m_i} + \omega_e\, f_R(e^i, \vec{e}^{\,i}_{m_i}) + \omega_{esp} \sum_{j=1}^{n\_esp_i} f_R(s^{ij}, \vec{e}^{\,i}_{m_i}) \\ &+ \omega_{ngh} \sum_{j=1}^{n\_ngh_i} f_R(e^j, \vec{e}^{\,i}_{m_i}) + \omega_e f_A(e^l, \vec{e}^{\,i}_{m_i}) \\ &+ \omega_{add} \sum_{\substack{j=1 \\ j \neq i}}^{n_e} f_A(e^j, \vec{e}^{\,i}_{m_i}).\end{aligned}$$

The gaps between edges left by edge extraction are completed by taking into account the fact that the total energy of the edges must be at a minimum. By doing this, one of the advantages of the model is achieved: we obtain regular boundaries. Some examples can be seen in the third step (red lines) of Figures 5.2 and 5.6.

Model fitting

Once the energy-minimization has been applied, the image is completely segmented in terms of the number of fields and their approximate boundaries. But, for each boundary that has been completed, it is still necessary to analyze the ratio between the length of the segments added and the one that comes

from the HED extraction. Also, isolated long edges are analyzed to decide whether they are clues to add the complete division of the field or if they represent just a groove. Moreover, it is necessary to take into account the length of the contours to avoid closing small regions over themselves. This is similar to the way that the edge completion step rejects reaching the segment that grows at the beginning of the process. These processes are done not only using the segments, but also taking into account the image information inside the output regions. An added line in a boundary is analyzed, and even modified, by a local radiometry growing process that uses a threshold (Add_{max}). The boundary modified is kept when the number of non-overlapping pixels, between the original region and the output one, is lower than Add^2_{max}. Otherwise, if the added pixels are lower than Add_{max}, the addition is admitted. Figure 5.6 depicts how the outputs of the energy minimization phase produces edges that will be included in the final results, depending on whether these additions stay under the Add_{max} parameter.



(a) (b) (c)

FIGURE 5.6: (a) Input image, (b) edges obtained with HED, and (c) completed edges. The red arrow points to the accepted added boundaries, whilst the blue one points to the area rejected.

5.4 Experimental results

This section includes a description of the dataset, the implementation settings, the statistical measures used for validation purposes, and the experimentation carried out to evaluate the proposed method, including a comparison with state-of-the-art methods and a discussion of the results.

5.4.1 Agricultural field dataset

To the best of our knowledge, there are no representative public datasets that fit our goal of segmenting agricultural fields from High-resolution Visible (HRV)

images. For this reason, we built a complete dataset composed of 1200 HVR images and evaluated our approach on it. Moreover, it is publicly available ([90], [91]) to serve as a benchmark for comparing the agricultural field segmentation of different algorithms. Our dataset is composed of 1200 HVR images (RGB, spatial resolution: 500×500 pixels), and their associated ground truth (GT) delineated by a human operator. Figure 5.7(a) shows thirteen original images and their corresponding ground truth delineated by a professional experienced in manual aerial boundary delineation.

The dataset is composed of images available on the Institut Cartogràfic i Geològic de Catalunya website [92], which are parts of 1:25.000 orthophotos. We have chosen areas with assorted agricultural field appearances and from the agricultural regions of Catalonia, such as la Plana de Lleida, Baix Camp and Penedès (Tarragona, Spain). The flights to obtain the aerial images which form orthophotos were taken under clear weather conditions. The growing state of the crop is not important as long as the fields can be distinguished from their surroundings or are surrounded by linear elements, such as roads or water streams. This is one of the main advantages of mainly relying on contrast lines instead of doing it with radiometric clues. We have selected several types of crops such as wheat, corn, hay, olive orchard, vineyard and fruit trees.

The images contain more than 3300 agricultural fields. It is worth noting the great variability of the images, not only in terms of crops or textures, but also in size, shape and different kind of elements acting as boundaries. Note also that the dataset contains fields with limits not completely defined, as well as others that contain some isolated elements, such as trees, bushes or grooves.

5.4.2 Implementation settings

The HED network delivered was trained from scratch, using an initial learning rate of 10^{-5}, which was lowered by 10 times each 1000 epochs. Note that the learning rate is important to start the training process, since the process does not converge when it is set to a value greater than 10^{-5}.

For experimentation purposes, the dataset has been split into train and test partitions with the following distribution: the training set contains 920 images, whilst the test set includes 280 images.

Data augmentation has proven to be a crucial technique to obtain more reliable results when dealing with DNs with a large amount of parameters, as a way to enlarge the dataset in order to train the network. In this sense, we

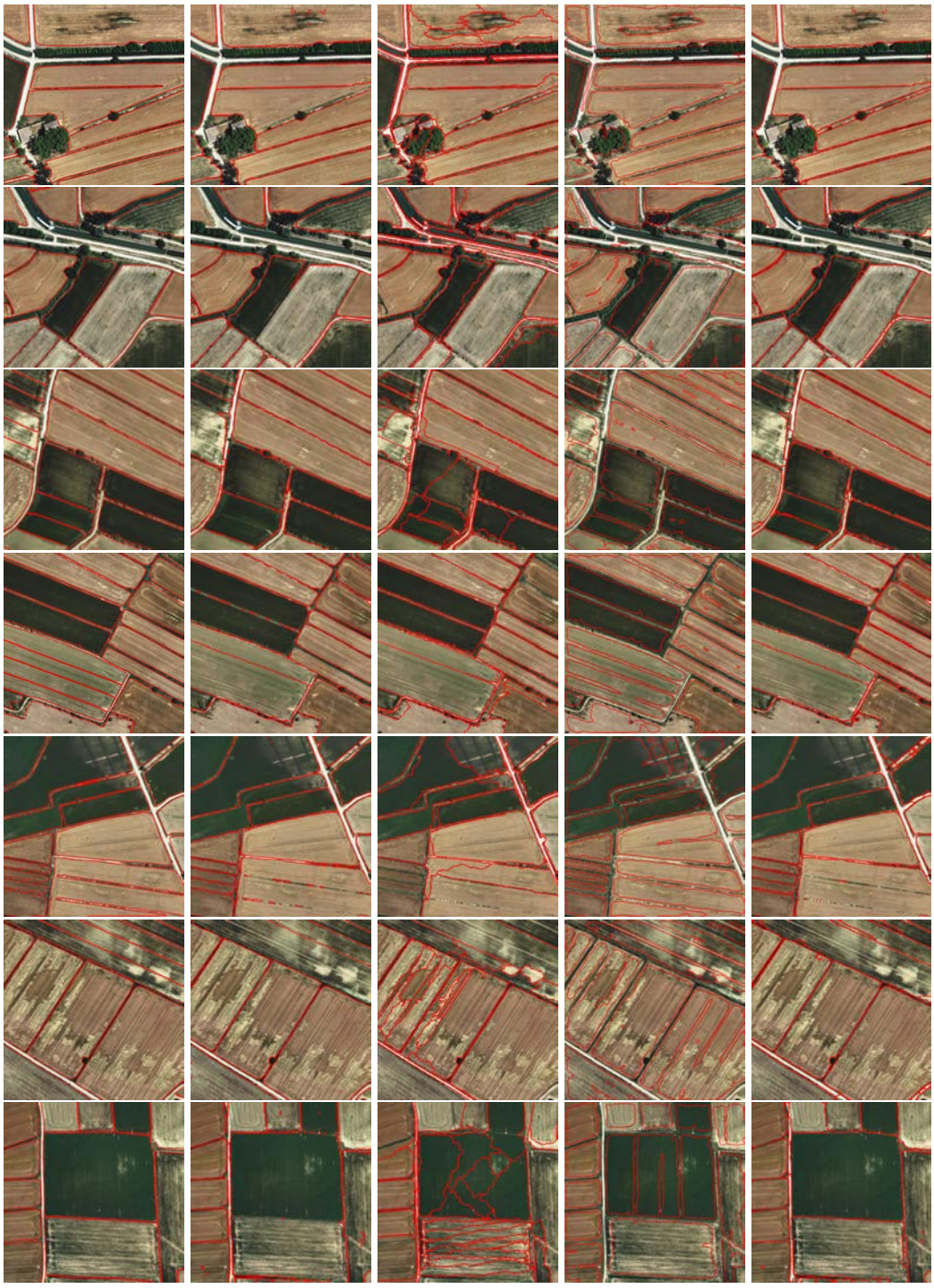
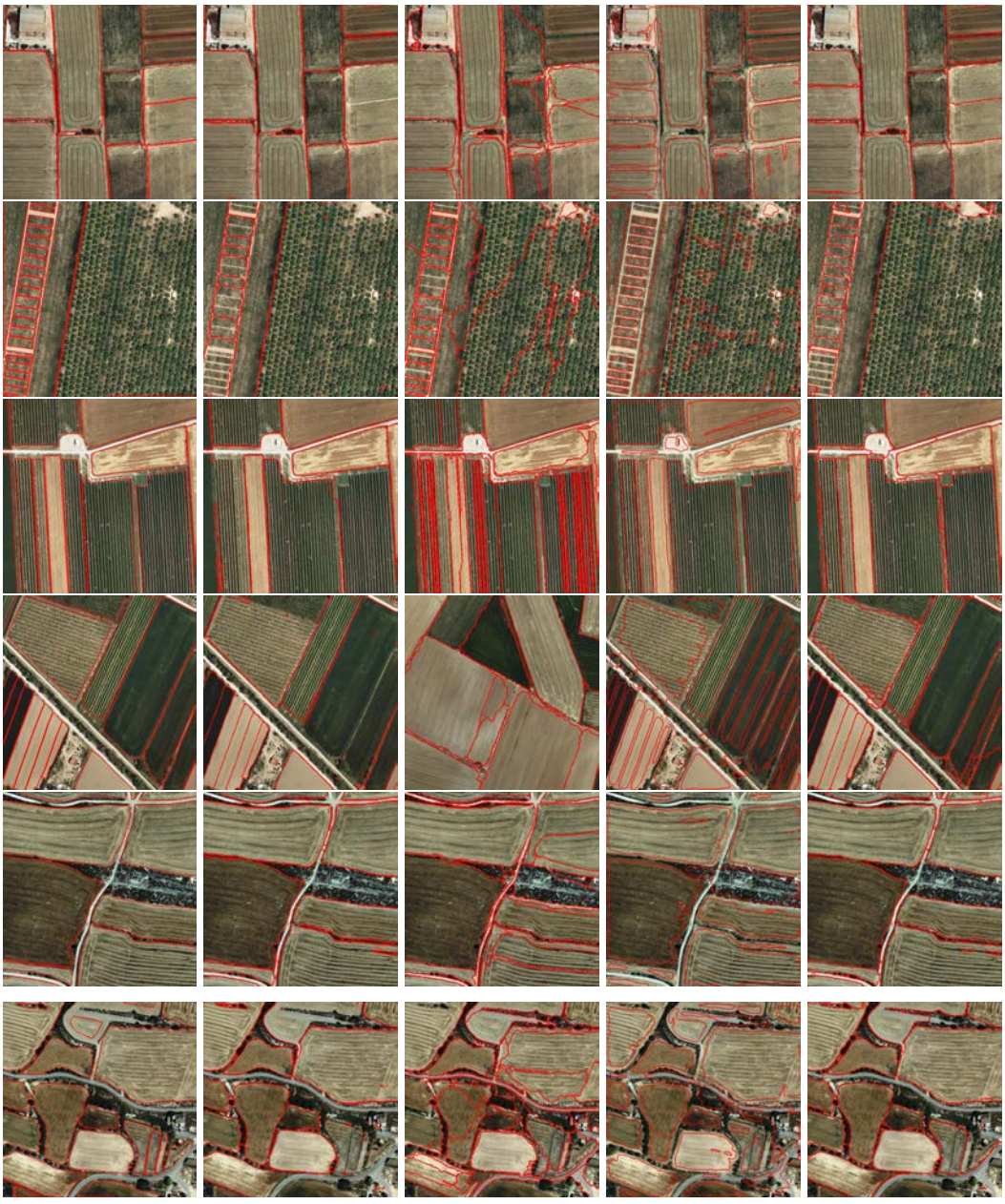


FIGURE 5.7: (a) Input images with boundaries drawn by a human operator (GT), (b) HED results, (c) eCognition results, (d) BDCN results, and (e) DeepNEM results. These images represent a wide range of agricultural fields, in terms of radiometry and texture.



(a) (b) (c) (d) (e)

FIGURE 5.7: (a) Input images with boundaries drawn by a human operator (GT), (b) HED results, (c) eCognition results, (d) BDCN results, and (e) DeepNEM results. These images represent a wide range of agricultural fields, in terms of radiometry and texture.

rotated the images at 16 different angles and cropped the largest rectangle in the rotated image. Moreover, we flipped these images at each angle, leading to an augmented training set that is a factor of 32 larger than the original one. All these images were scaled to the half and to the double. To sum up, we used a total of 96×920 training images. After the training process with the augmented data, the method was tested over the 280 images, and the results were compared to the GT.

5.4.3 Performance measures

Some quantitative metrics were used to evaluate the performance of DeepNEM. On the one hand, we consider the Jaccard distance (JD) between two regions A and B, which represents the number of pixels that fall into the intersection of both regions, normalized by the pixels counted by the union (also known as *intersection over union*):

$$JD = \frac{|A \cap B|}{|A \cup B|}. \tag{5.1}$$

For a given image, we use this equation to compute the JD between all the regions extracted from an input image and its corresponding GT regions. The average JD is calculated as the mean value of the Jaccard distances computed for all the pairs of fields in the image.

On the other hand, the under- and over-segmentation must also be considered in the problem at hand. For this purpose, we defined three different types of regions:

- Type A: one extracted region that fully represents one GT field.
- Type B: two or more extracted regions that represent one GT field (over-segmentation).
- Type C: one extracted region representing two or more GT fields (under-segmentation).

In this case, we calculated the number of regions for each type. It should be noted that for regions of type A, the higher the number, the better; while for the regions of type B and C, the lower, the better.

Notice that these distinctions among the different ways of recovering fields are necessary and highly related to human interpretation, as shown in Figure 5.8. Some fields are clearly defined but, in others, the human operator may add some lines or may erase others. The automatic process needs clues and only

rejects them when they do not follow the model clearly, not by interpretation. The red areas in GT are fields recovered only by a single region. On the other hand, green ones are fields clearly split into two or more areas by DeepNEM, whilst the opposite phenomenon is shown in blue color. For these reasons, it is necessary to compute the JD not only when the correspondence falls in Type A category, but also when fields have been clearly split or merged with neighbors. In some of these last categories, human delineation could have split or merged delineations.

(f) (g) (h)

FIGURE 5.8: (a) Input image with delineated GT, (b) results obtained with DeepNEM, and (c) results in terms of over- and under-segmentation: regions type A in red, type B in green, and type C in blue.

Additionally, we consider three region-based metrics ([93], [94]) commonly used in different segmentation problems:

- *Covering (CO)*. It represents the level of overlapping between each pair of regions (R and R') corresponding to the ground truth (GT) and the output (O) images:

$$CO = \frac{1}{N} \sum_{R \in GT} |R| \cdot \max_{R' \in O} \frac{|R \cap R'|}{|R \cup R'|} \tag{5.2}$$

  where N is the number of pixels of the image.

- *Rank index (RI)*. It represents the compatibility of assignments between pairs of elements in the ground truth and the output images:

$$RI = \frac{1}{\binom{N}{2}} \sum_{i<j} \left[\mathbb{I}(t_i == t_j \wedge p_i == p_j) + \mathbb{I}(t_i \neq t_j \wedge p_i \neq p_j) \right] \tag{5.3}$$

where $\binom{N}{2}$ is the number of possible unique pairs among the N pixels of each image, and $\mathbb{I}$ is the identity function.

- *Variation of information (VI)*. It represents the distance between the ground truth (GT) and the output (O) images in terms of their average conditional entropy:

$$VI = H(O) + H(GT) - 2 \cdot MI(O, GT) \tag{5.4}$$

where H and MI are the entropy and the mutual information, respectively. In this case, the lower the better.

Finally, we also consider three boundary-based metrics [94]. For this purpose, we used the edges (boundaries of the regions) to compute three standard measures commonly used in different learning tasks:

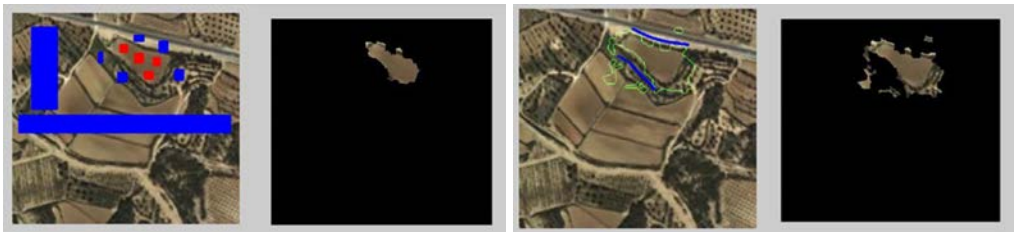
- *Recall*. It represents the proportion of true positives correctly classified.
- *Precision*. It represents the proportion of true positives against all the positives.
- *F-measure*. It represents the harmonic mean of precision and recall.

Notice that the three region-based metrics and the three boundary-based measures were calculated for each single image, and then the mean across images was computed.

5.4.4 Validation

Other semi-automatic flooding approaches, such as an unified graph-based framework, have been also tested. Figure 5.9 shows the obtained results in terms of areas which look similar, although the watershed approach delivers smoother boundaries. Even with more seeds or more cues to delimit the regions, the results do not improve, as can be seen in Figure 5.9(b), due to the fact that the difference between the representative grey levels associated with a set of contiguous fields is very small. We have compared our DeepNEM method with power watershed that comes from the unified framework [95] and it outperforms the previous image segmentation grouping techniques. This unified power watershed framework offers the possibility of refining the results by delivering more image cues into the segmentation process. The benefits of the efforts to automatically place the seeds, in order to address the flooding process, are visible when comparing the DeepNEM results with those obtained

when seeds are initially placed randomly. Figure 5.10 illustrates the process, with the random seeds placed over the original image (first row) and the results of the watershed flooding (second row), which are significantly different from those obtained by the DeepNEM (right column). In terms of over/under-segmentation, our proposal also delivers better numbers as shown in Table 5.1 whose fields are in Figure 5.11.



(a) Some initial seeds were delivered to distinguish between the field and its surroundings: results on the right.

(b) Refinement is applied to the results, by drawing some blue lines that delimit areas that do not belong to the region.

FIGURE 5.9: Graph cut by delivering cues either in seed form at the beginning of the process, or providing some lines to refine in the second step of the energy-minimization process.

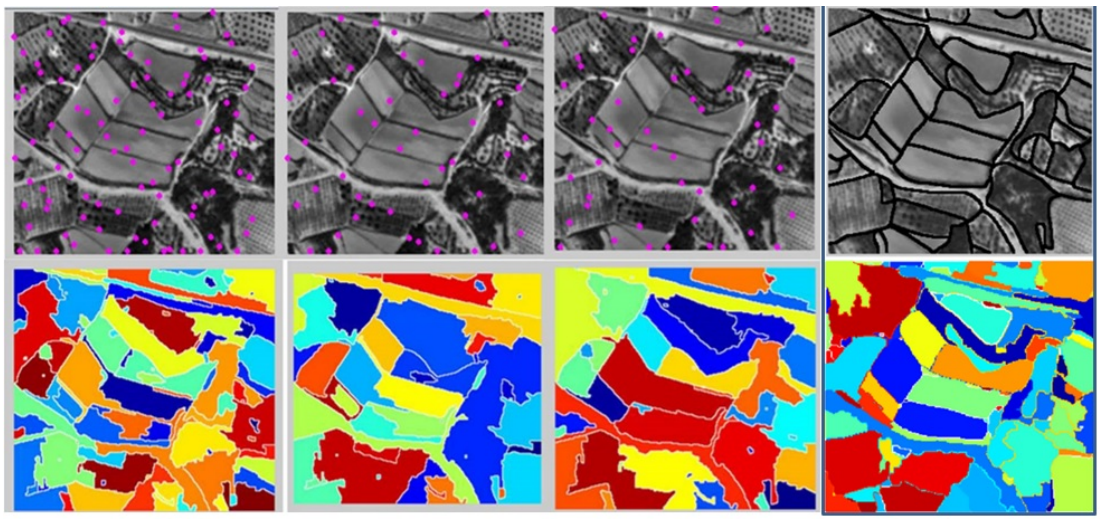


FIGURE 5.10: Comparison of the original algorithm watershed with different random seeds location (the first three columns of both rows) and our DeepNEM approach shown in the last column: on the top the image with the ground truth, on the bottom the result of DeepNEM.

FIGURE 5.11: From left to right: the original image, the ground truth, the results obtained with the power watershed, and the results provided by our proposal.

	Type A	Type B	Type C
PowerWatershed	11	0	10
DeepNEM	20	0	9

TABLE 5.1: Distribution of the fields, depending on the qualitative measures, relating to the extraction represented in Figure 5.11.

5.4.5 Results and Discussion

Robustness of the proposed DeepNEM

In terms of DeepNEM algorithm, there are three parameters that affect the capacity of the method to complete the field boundaries and to provide the final segmentation:

- A_{min}, the minimum isolated length element.
- T_{min}, the length of spurious segments.
- Add_{max}, the number of pixels allowed to be added for each boundary.

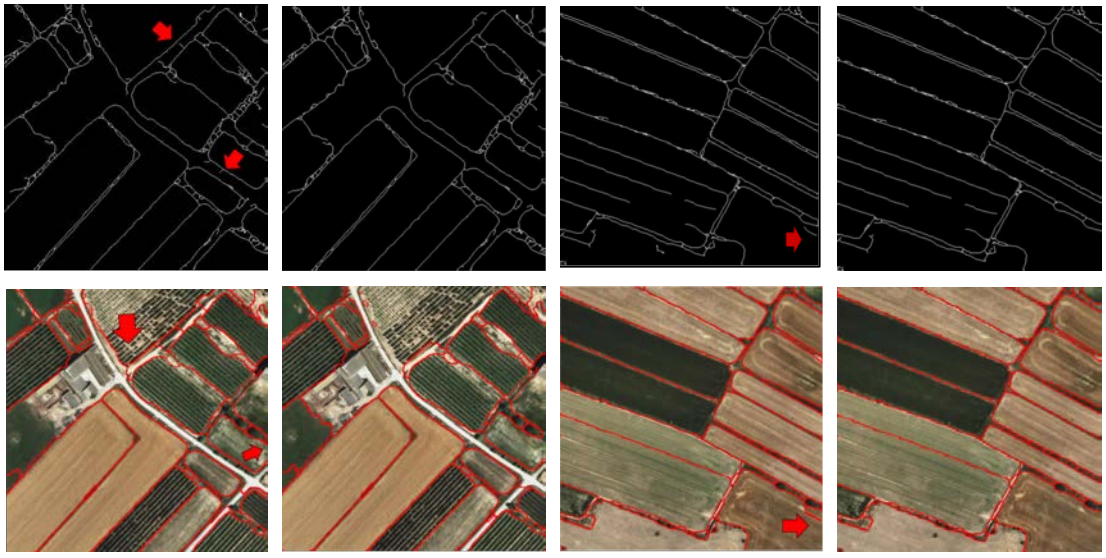
In order to analyze their impact on the performance results, we tested different values for them. Table 5.2 shows these values and the measures obtained for each parameter configuration. As can be observed, DeepNEM is a robust method that provides very competitive and stable results regardless of small changes in the parameters. Depending on the purpose of the segmentation, priority may be given to obtaining more Type A regions or having more regions with a distance of JD above 0.9. Taking into account the problem at hand, the best trade-off for all the metrics evaluated is achieved when the parameter configuration is: $A_{min} = 40$, $T_{min} = 8$ and $Add_{max} = 20$.

Finally, Figures 5.12 and 5.13 illustrate the impact of these parameters by means of some representative examples. Firstly, Figure 5.12 depicts the effects of the parameters A_{min} and T_{min}. As can be observed, the lower they are,

A_{min}	T_{min}	Add_{max}	average JD	No. of regions				
				Type A	Type B	Type C	JD $\geq$ 0.9	JD $<$ 0.7
40	8	10	0.9032 ± 0.0277	**2087**	**441**	327	1745	458
		15	0.9044 ± 0.0267	2068	504	290	1766	451
		20	0.9047 ± 0.0265	2030	570	273	1790	440
		25	0.9044 ± 0.0266	2021	600	266	1797	426
		30	0.9049 ± 0.0263	2015	623	**251**	**1804**	**424**
40	6	20	0.9047 ± 0.0267	2021	581	273	1802	438
	8		0.9047 ± 0.0265	2030	570	273	1790	440
	10		$\mathbf{0.9050 \pm 0.0263}$	2019	562	287	1791	445
	12		0.9045 ± 0.0265	2003	547	300	1775	463
	14		0.9045 ± 0.0265	2003	547	300	1775	463
20	8	20	0.9049 ± 0.0264	2024	595	278	1803	436
30			0.9047 ± 0.0265	2011	584	281	1794	437
40			0.9047 ± 0.0265	2030	570	273	1790	440
50			0.9046 ± 0.0264	2035	565	272	1790	441
60			0.9047 ± 0.0263	2040	556	274	1791	443

TABLE 5.2: Results obtained by DeepNEM using different values for the three parameters: A_{min}, T_{min}, and Add_{max}. Blanks correspond to the same value that heads the column.

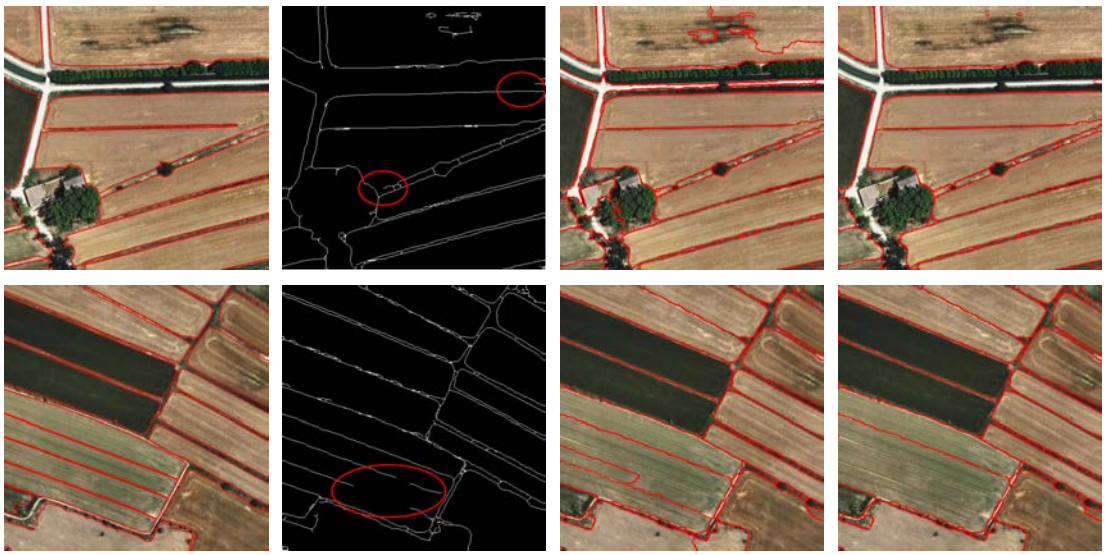
the more details are kept and the more likely the results are to present over-segmentation.



(a) (b) (c) (d)

FIGURE 5.12: Output edges obtained after the edge map construction (top), and their corresponding regions (bottom). (a,b) Results obtained with parameter $A_{min} = 40$ and different T_{min} values: (a) $T_{min} = 6$ allows to keep more detailed segments and to complete the two boundaries pointed out by the arrows (JD = 0.8950, Type A = 15), (b) $T_{min} = 14$ provides less detailed segments and so these two boundaries are lost (JD = 0.8937, Type A = 14). (c,d) Results obtained with parameter $T_{min} = 8$ and different A_{min} values: (c) $A_{min} = 20$ generates more division inside fields (JD = 0.9053, Type A = 12), (d) $A_{min} = 60$ deletes pixel areas (JD = 0.9051, Type A = 13). Since the minimization process acts globally, some details that are kept depending on the parameters will affect other neighboring fields, as it is pointed out with the red arrows in the final results (a). Some parts of fields are completed meanwhile these regions will not be recovered without these clues (c).

Regardless of the established parameters, it is difficult to recover human interpretation. DeepNEM can reinforce this division or just ignore it, because there are not enough clues for it. This fact is recovered by fitting the results to a model, as it is shown in Figure 5.13 (top). The opposite phenomenon is also shown in Figure 5.13 (bottom): the division, due to minimization process, delivers a segmentation coincident with the GT in terms of region number, but their boundaries are different since the process is driven by radiometric information. The final model constraint will erase this division because the number of boundary pixels added is greater than a Add_{max}.



(a) (b) (c) (d)

FIGURE 5.13: Examples of over-segmentation (top) and under-segmentation (bottom): (a) input images with their GT, (b) HED results, (c) DeepNEM results after energy-minimization, and (d) final DeepNEM results. (Top) The JD is similar in (c) and (d), 0.8860 and 0.8853, respectively; whereas the number of Type A regions increases from 7 to 8. (Bottom) The JD improves from (c) to (d), 0.8947 and 0.9051; while the number of Type C regions increases from 0 to 3, because the three fields highlighted (red circle) are recovered as only one.

Comparison to the state-of-the-art

We compared DeepNEM with three state-of-the-art methods for aerial segmentation: 1) HED [39], in order to reveal the improvement achieved by adding the proposed energy-minimization framework; 2) eCognition [77], the commercial software most widely used in remote sensing field; and 3) BDCN [84], one of the most recent frameworks for edge detection and segmentation. We applied DeepNEM as well as these three approaches to the 280 test images of the dataset described in Section 5.4.1. Regarding the parameter settings of DeepNEM, we used the most competitive ones according to the experimentation presented in Section 5.4.5: $A_{min} = 40$, $T_{min} = 8$ and $Add_{max} = 20$. With respect to the BDCN network, it was trained from scratch using an initial learning rate of 10^{-6}, which was reduced by 10 times each 10000 epochs. Other configuration parameters include a momentum of 0.9, and weight decay of 2^{-4}.

Table 5.3 includes the results of the four methods considered (HED, eCognition, BDCN, and DeepNEM) in terms of the average JD calculated over the 280 test images. As can be seen, the worst results are obtained with eCognition; whilst the best results are achieved when using DeepNEM, demonstrating the

adequacy of the energy-minimization process applied to the edges provided by HED. The results achieved by DeepNEM are not only better on average (mean), but also have a lower standard deviation.

Figure 5.14 shows the results obtained with the four different methods in four intervals of JD: from greater than or equal to 0.9, to lower than 0.7. As can be observed, DeepNEM obtains the best results by detecting 1790 regions with a JD $>=$ 0.9, which represents 54.03% of all the fields detected. This method is followed by HED, demonstrating once again the adequacy of using DNs in the problem at hand, and providing a better performance than the commercial software eCognition and the novel BDCN, which has a lower performance in this case. Analyzing the fields with a low JD, it should be highlighted that only 13.28% of the regions detected by DeepNEM have a JD $<$ 0.7.

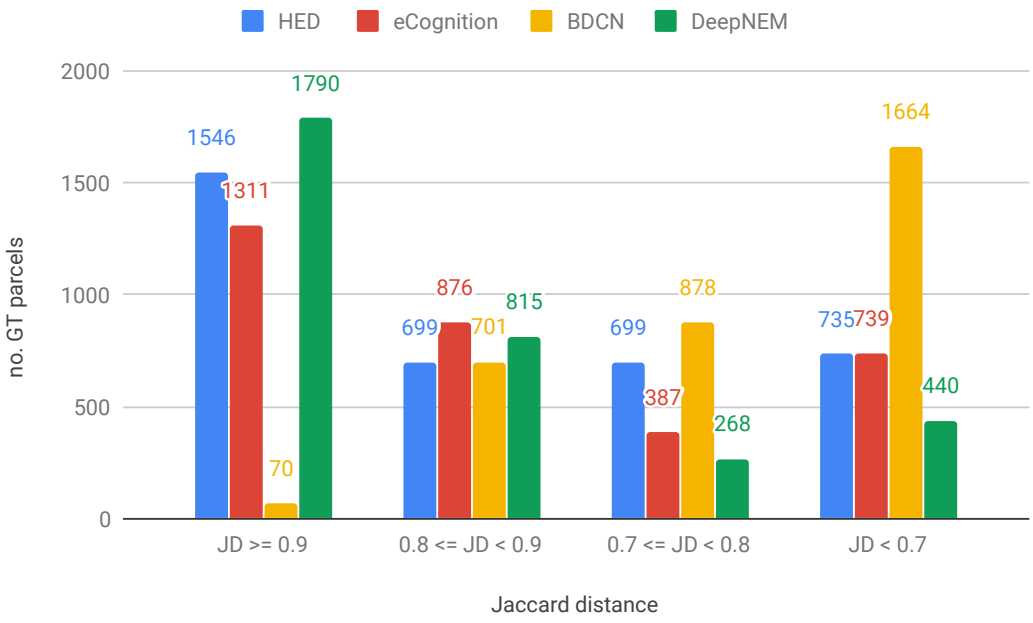


FIGURE 5.14: Performance of the different methods in terms of the number of ground truth fields segmented, for different values of JD.

Table 5.3 also shows the results in terms of the different types of regions (A, B, and C). Regarding the fields successfully recovered (type A), the numbers obtained when applying eCognition are noticeably lower than when using DNs, showing that eCognition tends to produce over- or under-segmentation. HED achieves the best results in terms of over-segmentation (type B), with only 63 regions showing that it tends to produce under-segmentation. On the other hand, BDCN provides the best results in terms of under-segmentation (type C), with only 112 regions. In fact, BDCN shows the best balance between type B and type C regions. However, the number of fields 1 to 1 is

	HED	eCognition	BDCN	DeepNEM
average JD	0.8983 ± 0.0367	0.8870 ± 0.0324	0.7916 ± 0.0892	$\mathbf{0.9047 \pm 0.0265}$
No. of type A regions (1 to 1)	1759	1301	1305	**2030**
No. of type B regions (over-segmentation)	**63**	795	232	570
No. of type C regions (under-segmentation)	756	478	**112**	273
Covering	0.590	0.609	0.589	**0.782**
Rank index	0.880	0.849	**0.888**	0.874
Variation of information	0.442	0.530	**0.387**	0.474
Recall	0.571	0.654	0.490	**0.679**
Precision	0.543	0.563	0.491	**0.581**
F-measure	0.557	0.604	0.491	**0.626**

TABLE 5.3: Performance measures obtained when applying the four different methods to the 280 test images. The average Jaccard distance is in terms of mean $\pm$ standard deviation, calculated across all the test images.

quite low, similar to the one achieved by eCognition. The most competitive results in terms of 1 to 1 regions (type A) are achieved by DeepNEM, which also provides a good trade-off with respect to the number of over- and under-segmented fields (types B and C, respectively). In order to illustrate the different types of fields associated to aerial image segmentation, Figure 5.15 uses different colors to show how DeepNEM works on three sample images. As can be seen, almost all the regions are extracted one to one with a JD ≥ 0.9.

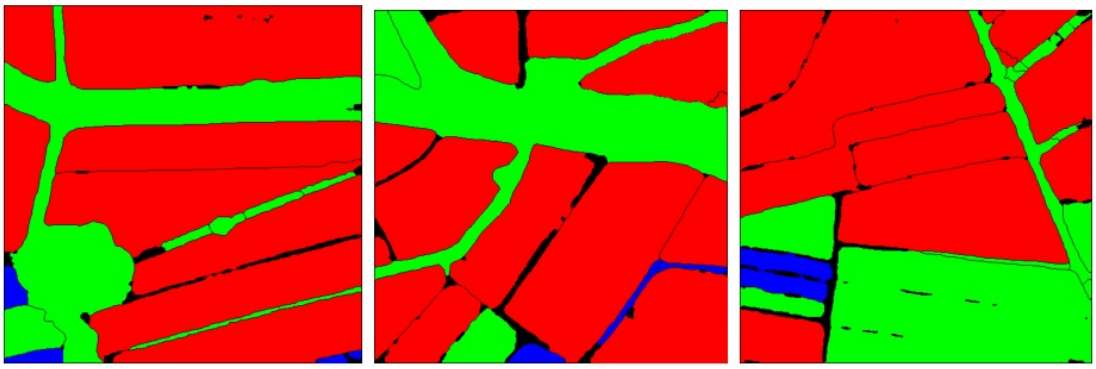


FIGURE 5.15: A graphical representation of JD for three of the thirteen images illustrated in Figure 5.7: JD $\geq$ 0.9 (green), 0.8 $\leq$ JD $<$ 0.9 (red), and 0.7 $\leq$ JD $<$ 0.8 (blue).

Regarding the region- and boundary-based metrics, they are also reported in Table 5.3. As can be observed, DeepNEM outperforms the other three methods in four out of the six measures (covering, precision, recall, and f-measure), followed by eCognition. With respect to the other two metrics (rank index and variation of information), the best results are provided by BDCN. However, it is worth noting that there are no significant differences in terms of the rank index, with very similar results achieved regardless the method considered.

Figure 5.7 depicts thirteen images for a qualitative comparison: Figure 5.7(a) shows the original images with their corresponding GT stacked, Figure 5.7(b) shows the results obtained with HED (using the fusion layer as output), Figure 5.7(c) shows the results obtained with eCognition, Figure 5.7(d) shows the results obtained with BDCN, and Figure 5.7(e) shows the results obtained with DeepNEM. As can be observed, DeepNEM keeps the boundaries more clearly delineated. On the other hand, there are more regions over-segmented with eCognition than with DeepNEM. This over-segmentation is due to radiometric variability inside each field, which is relevant enough to be recovered by a software that relies on radiometry. On the other hand, DeepNEM relies more on linear elements, since it has been trained to find these clues, and introduced constraints to keep these elements in the energy-minimization process.

Note that eCognition is used in productive environments as a classification-segmentation software. So, it is very important to evaluate results in terms of overlapping areas, as well as to analyze the way in which areas are recovered: under- and over-segmentation. This fact determines the amount of manual edition necessary to obtain a final product. Note that DeepNEM reduces by a quarter the necessary edition obtaining much less under- and over-segmented images.

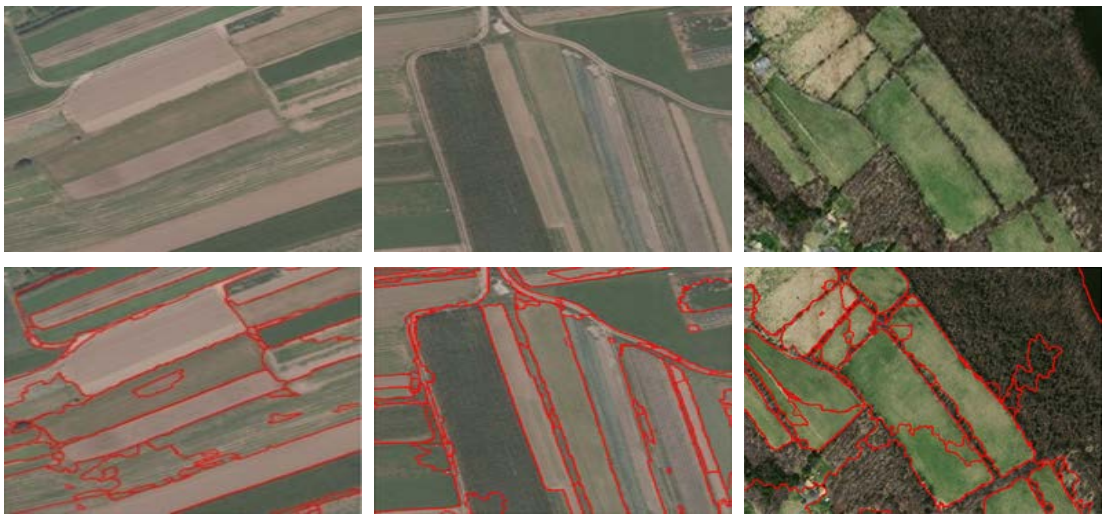
Aerial datasets

As far as the authors know, there is no public dataset for aerial segmentation. However, we have found two aerial datasets for object detection (instead of field segmentation): AIRS [96] and Massachusetts Buildings [97]. Despite this, it is possible to identify some agricultural regions among their images. We used them in order to check how robust is our algorithm run on different datasets.

From these datasets, we selected the images that contain fields, and ran our DeepNEM. The results are shown in Figure 5.16. Due to the fact that their resolution is smaller than that associated with our training dataset, DeepNEM has to rely more on constraints associated with the model than on the edges extracted by the HED. For this reason, the results are slightly over-segmented in fields that are highly textured. Anyway, the improvement of the results over these images makes necessary a GT associated to these agricultural fields, and train the DN with them.

5.5 Conclusions

We present a joint venture between a deep network and an energy-minimization model-guided radiometric method that improves the benefits of each component. The two-step process we proposed, represented by DeepNEM, has been trained and tested over a new public aerial dataset of 1200 images. The contours delivered by our DeepNEM are really close to the GT both in area and shape. Furthermore, it is possible to take advantage of the by-products in order to trigger other semi-automatic segmentation processes, as we have demonstrated in the validation section. The two step process can improve as other networks deliver better edges. DeepNEM has been tested over a variety of natural areas and compared with other region extraction algorithms. This has demonstrated that DeepNEM eliminates the need for human interaction



(a) (b) (c)

FIGURE 5.16: Original images (top) and DeepNEM results (bottom): (a,b) original images from the AIRS dataset, and (c) original image from the Massachusetts dataset.

and obtains smoother and more reliable results. When the image is a continuum of regions, DeepNEM will pull them apart, whether or not there is any evidence of border reliable enough. Moreover, if inside the fields there are some trees or bushes, which are not large enough to become an isolated entity, the process will not consider them.

The whole process has been published in the paper:

- Margarita Torre, Beatriz Remeseiro, Petia Radeva and Fernando Martínez. DeepNEM: Deep Network Energy-Minimization for Agricultural Field Segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 726-737, 2020.

Chapter 6

Towards DEM construction from SAR images

One of the great problems of monitoring the movements of a given area of land, what we call subsidence, is the great economic and human effort necessary to establish and regularly measure a series of control points on it.

If the area to be studied is very large, the problem can be multiplied by several orders of magnitude, and even more, if we do not have the absolute certainty that deformations of the land are taking place, this campaign of measures will surely never be carried out.

Taking into account all these limitations and with the need to control these movements due to the great problems that their lack of control can cause, it can be concluded that an automatic system capable of generating deformation maps without the need to physically access would be of great interest to the area under study. With this system, it would be possible to periodically monitor large areas for risk control at a much lower cost than necessary using field measures.

At the Institut Cartogràfic i Geológic de Catalunya we developped an approach to manage a new way of observing these phenomens known as SAR Differential Interferometry (DInSAR) [98]. This technique, by means of the use of radar images of the reflectivity of the terrain acquired by satellite, is capable of generating maps of deformation of the terrain with an accuracy of millimetres.

6.1 Differential interferometry (DInSAR)

DInSAR techniques consist of the combination of two SAR images of the same area acquired from slightly different positions, as can be seen in Figure 2.4. The result of this combination is a new image known as an interferogram, whose phase variation between neighboring pixels con be expressed as

$$\delta\phi_{\text{TopoDInSAR}} = \delta\phi_{\text{mov}} + \delta\phi_{\text{topo}} + \delta\phi_{\text{atm}} + \delta\phi_{\text{noise}} \tag{6.1}$$

where $\delta\phi_{\text{topo}}$ is the topographic phase, $\delta\phi_{\text{mov}}$ is the component due to the displacement of the terrain in range direction [line of sight (LOS)] between both SAR acquisitions, $\delta\phi_{\text{atm}}$ is the phase related with atmospheric artifacts, and $\delta\phi_{\text{noise}}$ comprises degradation factors related with temporal and spatial decorrelation and thermal noise.

If a set of TopoDInSAR interferograms of the same area is used, a model, which considers a linear velocity deformation and topography, can be fitted to the stack of interferograms with different spatial and temporal baselines.

The TopoDInSAR model cannot be applied to all the pixels within the area under study, since only a part of them have the sufficient phase quality due to decorrelation. If short temporal baseline interferograms are used, the percentage of useful pixels will be very high and enough for topographic purposes. In most cases, the deformation term will be very low in comparison with the topographic one. Nevertheless, if terrain movement is very strong, interferograms with short temporal baselines will present deformation fringes, and the movement term must be computed for precise estimation of topography.

In addition, to obtain soil deformation measures we will have to cancel or minimize the effects of unwanted components, which will be topography, atmospheric effects and thermal noise. When working with classic DInSAR techniques, the main problem lies in the presence of atmospheric artefacts, difficult to eliminate using a single interferometric pair. However, the term related to the topography of the land may be canceled with the help of a Land Elevation Map (MET) and the orbital parameters of the acquisitions.

In any case, both the impossibility of eliminating the atmospheric component and the inaccuracies of the MET will largely determine the precision obtained in the measurement of the movement of the terrain. For this reason, advanced techniques such as the one presented below are necessary.

6.2 DISICC development

The DISICC software package has been created to overcome the intrinsic limitations of the classic Differential Interferometry. His operation is based not on creating a only interferogram (two SAR images), but in generating a set of interferometric pairs with images taken in different dates. With this preamble

we get a redundancy of the data obtained that will allow the minimization of errors topographic and atmospheric artifacts.

The technique in three steps:

1. Selection of those pixels of the image that present a good quality for the measurement of soil deformation. This pixel selection uses information from coherence of the set of interferograms available through a threshold selected by the user. This is necessary since depending on the type of land (urban, wooded, desert,...) the quality of the interferometric phase will vary considerably. For example, urban land usually provides great signal quality even for peers interferometric with temporal separations of various years. On the contrary, wooded soils they can lose the quality of measurement in pairs separated only a few days.

2. Surface triangulation, where each vertex of the triangles corresponds to one of the pixels selected in the previous stage. DISICC calculates the gradients of deformation of the land for each edge of the triangulation network, since working on this form the interferometric phase (in this case phase increments) presents a higher quality.

   As the topography and linear velocity are constant in the whole set of differential interferograms, it is possible to retrieve a good estimation of them, adjusting the following phase model to data

$$\begin{aligned}\delta\phi_{\text{model}}(x_{\text{m}}, y_{\text{m}}, x_{\text{n}}, y_{\text{n}}, T_i) = \frac{4\pi}{\lambda} T_i \left[v_{\text{model}}(x_{\text{m}}, y_{\text{m}}) - v_{\text{model}}(x_{\text{n}}, y_{\text{n}})\right] \\ + \frac{4\pi}{\lambda} \frac{b(T_i)}{r(T_i)\sin(\theta_i)} \left[h_{\text{model}}(x_{\text{m}}, y_{\text{m}}) - h_{\text{model}}(x_{\text{n}}, y_{\text{n}})\right] \quad (6.2)\end{aligned}$$

3. Once the gradients of the movement have been calculated, it is possible to calculate the absolute speed of each pixel by the integration of the increments obtained from the adjustment of Equation (6.2). After this process, the DISICC software returns to the user the map of average ground speed for the temporal interval between the first and the last SAR image used in the study.

6.3 Results

Sallent, in the Bages region, is a very interesting area since one of its neighborhoods (the Barri de l'Estació) is affected by strong subsidences of the land that

have occurred in recent years. In addition field measurements are available through precise leveling.

As can be seen in Figure 6.1 this comparison allows us to observe how the process deformation has accelerated in recent years (with a maximum of -2 cm / year before 1999 and -4 cm/year for 2004). It proves how the process was fully active at that time. Likewise, it was shown as the pattern of deformation in the Barri de l'Estació had a perfect correspondence between the result obtained by the DISICC and the field measurements. So, it confirms the optimal operation of the ICC software.

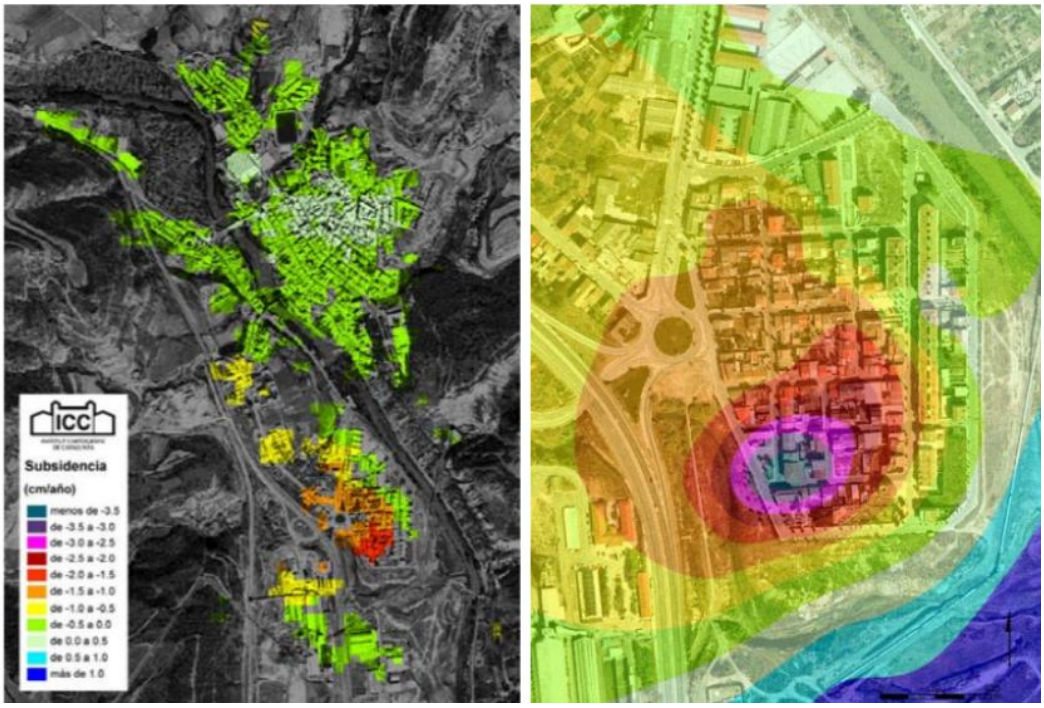


FIGURE 6.1: Left: Deformation of the land (cm/year) in Sallent. ERS data from november 1992 to december 1999. Right: Deformation measures on the Sallent Station neighborhood during 2004

Chapter 7

Conclusions

In an environment where images proliferate every day, and in view of the growing need to extract quality information from them, the tools that facilitate this transformation of data into semantic information are increasingly necessary. In addition, we must take into account the growing importance that GIS are acquiring to locate and plan a wide range of actions, both at the state and private levels. Among all the geographical elements we highlight two: the transportation network and the fields. Both, together with buildings are the what need more human operation time to extract and delineate from aerial images. In addition, their applications exceed the scope of the mere representation on a topographic map. Roads and paths are the reference system to follow for human mobility as well as goods delivery. In the case of fields, their exact delineation is decisive for the cadastral map, as well as for the records of subsidies for agricultural production policies. Among the automation approaches analyzed, those that need a thorough review by operators cannot be reliably introduced into a production environment. Human operators, if they do not have a guarantee of total success, prefer to have control of the digitization tools.

Initially, we presented a new methodology for semi-automatic boundary fields extraction. The approximation is based on a generalization of region growing techniques combined with deformable models, namely region competition. In a first approximation, the seed provided by the operator inside the field was made to grow by the radiometric information of the image. The incorporation of the high contrast lines of the images –edges– improves the results. Deformable models serve to introduce knowledge of what the contours of the fields look like. The concept of piece-wise smoothness incorporated in a B-spline model provided us with the required softness between the points of sharp curvature. We tested the approach in aerial photographs and orthophotos, and 80% of fields were properly extracted.

As the radiometry of the roads seemed to be a good element to help differentiate them from their surroundings, we applied a variation of the region competition algorithm to the roadsides. Our proposal was a path optimizer because the role of the human operator is focused on the provision of significant seeds and further results supervision. In this case the model is able to recover the centerline and the roadsides. Therefore, it is important that the seed points be delivered in regions where changes in radiometry and in curvature occur. The information associated with the seeds defines the statistical parameters that drive the region competition process.

The majority of the roads that appear on aerial images can be extracted, because the algorithm is mainly based on their light appearance and on two characteristics: small changes in radiometry and in curvature. Some of the roads that cannot be completely extracted only need some minor edition tasks.

Since DL has recently provided image classifications with a very high rate of success in other domains, we decided to approach the aerial image analysis in an end-to-end framework in order to obtain the full automation of fields extraction. Instead of relying on the radiometric classification that often leads to over-segmentation, we propose a joint venture (DeepNEM) based on a deep network that extracts edges and an energy-minimization model-guided radiometric method that improves the benefits of each component. Although the deep network provides excellent results for image edge detection, it cannot ensure obtaining all the complete boundaries of fields. So, the edges obtained with HED are processed, categorized and completed within an energy-minimization framework to assure straightforward extraction of agricultural fields.

One of the main difficulties we faced was testing the different neural network algorithms in the absence of datasets that collect a broad range of fields and their associated GT. The proposal has been trained and tested over a new public aerial dataset of 1200 images. DeepNEM eliminates the need for human interaction and obtains reliable results.

7.1 Further research

Given the success obtained in automating the extraction of fields in aerial images, the next step would be to try to apply a similar approach for the automatic extraction of linear man-made objects.

The coupled approach of neural network and energy-minimization method leaves the door open to test further potential progress in networks to extract

edges [84] and regions. These improvements can result in reducing and reformulating the edge map construction at the beginning of the energy-minimization step. Moreover, it would also be suitable to analyze how to implement this step and the energy-minimization within an end-to-end framework.

Although the proposed approach has reached a high degree of success, there still remain some types of regions that have not been included in the training step. There is room for improvement in semi-supervised or even unsupervised segmentation to detect fields that have not been used in the training process using techniques such as novelty detection.

The dataset of images with which we have trained and tested our approach is of a certain image scale. There is room for improvement in range of scales to consider. Moreover, another possibility is a multi-scale system, so that the same network can tackle several scales at the same time. For example the scale-adaptive networks [99], which learn the convolutional parameters and scale coefficients in an end-to-end way. Another possibility also remains to analyze if some layers could be present or not, depending on the input image scale. A further line of research is how to introduce other ways of carrying out context analysis, in order to disambiguate the interpretation of the agricultural fields that can only be recognized by contextual reasoning.

Bibliography

[1] M. Diaz, V. B. Manian, and R. Vásquez, "Wavelet features for color image classification", in *Imaging and Geospatial Information Society Annual Conference*, 2000.

[2] L. Lucchese and S. K. Mitra, "Filtering color images in the xyy color space", in *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, IEEE, vol. 3, 2000, pp. 500–503.

[3] O. Heyman, "Automatic extraction of natural objects from 1 m remote sensing images", in *IUCGIS Summer Assembly*, 2001.

[4] *ERDAS IMAGINE*, https://www.hexagongeospatial.com/products/power-portfolio/erdas-imagine/erdas-imagine-remote-sensing-software-package/, Last reviewed: March 2020.

[5] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[6] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition", *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[7] D. Comaniciu and P. Meer, "Mean shift analysis and applications", in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, 1999, pp. 1197–1203.

[8] M. Haggag, M. Zahran, and M. Salah, "Towards automated generation of true orthoimages for urban areas", *American Journal of Geographic Information System*, vol. 7, no. 2, pp. 67–74, 2018.

[9] W. Förstner, "A framework for low level feature extraction", in *European Conference on Computer Vision*, Springer, 1994, pp. 383–394.

[10] *Inpho-Trimble*, https://geospatial.trimble.com/products-and-solutions/inpho/, Last reviewed: March 2020.

[11] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets", *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[12] S. P. Lloyd, "Least square quantization in pcm. bell telephone laboratories paper. published in journal much later: Lloyd, sp: Least squares quantization in pcm", *IEEE Transactions on Information Theory (1957/1982)*, vol. 18, 1957.

[13] J. MacQueen *et al.*, "Some methods for classification and analysis of multivariate observations", in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, 1967, pp. 281–297.

[14] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm", *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.

[15] C. Yang, R. Duraiswami, D. DeMenthon, and L. Davis, "Mean-shift analysis using quasinewton methods", in *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)*, IEEE, vol. 2, 2003, pp. II–447.

[16] P. Doucette, P. Agouris, A. Stefanidis, and M. Musavi, "Self-organised clustering for road extraction in classified imagery", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 55, no. 5-6, pp. 347–358, 2001.

[17] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers", in *Proceedings of the fifth annual workshop on Computational learning theory*, 1992, pp. 144–152.

[18] N. Yager and A. Sowmya, "Support vector machines for road extraction from remotely sensed images", in *International Conference on Computer Analysis of Images and Patterns*, Springer, 2003, pp. 285–292.

[19] J. Canny, "A computational approach to edge detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[20] L. Bruzzone and D. F. Prieto, "Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 2, pp. 456–460, 2001.

[21] J. Zhou, W. F. Bischof, and T. Caelli, "Road tracking in aerial images based on human–computer interaction and bayesian filtering", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 2, pp. 108–124, 2006.

[22] Z. Hong, D. Ming, K. Zhou, Y. Guo, and T. Lu, "Road extraction from a high spatial resolution remote sensing image based on richer convolutional features", *IEEE Access*, vol. 6, pp. 46 988–47 000, 2018.

[23] W. Wang, N. Yang, Y. Zhang, F. Wang, T. Cao, and P. Eklund, "A review of road extraction from remote sensing images", *Journal of Traffic and Transportation Engineering (english edition)*, vol. 3, no. 3, pp. 271–282, 2016.

[24] V. P. Desai and H. Vala, "Survey on methods of road extraction using satellite image", *International Journal of Engineering Research and Technology*, vol. 3, no. 11, pp. 1422–1424, 2014.

[25] C. Y. Wang, "Edge detection using template matching (image processing, threshold logic, analysis, filters)", *Duke University*, vol. 288, 1985.

[26] T. Nagao, T. Agui, and M. Nakajima, "Automatic extraction of roads denoted by parallel lines from 1/25,000-scaled maps utilizing skip-scan method", *Systems and Computers in Japan*, vol. 21, no. 11, pp. 96–105, 1990.

[27] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models", *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[28] I. Laptev, H. Mayer, T. Lindeberg, W. Eckstein, C. Steger, and A. Baumgartner, "Automatic extraction of roads from aerial images based on scale space and snakes", *Machine Vision and Applications*, vol. 12, no. 1, pp. 23–31, 2000.

[29] J. Wang and M. F. Cohen, "An iterative optimization approach for unified image segmentation and matting", in *Tenth IEEE International Conference on Computer Vision*, IEEE, vol. 2, 2005, pp. 936–943.

[30] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines and curves in pictures", *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[31] X. Huang and L. Zhang, "Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines", *International Journal of Remote Sensing*, vol. 30, no. 8, pp. 1977–1987, 2009.

[32] Y. LeCun, Y. Bengio, *et al.*, "Convolutional networks for images, speech, and time series", *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[33] *Matlab*, http://www.mathworks.com/, Last reviewed: March 2020.

[34] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[35] S. Piramanayagam, E. Saber, W. Schwartzkopf, and F. W. Koehler, "Supervised classification of multisensor remotely sensed images using a deep learning framework", *Remote Sensing*, vol. 10, no. 9, p. 1429, 2018.

[36] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with convolutional neural networks", *Electronic Imaging*, vol. 2016, no. 10, pp. 1–9, 2016.

[37] G. Cheng, Y. Wang, S. Xu, H. Wang, S. Xiang, and C. Pan, "Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 6, pp. 3322–3337, 2017.

[38] S. E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture", in *Advances in neural information processing systems*, 1990, pp. 524–532.

[39] S. Xie and Z. Tu, "Holistically-nested edge detection", *International Journal of Computer Vision*, vol. 125, no. 1-3, p. 3, 2017.

[40] G. H. Ball and D. J. Hall, "Isodata, a novel method of data analysis and pattern classification", Stanford research inst Menlo Park CA, Tech. Rep., 1965.

[41] Y. Choung, "Mapping levees using lidar data and multispectral orthoimages in the nakdong river basins, south korea", *Remote Sensing*, vol. 6, no. 9, pp. 8696–8717, 2014.

[42] S. Chen and D. Zhang, "Robust image segmentation using fcm with spatial constraints based on new kernel-induced distance measure", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 4, pp. 1907–1916, 2004.

[43] T. S. Lee, "Image representation using 2d gabor wavelets", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, 1996.

[44] M. Schaale, "Land cover texture information extraction from remote sensing image data", in *Proceedings of the ASPRS-RTI annual conference, 2000*, 2000.

[45] M. M. Alemu, "Automated Farm Field Delineation and Crop Row Detection from Satellite Images", PhD thesis, University of Twente, 2016.

[46] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, "Lsd: A line segment detector", *Image Processing On Line*, vol. 2, pp. 35–55, 2012.

[47] M. P.-D. Isabelle Cléry and B. Vallet, "Automatic georeferencing of a heritage of old analog aerial photographs.", *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 2, no. 3, 2014.

[48] A. K. Sinop and L. Grady, "A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm", in *IEEE International Conference on Computer Vision*, 2007, pp. 1–8.

[49] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient nd image segmentation", *International Journal of Computer Vision*, vol. 70, no. 2, pp. 109–131, 2006.

[50] L. Grady, "Random walks for image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 11, pp. 1768–1783, 2006.

[51] E. W. Dijkstra *et al.*, "A note on two problems in connexion with graphs", *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[52] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks", *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 10, pp. 1797–1801, 2014.

[53] T.-H. Hong and A. Rosenfeld, "Compact region extraction using weighted pixel linking in a pyramid", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 222–229, 1984.

[54] S. C. Zhu and A. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 9, pp. 884–900, 1996.

[55] C. H. van Kemenade, H. La Poutre, and R. J. Mokken, "Density-based unsupervised classification for remote sensing", in *Machine vision and advanced image processing in remote sensing*, Springer, 1999, pp. 248–258.

[56] S. Dellepiane, "Detail-preserving processing of remote sensing images", in *Machine Vision and Advanced Image Processing in Remote Sensing*, Springer, 1999, pp. 23–36.

[57] R. Adams and L. Bischof, "Seeded region growing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 6, pp. 641–647, 1994.

[58] J. Rissanen, "Modeling by shortest data description", *Automatica*, vol. 14, no. 5, pp. 465–471, 1978.

[59] A. Ruiz and M. Torre, "Experiences with match_t for orthophoto production", in *Procceedings of OEEPE-Workshop on Application of Digital Photograrnmetric Workstations.1996*, 1996.

[60] E Gülch, "Application of semi-automatic building acquisition, automatic extraction of man-made objects from aerial and space images (ii)", in *Proceedings of the Ascona Workshop*, 1997.

[61] A. Gruen, *Automatic extraction of man-made objects from aerial and space images (II)*. Springer Science & Business Media, 1997.

[62] M. Torre and P. Radeva, "Agricultural field extraction from aerial images using a region competition algorithm", *International Archives of Photogrammetry and Remote Sensing*, vol. 33, no. B3/2; PART 3, pp. 889–896, 2000.

[63] R. H. Bartels, J. C. Beatty, and B. A. Barsky, *An introduction to splines for use in computer graphics and geometric modeling*. Morgan Kaufmann, 1995.

[64] T. Ohlhof, E. Gulch, H. Muller, C. Wiedemann, and M. Torre, "Semi-automatic extraction of line and area features from aerial and satellite images", *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 34, no. Part XXX, 2004.

[65] A. Gruen and H. Li, "Semi-automatic linear feature extraction by dynamic programming and lsb-snakes", *Photogrammetric Engineering and Remote Sensing*, vol. 63, no. 8, pp. 985–994, 1997.

[66] P. Radeva, A. Solé, A. M. López, and J. Serrat, "Nets of linear structures in satellite images", in *Machine Vision and Advanced Image Processing in Remote Sensing*, Springer, 1999, pp. 304–316.

[67] D. M. McKeown and J. L. Denlinger, "Cooperative methods for road tracking in aerial imagery", in *Proceedings CVPR'88: The Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 1988, pp. 662–672.

[68] A. Baumgartner, S. Hinz, and C. Wiedemann, "Efficient methods and interfaces for road tracking", *International Archives of Photogrammetry Remote Sensing and Spatial Information Sciences*, vol. 34, no. 3/B, pp. 28–31, 2002.

[69] C. Zhang, E. Baltsavias, and A. Grün, *Updating of cartographic road databases by image analysis*. Balkema Publishers, Lisse, The Netherlands, 2001.

[70] A. Baumgartner, C. Steger, H. Mayer, W. Eckstein, and H. Ebner, "Automatic road extraction in rural areas", *International Archives of Photogrammetry and Remote Sensing*, vol. 32, no. 3; SECT 2W5, pp. 107–112, 1999.

[71] L. D. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-d and 3-d images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1131–1147, 1993.

[72] W. M. Neuenschwander, P. Fua, G. Székely, and O. Kübler, "From ziplock snakes to velcro^{TM} surfaces", in *Automatic extraction of man-made objects from aerial and space images*, Springer, 1995, pp. 105–114.

[73] R. Toledo, X. Orriols, P. Radeva, X. Binefa, J. Vitria, C. Canero, and J. Villanuev, "Eigensnakes for vessel segmentation in angiography", in *Proceedings 15th International Conference on Pattern Recognition*, IEEE, vol. 4, 2000, pp. 340–343.

[74] R. Alamús, W. Kornus, M. Pla, and J. Talaya, "Evaluación de la precisión en tres años de campañas con cámara digital en el ICC.",

[75] M. D. Hossain and D. Chen, "Segmentation for object-based image analysis (obia): A review of algorithms and challenges from remote sensing perspective", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 115–134, 2019.

[76] *eCognition | Trimble Geospatial*, http://www.ecognition.com/, Last reviewed: March 2020.

[77] Y. Zhang and T. Maxwell, "A fuzzy logic approach to supervised segmentation for object-oriented classification", in *ASPRS 2006 Annual Conference.*, 2006, pp. 1–5.

[78] S. Bhardwaj and A. Mittal, "A survey on various edge detector techniques", *Procedia Technology*, vol. 4, pp. 220–226, 2012.

[79] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, p. 436, 2015.

[80] S. Xie and Z. Tu, "Holistically-Nested Edge Detection", in *IEEE International Conference on Computer Vision*, 2015, pp. 1395–1403.

[81] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", in *International Conference on Learning Representations*, 2015.

[82] Y. Liu and M. S. Lew, "Learning relaxed deep supervision for better edge detection", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 231–240.

[83] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3000–3009.

[84] J. He, S. Zhang, M. Yang, Y. Shan, and T. Huang, "Bi-directional cascade network for perceptual edge detection", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3828–3837.

[85] A. Troya-Galvis, P. Gançarski, and L. Berti-Équille, "Remote sensing image analysis by aggregation of segmentation-classification collaborative agents", *Pattern Recognition*, vol. 73, pp. 259–274, 2018.

[86] O. Csillik, "Fast segmentation and classification of very high resolution remote sensing data using slic superpixels", *Remote Sensing*, vol. 9, no. 3, p. 243, 2017.

[87] H. Gu, Y. Han, Y. Yang, H. Li, Z. Liu, U. Soergel, T. Blaschke, and S. Cui, "An efficient parallel multi-scale segmentation method for remote sensing imagery", *Remote Sensing*, vol. 10, no. 4, p. 590, 2018.

[88] L. Mou, Y. Hua, and X. X. Zhu, "A relation-augmented fully convolutional network for semantic segmentation in aerial scenes", in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 416–12 425.

[89] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu, "Deeply-supervised nets", in *Artificial Intelligence and Statistics*, 2015, pp. 562–570.

[90] *AgriculturalField-Seg DataSet*, https://www.aic.uniovi.es/bremeseiro/agriculturalfield-seg/, Last reviewed: March 2020.

[91] *AgriculturalField-Seg DataSet*, https://mat-web.upc.edu/people/fernando.martinez/dataset_af-seg.html, Last reviewed: March 2020.

[92] *Institut Cartogràfic i Geològic de Catalunya. Last reviewed: March 2020*, https://icgc.cat/, Last reviewed: March 2020.

[93] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 898–916, 2011.

[94] G. Ciocca, P. Napoletano, and R. Schettini, "Food Recognition: A New Dataset, Experiments, and Results", *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 3, pp. 588–598, 2017.

[95] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power watershed: A unifying graph-based optimization framework", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 7, pp. 1384–1399, 2011.

[96] Q. Chen, L. Wang, Y. Wu, G. Wu, Z. Guo, and S. L. Waslander, "Aerial imagery for roof segmentation: A large-scale dataset towards automatic mapping of buildings", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 147, pp. 42–55, 2019.

[97] V. Mnih, "Machine Learning for Aerial Image Labeling", PhD thesis, University of Toronto, 2013.

[98] R. Arbiol, V. Palà, F. Pérez, M. Castillo, and M. Crosetto, "Aplicaciones de la tecnología insar a la cartografía", in *IX Congreso Nacional de Teledetección*, 2001.

[99] R. Zhang, S. Tang, Y. Zhang, J. Li, and S. Yan, "Scale-adaptive convolutions for scene parsing", in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2031–2039.