Clustering algorithms for aerial photographs and high resolution satellite images

Matthias Zerbst[1]; Lars Tschiersch[1]; Mohamed Talbi[2]; Gabriela Guimarães[3]; Wolfgang Urfer[1]

[1] Department of Statistics, Universität Dortmund, Germany
[2] Remote Sensing Laboratory, Institut des Regions Arides.4119.Medenine, Tunisia
[3] CENTRIA, Department of Computer Science, Universidade Nova de Lisboa, Portugal

1 Introduction

This work, and specially, the use of clustering algorithms was motivated by the need to perform a field-study with erosion data from arid areas. Using data obtained from analyzing erosion, land degradation and desertification phenomena will show some limitations. If only terrestrial observations are considered. Specially, if we are interested in, for instance, forecasting problems of erosion spread. An improvement of the data is possible, if aerial photographs and recent high resolution satellite images are additionally taken into account.

The uprising problem with such images is that they contain a huge amount of information, and standard processing algorithms are, in most cases, unable to answer the analyst needs. In order to solve these problems, a compression and suitable selection of the underlying information is needed. Although the development of computer has reached a stage that enables the handling with huge data-sets, considerations concering time complexity are still relevant.

In this paper, we present the developed algorithms and discuss possible improvements to attein our aim in performing a classification within a suitable computational time.

In section 2, we describe algorithms, such as ISODATA and PHASE, that are based on the classical k-means algorithm. Section 3 describes two ways of finding a set of good starting seeds (centroids) for classification with an adapted method from the known single linkage and the Kohonen networks, as well. Section 4 presents the application of the methods from section 3 to aerial photographs and high resolution satellite images.

2 Basic concepts

In this section, a few fundamental terms associated with photographs and images are presented.

It is important to distinguish between the terms photographs and images in remote sensing. An image refers to a pictoral representation, regardless of what wavelength or remote sensing devises has been used to detect or record the electromagnetic energy. A photograph refers specifically to images that have been detected as well as recorded on photographic film, generally in the visible or near infrared parts of the spectrum, and are normally recorded over the wavelength range from 0,3 µm to 0,9 µm. Based on these definitions, we can consider that all photographs are images, but not all images are photographs. A photograph could also be represented in a digital format by subdividing the image into small equal sized and shaped areas, called picture elements or „pixels“, representing the brightness of each area with a numeric value or digital number. Therefore, photographs have to be scanned and subdivided into pixels, with each pixel assigned to a digital number representing its relative brightness. For panchromatic or black & white images, we usally use one channel or band, and combined brightness levels of each pixel. It is the same for each primary color (R,G,B) and shows various shades of gray from black to white. For color images, we display more than one channel (generally three) as a different primary color (R,G,B), then the brightness is generally different for each channel / primary color combination. Each combination forms a color image.

3 Unsupervised methods

When analysing aerial photographs as well as satellite images, several difficulties occur that have to be solved in order to interpret the images from a statistical point of view. To handle such problems, aerial photographs as well as satellite images have to be classified into sections of interest. Therefore, a classification algorithm is needed. Classification, in general, is a pattern-based process that assigns individual pixels to categories based on spectral properties.

Usually pictures contain a lot more information than one can take into account. Therefore, the technique for data compression have been used. For this kind of application, methods like principal component analysis have been used. The first principal component is the best linear transformation of the multivariate data with respect to the variance. Essential landscape information will be lost by this method such that it becomes necessary to use other algorithms. In this context, classification methods seems to be appropriate.

Image classification is used to digitally identify and classify pixels in the data. Classification is usually performed on multi-channel data sets, and this process assigns each pixel from an image to a particular class or theme, based on statistical properties of the pixel brightness values.

There are a variety of approaches that perform a digital classification. We will briefly describe the two generic approaches which are often used, namely, supervised and unsupervised classification.

In a supervised classification, the analyst identifies from an image homogeneous representative samples of the different information classes of interest. These samples are referred to as “training areas“. The selection of appropriate training areas is based on the analyst’s familiarity with the subject and the information present in the image.

The numerical classes given by the analyst are used to “train“ the computer to recognize spectrally similar areas for each class. The classification is done by comparing each pixel from the image to these signatures for each training class (Talbi (1999)).

The unsupervised classification reverses the process. First, spectral classes are grouped, based solely on the numerical information in the data, and then are matched by the analyst to the information classes. “Clustering algorithms“ determine statistical groupings or structures in the data. This iterative clustering process results in some clusters that the analyst could combine or break down further (Talbi (1993)).

A standard clustering technique is the iterative self-organizing data analysis (ISODATA), developed by Tou and Gonzales (1974). The ISODATA clustering technique initially divides the image data into equal areas of data variance, based on the output number of classes designated by the user. For each pixel the distance to each cluster mean according to a given metric is calculated and assigned to the nearest cluster. Each pixel’s spectral (euclidean) distance is reevaluated after adjusting the cluster means (centroids) to better fit the image variance. This process is recursive. It starts with a predefined number of starting seeds for the required – user defined – amount of categories. The algorithm can be mathematically described as follows.

Throughout this paper let Ω be the parameter space of the values describing a pixel (Note that: Each pixel is a data representation. Take the “content“ of it.). Let $C = \{c_1, ..., c_p\}$ be clusters, $z_1, \dots, z_p \in \Omega$ be the cluster centroids and $x \in \Omega$ a vector of a point from the underlying image. Choose $z_1, \dots, z_p$ randomly, for instance by simple random sampling. Then assign x into C, if and only if,

$$x \in C_i \Leftrightarrow z_i^{(0)} = \arg \min_{z_1^{(0)}, ..., z_p^{(0)}} \left\| x - z_k^{(0)} \right\|_2 , \qquad (3.1)$$

where $z_i^{(o)} \in Z^{(0)} = \{z_1^{(o)}, ..., z_p^{(o)}\}$ indicates one element of the first set of centroids.

After creating all p clusters $C_1, \dots, C_p$, calculate the mean according to

$$\overline{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} , \qquad i = 1,...,p . \qquad (3.2)$$

Now derive new starting seeds (centroids) from $z_i^{(1)} := \overline{x}_i , \quad i = 1,..., p$.

For a new recursive step start with (3.1).

The centroids are moved through the multidimensional space in direction to the mostly densed sections of the image. After a user specified number of iterations the algorithm will terminate. The calculated means of the last step are the centriods of categories.

The k-means algorithm may be regarded as an improvement of the ISODATA method (Hartigan (1975)). The advantage of this approach lies in convergence of the sequence of starting seeds (centroids), derived from the calculation of (3.2).

The convergence is independent from the initial choice of centroids and can directly be derived from (3.1) and (3.2). Bock (1998) demonstrated the equivalence to the minimum variance problem or SSW clustering criterion

$$g_n(C, Z) = \frac{1}{n} \sum_{i=1}^{p} \sum_{k \in C_i} \left\| x_k - z_i \right\|^2 \to \min_{C,Z} \quad , \qquad (3.3)$$

where $Z = \{z_1, ..., z_p\}$ is p-tupel of class centers with $z_1, ..., z_p \in \Omega$.

According to (3.1) and (3.2) a data-vector will be assigned to a centroid when it has the minimal euclidean distance. For each step an improvement is performed, until no further improvement is achieved.

Even if this strategy seems to be a suitable method, there are a couple of drawbacks. First, aerial potographs or satellite images usually contain some million pixels. This algorithm can easily be take a few weeks or months, even with very fast computers.

Second, small but essential objects may be ignored during classifying. To overcome this disadvantage, an approach was developed which uses hyperclusters.

Kelly and White (1993) developed an approach based on hyperclusters, that divides the picture in far more categories than needed. An improvement of this approach was obtained by Myers et al (1997), named Pixel Hyperclusters Approximating Spatial Ensembles – short PHASE.

It maximizes the minimum euclidian distance between the centroids, such that distinguishable categories without a loss of important information are found. A way of controlling the minimum cluster size without reducing the number of clusters has been provided by introducing a dynamic cluster splitting facility. Miniscule clusters have the centroids replanted in a manner that split clusters have a larger sum of squared deviations from the centroid. The latter is accomplished by tracking the migration of cluster seeds through multidimensional space (MYERS ET AL (1997)).

Usually PHASE clusters will be too numerous to serve directly for further analysis. Since PHASE clustering of a large spatial dataset may require computer runs extending over a couple of days, reworking of the original data would be a last resort. An alternative is to cluster the clusters. Thereforem, a weighted re-clustering of the cluster mean vectors are treated as data vectors and the cluster frequencies become the weights. That means that cluster of clusters become superclusters. Consequently, several recursive runs of the algorithm are possible. This reduces the number of clusters step by step to the needed amount of categories.

The advantage of PHASE is that the image is accessed just once for the analysis. Hereafter, all information is stored into relational tables, known form computer database environments such as Oracle. Relational tables are used, when information about an object is stored in one table and further detail information is stored in some other tables. All information can be connected together by keyelements of the tables, which define a relation between them.

The second advantage lies in a „starting seed"-dispersion-algorithm of PHASE. This algorithm disperses the starting seed $z^{(0)}$ over the picture arbitrarily. According to Cressie (1993) and Myers et al (1997) the selection of the first centroids $z^{(0)}$ should maximize the minimal euclidean distance between the centroids for a good classification. Mathematically,

$$\max_{\{z_1^{(0)},\ldots,z_p^{(0)}\}\in\Omega} \min_{z_1^{(0)},\ldots,z_p^{(0)}} \left\| z_i^{(0)} - z_j^{(0)} \right\|_2 , \tag{3.4}$$

where Ω is the parameter space.

A further improvement of PHASE is that far to small categories will be avoided. For possible too small categories, the centroids will taken away and the pixels will be assigned to the next nearest centroid. The classification algorithm is analogous to the k-means algorithm. For each category $C_1,\ldots,C_p$ a mean vector will be calculated and the centroids will be numbered due to

$$\mathcal{NO}(c_i) := \text{Rank}\ \left(\|z_i\|\right) \quad .$$

Additionally, the inter class variance (SSB)

$$\mathcal{SSB} = \sum_{i=1}^{p} n_i \left(\overline{x}_i - \overline{x}\right)^2 , \tag{3.5}$$

with $\overline{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$, $i = 1,\ldots,p$ x_{ij} the j-th observation in the i-th cluster and $\overline{x}$ as the overall mean; and the intra class variance (SSW)

$$\mathcal{SSW} = \frac{\sum_{i=1}^{p}\sum_{j=1}^{n_i}\left(x_{ij} - \overline{x}_i\right)^2}{\sum_{i=1}^{p} n_i} \tag{3.6}$$

is calculated.

The PHASE approach is only created for analyzing grayscale pictures. However, our aim is to get color images from our analysis, since colored pictures provide a lot more information than grayscale pictures.

4 Identification of cluster centers

The presented methods of chapter 2, will be modified in that manner that two essential innovations will be introduced.

First, we will convert the approach from gray-scale to color pictures and secondly the correct set of „starting seeds“ – the first centroids of the clusters – will be improved.
When analysing color pictures, a special color format has to be chosen. In our case and for most pictures of this type a format called RGB is appropriate.

4.1 RGB color model

When we talk about color, we have to keep in mind that color can be realized in different ways. That are the physical aspect, the physiological aspect, the technical aspect and the color models which are based on that facts.
To see how a RGB-model works, we first have to look at the technical aspect of color.
On the basis that the human eye perceives color reaction through a mix of red, green and blue signals, it was thought that each color can be mixed with only three primary colors. Unfortunately, there are no such colors which will do this.
In order to have a standardized way of mixing colors, the Commision Internationale de L'Eclaire (CIE) creates three artificial primary colors. Let these colors be noted as X, Y and Z and be characterized by its energy distribution. (Wyszecki (1982)). The CIE developed a standardized color table with which every color can be calculated by

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

Note, that $z = 1 - x - y$. With regard to this it is only necessary to have the (x, y) coordinates.
If a color is given by its (x, y) coordinates one needs, to calculate the contribution for primary colors, the brightness. But due to definition of CIE this is exactly the Y-Value. Hence, the coordinates X and Z can then be derived by

$$X = x\frac{Y}{y}, \qquad Z = z\frac{Y}{y}. \tag{4.1}$$

With this basic definition we are now able for a short look on a RGB color model.
Due to the CIE definition, every color can be derived as a linear combination of the three primary colors. Therefore, the RGB color space is represented by a 3-dimensional, cartesian coordinate system with standardized componets R, G and B to lie between zero and one.
When using such a model, one problem has to been solved. When mixing all colors one needs to get the color white. To reach this aim in a RGB model, one has to fix one point and refer to this as white. The usually choice is to take

$x=0.313$, $y=0.329$ and $z=0.358$. This has its origin in physics, where this is the white of a black body heated to 6500 celvin.
With a brightness of $Y=1$ and with (4.1) the contribute of X and Z to the color white are $X=0.951$, $Z=1.088$.
If a color in the RGB color space is given by the coordinates (R,G,B), the X,Y and Z coordinates be derived from

$$\begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \end{pmatrix} = \begin{pmatrix} 0.478 & 0.299 & 0.175 \\ 0.263 & 0.655 & 0.081 \\ 0.020 & 0.160 & 0.908 \end{pmatrix} \begin{pmatrix} \mathcal{R} \\ \mathcal{G} \\ \mathcal{B} \end{pmatrix}$$

and vice versa

$$\begin{pmatrix} \mathcal{R} \\ \mathcal{G} \\ \mathcal{B} \end{pmatrix} = \begin{pmatrix} 2.741 & -1.147 & -0.426 \\ -1.118 & 2.028 & 0.034 \\ 0.137 & -0.332 & 1.105 \end{pmatrix} \begin{pmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \end{pmatrix}.$$

To define a computer based RGB model the continuous model can be transformed to a discrete model on the parameter space $\{0,1,...,255\}$ for each component. Hence, a computer based RGB model is stored in a 24-Bit environment implying $2^{24}=16.7$ million colors. Note, that in a computer based RGB model, the RGB value from each pixel of a picture is stored separately.

Let us now return to the above mentioned problems.
To analyse the image, we start to calculate a frequency table of the RGB-values. This values will be our database for further analysis. The analyst has to specify the number of finally needed categories. This choice has to be specific for the underlying formulation of questions.
After doing so, one has to specify which points of the picture of interest, more precisely which RGB vectors, are the first centroids $z^{(0)}$. To reach this aim, we would like to suggest a method which is related to the well known multivariate single linkage method. This new algorithm is called maximum linkage.

4.2 Maximum Linkage

The mentioned single linkage (see Johnson (1992)) analyses the similarity structure of a set of points in IR^n, by joining the points or clusters with the minimal euclidean distance in each step.

The idea of maximum linkage is to summarize the points stepwise within only one cluster, so that the minimum euclidean distance between the points within the cluster in each step is maximal with respect to all possible alternatives. Let $P_1,\ldots,P_r$ be the points and $\{P_1,\ldots,P_v\}$ the cluster. Then

$$\min_{P_i,P_j\in\{P_1,\ldots,P_v\}} \left\|P_i - P_j\right\|_2 \rightarrow \max_{\{P_1,\ldots,P_v\}\subset\{P_1,\ldots,P_r\}} \qquad (4.2)$$

has to be derived.

That means that there will be only one cluster at all time and all remaining points, while using single linkage there can be more than one cluster.

Hence, one needs to calculate the distance d_{ij} between all n possible points. The distances will be collected in a distance matrix D. Therefore:

$$D = (d_{ij}),\quad i,j = 1,\ldots,r\,.$$

Now, the maximum of all distances d_{ij} in D is taken and the corresponding points summarized within a cluster. Further one needs to calculate the distances of each point P_i to the points P_j of the cluster and the results should be put in a new distance matrix, which by now is only a vector.

Define Q as the set of indices which belong to the points in the cluster and R as set of indices which belongs to the points which are not in the cluster, so $R = \{1,\ldots,r\}\backslash Q$. The distance is now given by

$$d_{iQ} = \min_{k\in Q} \left\|P_i - P_k\right\|_2, \quad i \in R\,. \qquad (4.3)$$

The maximum of the distances d_{iQ} is taken and the corresponding pixel is added into the cluster. The last step of the above described steps will be repeated as long as the specified amount of points within the cluster is reached.

4.2.1 A two dimensional example

Given eight points with the coordinates and distances of each point to another. The distances are collected in the first distance matrix.

	Points	1 (1,3)	2 (1,5)	3 (2,1)	4 (3,3)	5 (3,5)	6 (4,1)	7 (4,4)	8 (5,3)
1	(1,3)	0							
2	(1,5)	4	0						
3	(2,1)	5	17	0					
4	(3,3)	4	8	5	0				
5	(3,5)	8	4	17	4	0			
6	(4,1)	13	25	4	5	17	0		
7	(4,4)	10	10	13	2	2	9	0	
8	(5,3)	16	20	13	4	8	5	2	0

Due to the maximum linkage method the points *(1,5)* and *(4,1)* will be summarized in the cluster. The distances are calculated using formular (4.3). By doing so one gets the following distance matrix. The gray marked area is not needed for further calculations.

	Points	1 (1,5),(4,1)	2 (1,3)	3 (2,1)	4 (3,3)	5 (3,5)	6 (4,4)	7 (5,3)
1	(1,5),(4,1)	0						
2	(1,3)	4	0					
3	(2,1)	4	5	0				
4	(3,3)	5	4	5	0			
5	(3,5)	4	8	17	4	0		
6	(4,4)	9	10	13	2	2	0	
7	(5,3)	5	16	13	4	8	2	0

The point *(4,4)* is selected and added to the cluster consiting of the points *(1,5)* and *(4,1)*. Due to that the following distance matrix results in the next step.

	Points	1 (1,5),(4,1),(4,4)
1	(1,5),(4,1),(4,4)	0
2	(1,3)	4
3	(2,1)	4
4	(3,3)	2
5	(3,5)	2
6	(5,3)	2

Now we have a special case. We have two maxima. The algorithm picks one of the two points, *(1, 3)* and *(2, 1)*, randomly.

4.3 Kohonen networks for selecting centroids

Beside the selection of centroids with the maximum linkage algorithm, it is worthwhile to consider neural networks, as well. Several ANNs, such as Multi Layer Perceptrons, Hopfield networks or Kohonen networks may be used for clustering tasks (Bock (1998)). However, we will focus on Kohonen networks, since experiences with other applications using Kohnen networks (Kohonen (1982)) for clustering have already been made (Guimarães and Urfer (2000)).

ANNs are biologically motivates. Consequently, terminologies such as neurons and weights are used here. The adaption of the parameters (weights) is called learning, and is usually performed, when an input pattern is to an ANN (on-line lerning). Kohonen networks adapt their internal structures (weights) to the structural properties (e.g. regularities, similarities, frequencies, etc.) of high-dimensional input data.

During learning Kohonen networks adapt their weights such that a n-dimensional input space is projected onto a m-dimensional lattice with $m<n$, preserving the neighborhood of the input data on the lattice. Usually, a two-dimensional lattice is chosen. All neurons on this lattice are connected with each other. The lattice is formed by the properties inherent to the data itself. If we consider classification from a statistical sense, by finding a distribution statement on the parameter space, each neuron is assigned to a point in the parametes space, namely a weight vector. As Bock (1998) states, this problem has been tackled also to projection pursuit or principal components.

During the learning phase the weights are changed at each iteration step. At each iteration step, a best-match is searched among the neurons on the lattice according to some distance measure, usually the euclidean distance. Next, during the same iteration step, the weight vectors of the best-match and all neurons in a neightborhood are moved towards the input vector according to a neighborhood function d.

Let $X=(x_1,\ldots,x_n)$, $x_i \in IR$, be the input vector of a Kohonen network, $W_i = (w_{i1},\ldots,w_{in})$ a set of weight vectors, with weight w_i belonging to neuron $i = 1,\ldots,k$ and k the number of neurons.

To find the neuron j with the minimal euclidean distance between input- and weight vector consider

$$j = \arg \min_{i=1}^{k} \left(\left\| X - W_i \right\| \right) .$$

The weight vectors within the neighborhood of the best-match are moved towards the input vector with regard to their distance to the best-match within the m-th iteration step as follows

$$W_i(m+1) = W_i(m) + \eta(m) \cdot d_{ij}(m) \cdot \left(X - W_i(m) \right), \quad \forall\, i = 1,\ldots,k. \qquad (4.4)$$

The learning rate in (4.4), notated by $\eta \in [0,1]$ decreases monotonically with $\eta(m+1) = \alpha\eta(m)$, and $\alpha \in (0,1)$. The factor α can be seen as the speed of convergence. The neighborhood function d in (4.4) is a distance function which describes the changes of the weight vectors of all neurons with respect to the best-match. There are many possible functions d, such as the gauss function or the mexican hat function. For example, the Gauss function

$$d_{ij}(m) = e^{-\frac{\left\| w_i - w_j \right\|^2}{\sigma^2(m)}} . \qquad (4.5)$$

judges the distances of the weight vectors between each other, where $\sigma^2(m)$ is a suitable variance function, which depends on m-th iteration step.
Bock (1998) shows that under the assumption of the choice of a suitable function d from (4.4) an equivalent to (3.3) is given.

5 Application

Referring to our above mentioned problem of classifying pictures, we are going to choose the "starting seeds" by using the maximum linkage algorithm. The used points within the method are RGB-vectors of a single point of the underlying picture. In contrast to the single linkage method the maximum linkage method terminates when the required amount of "starting seeds" (centroids) is reached.
The problem arises that, regarding to the frequency distribution of the picture points, too many points, will be lying among the edge of the distribution. And a smaller amount of points will be within the center. Therefore, we suggest two separate sets of centroids $z^{(0)}$. The first set will be calculated due to the maximum linkage method mentioned above. For the calculation of the second set of centroids one needs to modify the method slightly. The distance d_{ij} is weighted with w_{ij}, the sum of the absolute frequency of the corresponding points within the picture. Formal $w_{ij} = \mathcal{H}_n(\mathcal{P}_i) + \mathcal{H}_n(\mathcal{P}_j)$.
Suppose we take the points $(1,5)$ and $(1,3)$ of example 4.2.1 . Furthermore suppose that this points have the weight 10 and 5, respectively. The factor which one needs to use in this case for weighting is 15.
Due to the usage of that modification one should be able to catch enough points either on the edge of the distribution (rare objects) and in the center (area of interest).
Now we know how we should select the sets of centroids $z^{(0)}$, but we also have to discuss how many centroid points we need in each of the two sets. To solve this, let us take a look on Figure 1.

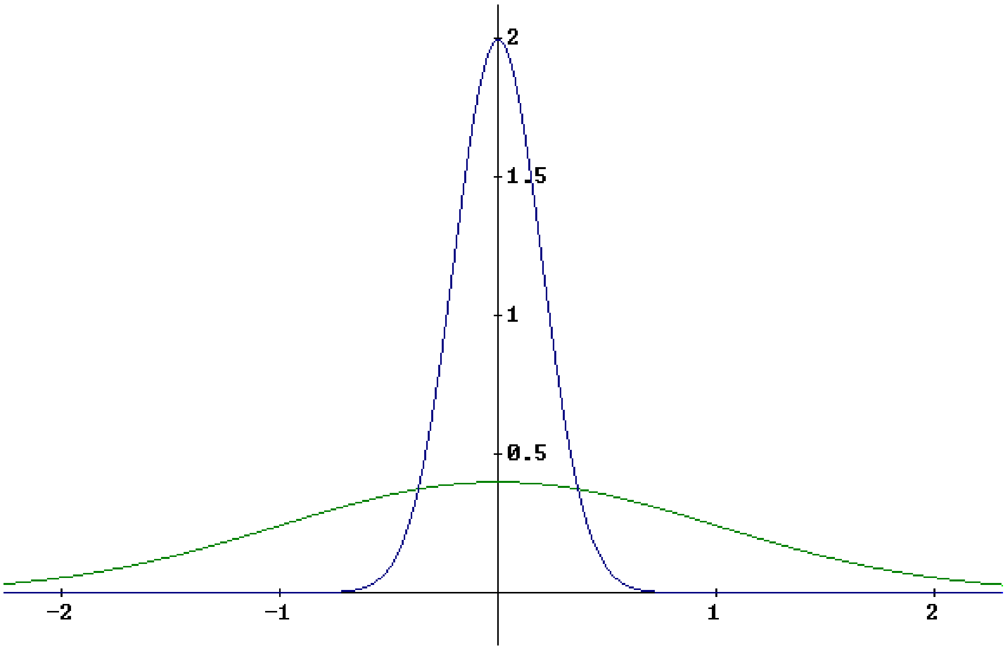

Fig.1: Scetched density function of the color values from two pictures.

Suppose we have a color distribution according to the steep curve. In this case the contribution of unweighted centroids has to be higher than the weighted ones. The reason is that their exist only a few values with a huge frequency.
In the case of color distribution regarding to the flat curve, one needs to take a higher number of weighted centroids. This implies that the variance is a measure for the determination of the contribution of weighted and unweighted centroids.
For the case of the application discussed here, the parameter space for our RGB-vectors will be $\Omega = \{0,1,...,255\}^3$. We now define a variance based parameter κ for selecting the number of weighted and unweighted starting seeds.
Due to the definition of the RGB model, each component of such a RGB-vector is independend of each other. Therefore, the covariance is zero. Hence, considering the variance of this components is sufficient. Each of the three variances are of the same scale. In order to get κ, we are going to sum over the variances for each component. κ is given by

$$\kappa = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)'\left(x_i - \overline{x}\right), \qquad (5.1)$$

with $x_i = (\mathcal{R}, \mathcal{G}, \mathcal{B})'$ as a 3-dimensional RGB color vector and $\overline{x} = \left(\overline{\mathcal{R}}, \overline{\mathcal{G}}, \overline{\mathcal{B}}\right)'$.

With respect to the bounded parameter space, κ can be assessed by

$$0 \le \kappa \le 48768.75 \,. \qquad (5.2)$$

In (5.2) zero is achieved when all color values are the same. The upper boundary is reached, when an equal number of values is located at both ends of the parameter space and n is even. This holds regardless of the magnitude of n.

According to this above consideration the following heurtistically derived function for calculating the contribution of weighted and unweighted centroids can be considered

$$f_w(\kappa, p) = \left\lfloor \left(\sin\left(\frac{\pi\kappa}{48768.75} - \frac{\pi}{2} \right) + 1 \right) \frac{p}{2} \right\rfloor , \quad (5.3)$$

$$f_{uw}(\kappa, p) = p - f_w(\kappa, p) , \quad (5.4)$$

with p = number of “starting seeds“ (centroids $z^{(0)}$) and κ as given in (5.1).
Figure 2 shows (5.3) for $p=10$.

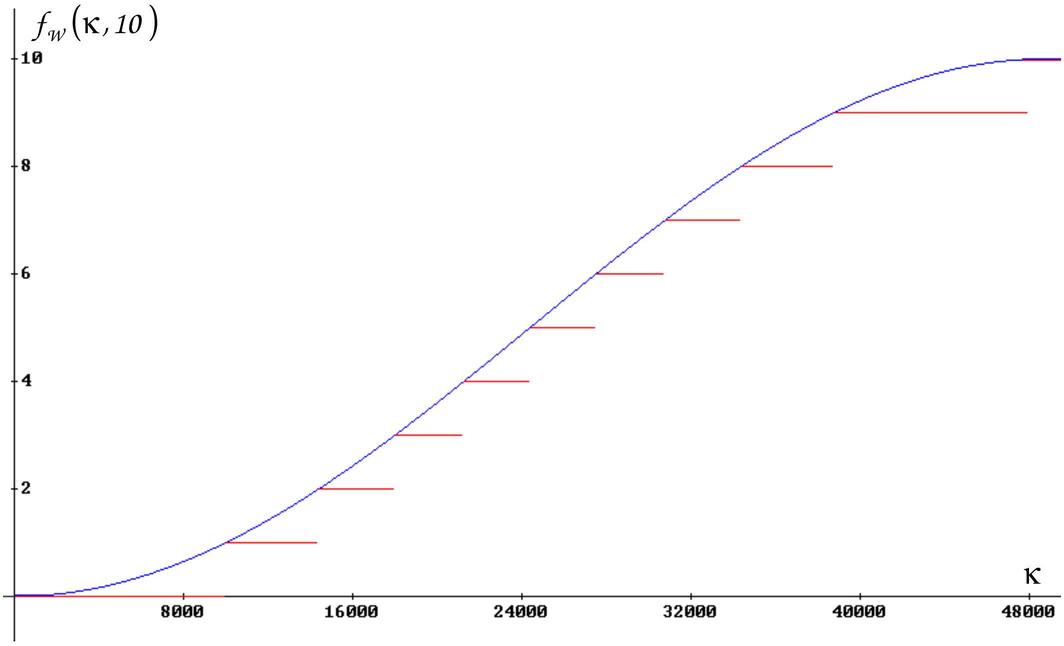

Fig.2: Visualization of (5.3) for $p=10$.

For selecting centroids $z^{(0)}$ with Kohnen networks we adapt the methods of section 4.3 for the parameter space $\Omega = \{0,1,\ldots,255\}^3$.
As Braun (1997) stated, we sprout a net star-shaped out of the center. For that reason, the mean is calculated across all pixels of the original picture (RGB-values). After that the first centroids are generated by adding a standard normal distributed random variable to the mean component of the RGB vector.

To start learning we have to specify $\eta(m)$ and d_{ij} from (4.4). We choose

$$\eta_b(m) = b^m , \qquad b \in (0,1) .$$

The constant b has to been choosen with respect to number of iteration steps. If one wishes to have a learning rate of 5-thousandth in the last iteration step, one has to select b in dependence of the maximal number of iterations steps m_{max},

$$b = (0.005)^{1/m_{\max}} .$$

The function d_{ij} will be used independently of m, in accordance to (4.5). In this case the neurons i will directly represented by itt weight vectors w_i. During the learning phase the RGB-values of each pixel will gradual be used as input for the network. The network will be moved step by step and will end up with a structure that represents the categories of the classification. After the learning phase each neuron will represent a category.

The network will then be used to assign all pixels x of the orginal picture to categories C_i, with respect to the minimal euclidean distance. Hence, we have

$$x \in C_i \Leftrightarrow w_i = \arg \min_{w_1,\dots,w_p} \| x - w_k \|_2 .$$

For practical consideration both methods, maximum linkage and selection using Kohonen network, will soon be applied within a computer environment to do some simulation regarding our classification problem.

To compare the quality of the reached classification we define in analogy to (5.1) two additional parameter for (3.5) and (3.6). Note, that we don't use RGB-values it self, but the frequency values h_{ij} of the RGB-values of the underlying picture. Therefore, the parameter is given as

$$\kappa_{SSB} = \sum_{i=1}^{p} n_i (\bar{x}_i - \bar{x})' (\bar{x}_i - \bar{x}) , \qquad (5.6)$$

with $\bar{x}_i = \sum_{j=1}^{n_i} h_{ij} x_{ij} \Big/ \sum_{j=1}^{n_i} h_{ij}$ and h_{ij} as the frequency of the j-th RGB-vector in cluster i and $\bar{x}$ as the overall mean-vector of all components instead of the inter class variance. An equivalent for the intra class variance for this problem is

$$\kappa_{SSW} = \frac{\sum_{i=1}^{p} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)' (x_{ij} - \bar{x}_i)}{\sum_{i=1}^{p} \sum_{j=1}^{n_i} h_{ij}} . \qquad (5.7)$$

6 Outlook

Further practical applications of this method may help to improve the data quality of a data analysis. This could be applied in an arid-ecosystem research project in cooperation with Tunisian Institut des Régions Arides (Environment Observatory

Research Program); using infrared aerial photographs (Urfer et al (1994) and Kötting at al (1998)) as well as available very high resolution satellite images from LANDSAT, IRS or space imaging IKONOS systems. This research project will contribute to a better knowledge of land degradation, erosion and desertification and is to be seen as a first step of a frontierless efford to study, monitor, observe and combat that phenomenon, in concordance with articles 8, 16, 17, 24 & 25 of the United Nations Convention to Combat Desertification (UNCCD (1994)).
UNCCD calls to improve understanding of causes and impacts, promote international cooperation in scientific research and observation of desertification, develop know how, draw scientific researchers into a network to support the Convention, and finaly enhance scientific and technology capacity building.

Acknowledgment
This work has been supported by the Collabrorative Research Centre "Reduction of Complexity in Multivariate Data Structures" (SFB 475) of the German Research Foundation (DFG).

References

Bock, H.-H. (1998): *Clustering and neural networks.* 6-th Conference of International Federation of classification Societies, Rom.

Braun, H. (1997): *Neuronale Netze: Optimierung durch Lernen und Evolution.* Springer, Berlin.

Cressie, N.A. (1993): *Statistics for spatial data.* Wiley, New York.

Guimarães, G.; Urfer, W.(2000): *Self-Organizing Maps and its applications in sleep apnea research and molecular genetics.* Technical Report 23/2000, SFB 475, Department of Statistics, University of Dortmund.

Kelly, P.; White, J. (1993): *Preprocessing remotely-sensed data for afficient analysis and classification. Application of artifical intelligence.* Proceedings, SPIE 1993, pp. 24-30, Knowledge Based Systems in Aerospace and Industry.

Kötting, J.;Bonney, G.E.; Urfer, W. (1998): *Disposition models for the analysis of dynamic changes in forest-ecosystems.* Technical Report 25/1998, SFB 475, Department of Statistics, University of Dortmund.

Hartigan, J.A. (1975): *Clustering algorithms.* Wiley, New York.

Kohonen, T. (1982): *Self-organization formation of topologically correct feature maps.* Biological Cybernetics43, 141-152.

Johnson, R.A.; Wichern, D.W. (1992): *Applied multivariate statistical analysis.* 3rd-Edition, Prentice Hall, Engelwood Cliffs, USA.

Myers, W.; Patil, G.P.; Taillie, C. (1997): *PHASE formulation of synoptic multivariate landscape data.* Technical report Number 97-1102, Center for statistical ecology and enviromental statistics, Department of Statistics, PennState University, USA.

Talbi, M. (1993): *Contribution á l'etude de la désertification par Télédétection dans la Jeffara: Sud est de la Tunisie.* These de Doctorat en Géographie. Université de Tunis-I.

Talbi, M. (1999): *Spectral signature for desertification pattern recognition in arid environment, using remote sensing.* Technical paper, Institut des Régions Arides. Tunisia.

Tou, J.T.; Gonzalez, R.C. (1974): *Pattern recognition principles.* Addison-Wesley, Reading, Mass.

UNCCD (1994): *United Nations Convention to Combat Desertification*

Urfer, W.; Schwarzenbach, J.; Kötting, J.; Müller, P. (1994): *Multistate models for monitoring individual trees in permanent oberservation plots.* Environmental and Ecological Statistics 1, pp. 171-199.

Wyszecki, G.; Stiles, W.S. (1982): *Color science, concepts and methods. Quantitative data and formula.* John Wiley, New York.