THE ROLE OF EDGE OBJECTS IN FULL AUTONOMOUS IMAGE INTERPRETATION

R. de Kok*, P.Wezyk**, M. Weidenbach***

- *www.progea.pl,**Agricultural University of Krakow, ***www.landConsult.de
- roeland_de_kok@hotmail.com

KEY WORDS: Autonomous classification, crucial features, anchor objects, transferability

ABSTRACT:

The autonomous image interpretation of large data volumes is state of the art in medical image analysis. Nowadays image data in earth observations have become ubiquotus, the image analysis could profit from strategies derived from the medical domain. The traditional 'ways of thinking' from the era of pixel classification should be put in the background and it is time to imagine objects. Geobia is much more than applying standard classification methods on segments instead of pixels . In classical pixel classification the dominant role of sampling (in the image domain) and defining the spectral mean of a class (in the feature space domain) has been a well known practice but these procedures should not be dominant anymore in Geobia procedures. They might still have a role to play but a minor one in full autonomous image classification processes.
In this paper alternative strategies for sampling and class descriptions are based upon histogram extremes. This is further clarified in a case study, which intends to be overall transferable. The chosen case study can be applied on any arbitrary VHR subset including those from Google-Earth. And so the assumptions on stability and transferability of the proposed sequential classification process can be verified by everyone on it's chosen global region of interest. The predictable behavior of edge objects in the histogram extremes of crucial features is a key factor in the full autonomous definition of representative classes.

1. CHANGING ASSUMPTIONS

1.1 No more sampling

Sampling inside the image domain by a remote sensing specialist has been general practice for decades. This procedure is too cumbersome in the face of the overwhelming load of imaging data available today. Visual sampling is a preparation to define a mean and standard deviation in feature space. The latter can also be achieved by unsupervised classification algorithms, where cluster analysis defines a mean per class. Considering the class mean value is not a crucial part anymore in the Geobia classification. Note there is a huge difference between unsupervised classification and autonomous image interpretation. The fact is that clustering of both pixels as well as image-objects exists within the feature space and therefore can easily be statistically analyzed. This has also been a reason why pixel based classification techniques have been used widely in segment based analysis. However, the cause of clustering inside feature space is not directly transferable in an explanation on the content of the image domain. The full automatic definition of a class mean value in a statistical analysis in feature space is not an explicit description of object features in the image domain. Only these explicit descriptions are considered stable and transferable

Autonomously defining the mean spectral value in an imaging band of a standardized class (not a cluster) in feature space is extremely difficult, unstable and hardly transferable. For example a mean value for an image-object 'Building' can be found neither in the Red nor Infrared imaging bands. However, representative objects per class are a must have for autonomous classification. Representative objects have unique properties. In our example of the object 'Bright Buildings' it is their sharp contrasting edges. Only properties which are stable and transferable allow for candidate objects being representative in autonomous classification procedures. Some object properties are completely depending on solar angle, growth season, sensor, incident angle etc. Some characteristics are directly related to the object's spectral response. One of them is the well known ratio between NIR and Red light for the class "Vegetation". The NDVI is a feature, transferable through scale and relative insensitive for solar angle and sensor type. The upper extreme in the histogram (upper Quantile) of the NDVI band is the part where vegetation can be located if present inside the image during the growing season. Besides vegetation, more classes display unique histogram extremes in crucial features. These histogram extremes now offer an alternative to mean values of visual sample classes. Most important, histogram extremes are much easier to find automatically. The sequential protocol decisions on full automatic image classification must in the first place find representative areas autonomously.
Some of these crucial features have specific histogram extremes which are decisive to classify anchor objects. Anchor object replace visual samples and are unique representative image objects for their class and should be detectable by robotic vision algorithms. The relative position of anchor objects in feature space remains always the same, but not their absolute values. Take for example the NDVI. The 75th Quantile of the upper part of the NDVI histogramm always contains vegetation if present in the image. However values can range from 0.15 up to 0.95 and still we talk about the upper 25 Quantile of the feature NDVI. The Quantile describes the distribution on the amount of image objects, not their area. A large water body can occupy a 25% of the image area but could only make up the lower 2% Quantile of the NDVI feature.

1.2 Contrasting Edges

It is necessary to concentrate on the idea of "intuitive appealing" of image objects (Blaschke et.al, 2006.) This is typical for landcover classifications where the human eye identifies a certain group of pixels as a smooth area where neighboring pixels have similar values. This is very typical in an agricultural field and only one single segmentation step is needed to establish the pixel populations within these predefined borders. Intuitive appealing here refers to the

similarity of the human vision to group things that seem to belong together. Neigbouring pixels that are spectral similar depict homogeneous areas in such a way and have received (too ?) much emphasis in image segmentation based on homogeneous objects. The next intuitive visual step is the recognition of edges. The outlines/edges of buildings, highways and other infrastructure are visible in processed VHRS imagery and can be used to reconstruct these edge features. Although edge filtering has been around for decades in remote sensing analysis (Haberäcker, 1995) , the classification of edges seems to be an undiscovered territory.
Edges are an example par excellence for being intuitively appealing. The understanding of images by the human eye is defined largely by the correct imaginable grouping of edges. The strong connection of edges and the human visual interpretation can easily be demonstrated with edge manipulation in art as demonstrated in the graphics of M.C. Escher. Correct grouping of edges by an autonomous classification process is difficult, however not impossible.

2. ARBITRARY IMAGERY TO DEMONSTRATE PRINCIPLES

2.1 Crucial features and anchor objects

Image understanding precedes full autonomous classification. An insight in the behavior of anchor objects is therefore essential. To demonstrate stability and transferability of the classification an example is chosen from the land cover class 'urban fabric' Here contrasting edges are a must have condition for autonomous building detection. The resulting classified image always returns the brightest buildings inside the land cover class 'urban fabric' regardless the sensor or season. The class of bright buildings function as seed areas on which cityfootprints are assembled through further neighborhood relations. The given protocol in the case study should always return "bright buildings" from any image including any subset-image derived from Google Earth with VHR data (below 2,5 meter resolution see also figure 1) if present in the image frame. The first step in image understanding is the explicit description of stable and transferable features of standardized classes. The absolute gray values for these classes within the image have little meaning.. Their relative gray values, especially towards their neighbors (contrast) are crucial. To clarify a crucial feature and anchor objects, the presented case study shows the feature, '*contrast to neighbor pixel in the edge detection image* (Frame) displayed in figure 2 and figure 4. And in the (lower) histogram-extreme of this feature the anchor object class "edges-of-buildings" can always be located if build up areas exist inside the image.
Be aware of the quality difference between the original Quickbird and IKONOS bands and the deteriorated RGB imagery as used in Goole-Earth (figure 1). Is not necessary helpful for the quality of the output but principle results are not effected. The case study on such an arbitrary subsets from figure 1 can be used as an example with the character of education material to the debutant remote sensing student and can also be used as an introduction at high school level for remote sensing applications using Geobia techniques. Remember this is only one single building block similar to a description of the principles of a piece of LEGO brick or Meccano part which lies the foundation of more elaborated constructions.
Because of the initial stage of this development, the paper can demonstrate only transferability and autonomy on a completely arbitrary image from a subset in Google Earth which should be available to almost anyone anywhere. The reader/user can directly apply the process described in the paper to any other area within Google Earth by simply making a screenshot and run the process described in the case study using the process tree from figure 3. By using these instructions, the user can immediately confirm or reject the capabilities of the process regardless the origin and quality of his/her subset from Google Earth.

Figure 1. Urban fabric along the African coast from a Google-Earth screen-dump. RGB from VHRS sensor data.

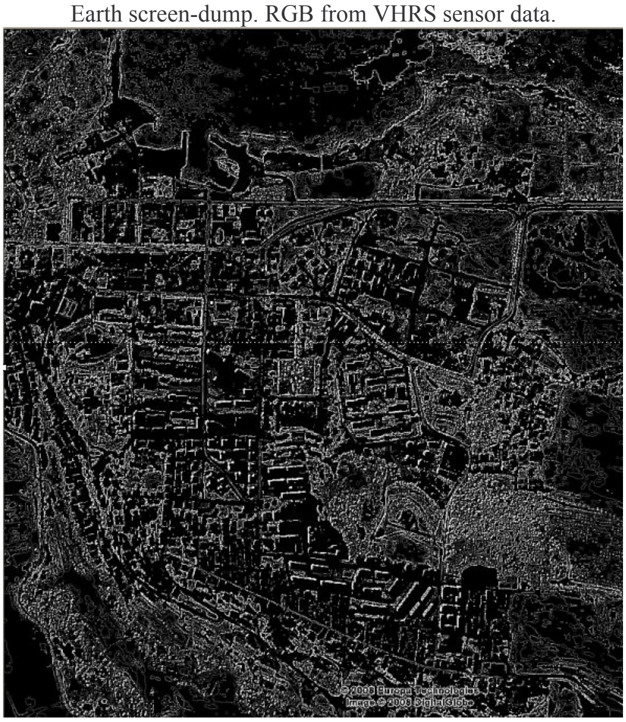

Figure 2 The crucial feature '*contrast to neighbor pixel (1) from the edge detection image* '***Frame***',See also . Detail in Fig.4

3. THE CASE STUDY

3.1 Preparations

An edge image is created with the intensity image as input. This can be a panchromatic band or in this case a (negative) first principle component (PC) image. After applying a Lee-Sigma filter on this (negative) PC image, the difference between the original and filtered image reveal all pixels which are mixed pixels on the edges of different image objects. The difference between an intensity image and it's Lee-Sigma filtered result is not the same for a bright object on a dark background compared to a dark object on a bright background. Therefore two edge images can be produced, which in our RS lab are named 'Border' and 'Frame'. 'Border' because they represent exactly those pixels of a bright object on a dark background and 'Frame' because they represent all pixels inside the dark edge around a bright object. In the latter 'Frame' highlights pixels which never would have existed as an object after a single segmentation. A typical artificial edge in 'Frame' is the inside of a shadow area. Normally the whole shadow surface has similar pixel values and is one segment. Due to creating an edge image from the edge-pixels of those dark areas, extra artificial objects start to appear.

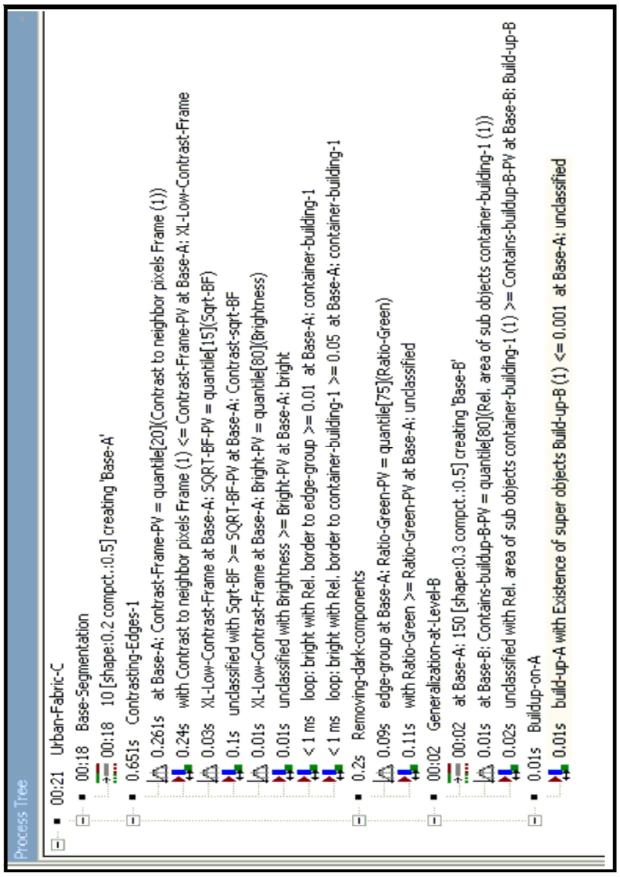

Figure 3. The Process tree in Definiens-Developer version 7

3.2 Segmentation Base-A and Base-B

Because it is most likely that the user is depending upon the facilities of Definiens Developer for reviewing this case study, the screendump of the process tree from this software is chosen to make it more recognizable for those users who like to run the process themselves. The chances are low other software allow to measure histogram extreme using Process Variables which are unique in newer versions of the Definiens Developer. The first step is a base segmentation using a quite low scale factor of 10.Only a weight factor of 3 for the first PC and a weightfactor 1 for Border and Frame. Users should be aware that for larger subsets (more than 3500*3500 pixel) processing time might run up.

Under *contrasting edges*-1 in the proces tree (Fig.3.) we start with measuring the 20 quantile of the histogram *<layer values/pixel based/> 'contrast to neighbor pixels Frame (1)'* on the population of all unclassified objects from the input image; Frame.Tiff (see also fig.4).

If we visualize this image in Figure 2. At first sight it might seem a pure Edge image. However this is not the case. Figure 2 is a visualization of a crucial feature. It shows the contrast of one object towards it's neighbor in a distance of 1 pixel. Because in the Frame.Tiff image the shadows have the highest values as this image are edges from the (principle component * (-1) +256 , 8 bit image) the resulting contrasting edges which are bright edges directly next to a strong shadow area have the lowest values. These are edge-objects in the lower 20 quantile of this crucial feature histogram. Here typical anchor objects like bright buildings can be located. The measurement returns a value for a Process-Variable. The value can change a lot over various scenes. The value of this Process Variable is now used in the second step for a classification of a category with the name *xl-low-contrast-frame*.

In the 3rd step under *contrasting edges*-1 we make a second measurement on the population of the category *xl-low-contrast-frame*. We measure the value of the crucial feature SQRT-BF which represents the squared difference of Border and Frame

(Square_Root((Border–Frame)^{2}*))*

Here we measure the 15th Quantile and use the value to classify the category of other contrasting edges. By applying this 15th Quantile, we ignore possible unwanted objects or failures of commission which are expected to be found in the lower 15th Quantile of this features histogram. In the 4th step this value is used to create the category *Contrast-SQRT-BF* from unclassified objects.

Now we measure brightness from the category *xl-low-contrast-frame*. Brightness is simply Red+Green+Blue/3.

With the 85 Quantile we are sure to capture only the value for the brightest buildings. We use this value in step 6 to classify all bright areas which have yet not been classified.

The first two categories *xl-low-contrast-frame and Contrast-SQRT-BF* are grouped in a category called *Edge-Group*.

All bright areas now classified are evaluated for being neighbor to any element from the *Edge-Group*. If they are neighbor they are assigned to the *container building-1* category. The next step is to let this *container building-1* category grow as long as bright objects are neighboring, using a loop. With these 8 steps under *contrasting edges*-1 we have a population of Bright buildings with a lot of noise.

The noise is removed in the next part under *removing dark components*. First we measure the value for the ratio Green (Green / Blue+Green+Red) in the edge-group population. This feature has normally high values for darker areas. The top 75 Quantile is classified back into the population of untouched unclassified objects.

Generalization takes place on a level above. A Super-Level, Base-B is created above Base-A. Using a scale factor of 150, only on the first PC, the segments start to appear about the size of city blocks around 2 –5 ha. The value for the upper 80 Quantile of containing subobjects from *container building* is used to classify this category into *Build-up-B* the 3rd step under *Generalization at level B*.

After Generalization at Base-B we return to the level below and remove all objects which have no superobjects on B with sufficient *container building-1* in the total area under a level B segment. All three categories *xl-low-contrast-frame, Contrast-SQRT-BF* and *Container Building* are part of the group *Build-up-A*.

3.3 Implications

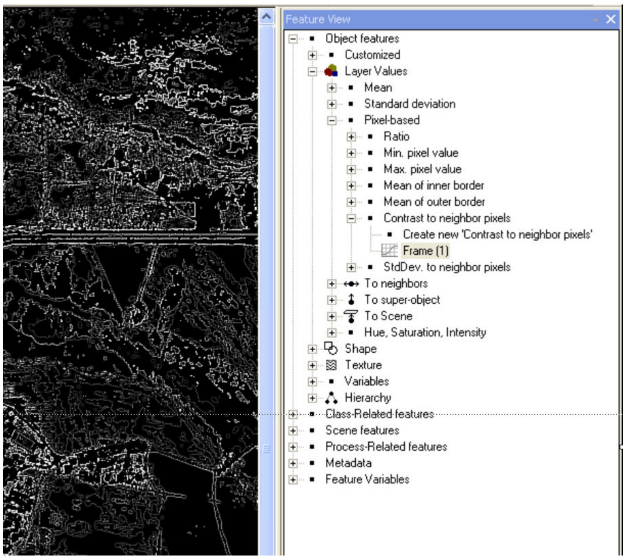

Figure 4 The crucial feature;
'*contrast to neigboring pixels Frame (1)*
Detail from Fig.2 in Definiens-Developer version 7

The core of the classification is the crucial feature *<layer values/pixel based/> 'contrast to neighbor pixels Frame (1)'*.(Fig.4.)
The resulting population contains all bright edges of buildings together with many unwanted objects. They needed to be removed using a ratio of spectral values (here from Ratio Green). The feature contrasting edges is more stable and reliable than ratio's among spectral values. The ratio's are also effected by unknown image manipulation techniques like pansharpening techniques. Therefore in a commercial environment an absolute control over the total data flow is a neccessity and subsets from Google Earth lack this information on preprocessing. In education environment we should be aware that this remains a weak link in the process chain. The features should always respond to buildings in any Google-Earth subset containing urban fabric. After a 2 level categorization we have a basic block to continue further classification.
There are no conditions inside the process tree for grouping. This is done separately in the classification tree and allows to make functional and strategic grouping according to user preference. In Base-B we define core areas over which a sufficient density of candidate building subobjects exist. This procedure removes a lot of noise of non-core areas. The aim is to have only build up areas with bright buildings as candidates. However we are not focused on the area the size of city blocks but on bright buildings themselves. That is why we return to classify level Base-A again. Here the proper sequence of classification shows it importance. The last process line is changing the original categorization on the Base-A level using results from the level (Base-B) above. This allows to use spatial features which are not directly related to neighboring objects, but to use the density of various image object categories from a much larger area and no necessity for direct spatial connection or contact exists. The conditions necessary to find spatial relationships in a larger neighborhood can not be found inside the level Base-A alone but needs the presence of a larger neighborhood defined by the Level Base-B Above.

4. DISCUSSION

The quality of subsets from Google-Earth are limited at the moment but high quality imagery will come available in due run. The quality is at least sufficient for demonstrating principles of Geobia to a much wider audience from first grade student level upwards.
Core issue remains the peer review by experts. This proposed principle of categorizing of contrasting edges is a case for further discussion among peers to be developed into teaching and demonstration material for full autonomous classification.
It would be a pity if segmentation is regarded as one of the most important breakthroughs since pixel classification and still the classification techniques derived from pixel based classification still continue such as sampling and maximum likelihood classification and cluster algorithms like K-nearest neighbor clustering analysis remain dominant. The need to try extensive experiments outside the statistical dominated feature space analysis into the realm of spatial classification rules is a condition sine qua non

Letting the computer calculation power decide on statistical decisions in spectral feature space risks to makes us lazy to search for further correct classification sequences and we need to explain our students how the brain processes imagery and how a machine should simulate this. The developemtn on strategies of sequencial classifications is only feasable by extending the knowledge of the expert.

Clusters can be found by statistical analysis in unsupervised classification. But close and far distances in normalized feature space have no effect at all on the proper design of a strategy using a sequence of categorization. This type of architecture is solely based upon case studies demonstrated and evaluated by peer-experts.

The case studies can develop empirical knowledge and their functionallity can be comprehensively demonstrated. Even if the theoretical knowledge on the reason why they exactly function are less clear. Low variance for textured vegetation in the Red Spectral Band for example is well known. Not it's biophysical cause in the interaction between sunlight and sensors IFOV towards the vegetation, but in it's application in classifying textured vegetation it demonstrates the functional contribution. It is a nice to discover the biophysical backgrounds of the interaction between Landcover objects and the light of the sun or their properties towards other wavelengths such as LIDAR intensity and RaDAR polarity.

Nevertheless if functionality can be demonstrated the procedures are still acceptable if they show functional classificication results on the large majority of the imaging data.

Sequential classification strategies are intuitive appealing as well as the focus on edge categorization. Both procedures are very close to normal visual interpretation by the operator. Also these procedures allow a peer group for critically evaluation.
Peer review allow such process to reach maturity. On selection of samples hardly any discussion is possible and thus never become mature on the other hand statistical cluster algorithms are mature already but prevent complete image comprehension.

The principles of crucial feature, anchor objects and sequential classifications can be demonstrated with such simplicity to allow debutants to work with the data. Full autonomous classification processes need a huge library of crucial features for anchor classes. Some can be demonstrated already , others are under construction.

5. OUTLOOK

The whole demonstration is to make clear how a principle can be developed to use a crucial feature and apply this to autonomously detectable anchor objects (bright buildings) regardless the image quality and scale from 2,5 Meters and higher resolutions all around the Globe.

Although the result of the case study is meager with only one core class, the bright buildings. These are the cornerstone for further development of the Landcover class *Urban fabric*. Adding height information makes much further differentiation possible. Using the seed areas for further growth would make up the core of an autonomous Urban footprint classification, full autonomously.

Objects of interest can be linked to a set of crucial features in the manner a look-up-table. This will make a Geobia recepy book feasable. Contributions on similar working case studies will advertise for Geobia classification methodes, which do not require operator interaction anymore especially no more visual sampling.

To apply these features, the user should start with an evaluation of the crucial feature *contrast to neigboring pixels Frame (1)* and check the lower Quantile of this Histogramm for response of existing buildings in his image of choise on any region of interest. We are quite sure bright buildings will respond as they did in our tests in Afrika, Asia and European environments. We are confident the user finds out himselve this feature is responding everywhere around the globe.

Making it work is an invitation to the audience, hoping a similar process can be part off a total free trial version for all (would be) users of Geobia.

Interested readers can be provided with the .DRP and .DCP files used in Definiens Developer 7, from the case study after sending an email request to the corresponding author

References from Books:

Haberäcker, P., 1995, *Praxis der digitale Bildverarbeitung und Mustererkennung*, Hanser, München-Wien.pp 108-114

References from Websites:

An Object based information extraction methodology incorporating a-priory spatial information;t ESA-EUSC 2006: Image Information Mining for Security and Intelligence EUSC, Torrejon air base - Madrid (Spain) Conference: November 27-28, 2006; Thomas Blaschke, Dirk Tiede, Stefan Lang http://earth.esa.int/rtd/Events/ES-USC_2006/ Oral/Ar06_Blaschke.pdf, accessed 28-april-2008