Methods for classification of agricultural fields in aerial sequences - a comparative study

Zweitze Houkes[a], Haijun Chen and Jan-Friso Blacquière,
University of Twente, Faculty of Electrical Engineering,
P.O.Box 217, 7500 AE ENSCHEDE, The Netherlands

ABSTRACT

A comparative study of a selection of classification methods for agricultural fields in sequences of aerial images is presented. The image sequences are acquired by an RGB-CCD video camera which is assumed to be on board of an airplane, moving linear over the scene. The objects in the scenes being considered are agricultural fields. The classes of agricultural fields to be distinguished are determined by the type of crop, e.g. potatoes, sugar beet, wheat, etc. In order to recognize and classify these fields obtained from the aerial sequences of images, a common approach is in the use of surface texture. Textural features are extracted from the images to effectively characterize the vegetation. Methods based on Circular Symmetric Auto-Regression, Co-Occurrence Matrix and Local Binary Patterns are selected for the comparative study. The experiments are carried out with image sequences taken from a scaled model of a landscape and a selection from the Brodatz set. A few training images are used to set up the model bases for the three methods. The methods are tested using the same regions from other images of the sequence, and other sequences of images of similar fields. Comparison of the methods is based on the confusion matrix. Sensitivity to variations in flight direction, variations in altitude and luminance conditions are being considered.

Keywords: texture, classification, vegetation, agricultural fields, aerial images, pattern recognition

1. INTRODUCTION

1.1 MAIS project

The research described in this paper is part of the project concerning Model based Analysis of Image Sequences (MAIS). The objective of this project is the design of a complete image analysis system for handling the data (storage, conversion, etc.) and the image sequence analysis [14]. The image sequences are assumed to be acquired by a video camera on board of an airplane, which moves linear over the scene. The position and orientation of the camera are assumed to be known as a function of time. The objects in the observed scenes are agricultural fields, roads, houses, forest etc. The goal of MAIS is twofold. The analysis should firstly yield a digital elevation map (DEM) of the landscape and secondly a classification and recognition of the agricultural fields in the scene. The classes of the agricultural fields are determined by the type of crop, e.g. grass, potatoes or wheat. This paper focuses on methods for classification of agricultural fields.

1.2 Problem description

A common approach in recognition and classification of the agricultural fields is in the use of surface texture. Texture provides essential structural information of regions in an image and is therefore used in many fields of applications. The bio-medical community uses texture to analyze microscopic images of cell structures or tissue samples, the machine vision community to analyze object surfaces. This leads to the objective of the research of this sub-project: "*Select and compare texture-based methods for the classification of agricultural fields in aerial sequences of images*".

Despite of its wide use, a formal approach or precise definition of texture doesn't exist. Granlund and Knutson [6] give the following informal definition: "An organized area phenomenon, composed of a large number of more or less ordered similar patterns or primitives, giving rise to a perception of surface homogeneity". Ehrich and Foith (1978) summarize the following issues in texture analysis:

Given a textured region, determine to which of a finite number of classes the region belongs.

Given a textured region, determine a description or model for it.

[a] For further information –
Z.H. (correspondence): e-mail: Z.Houkes@el.utwente.nl; WWW: http://utelmi01.el.utwente.nl/

Part of the EUROPTO Conference on Remote Sensing for Agriculture, Ecosystems, and Hydrology • Barcelona, Spain • September 1998
SPIE Vol. 3499 • 0277-786X/98/$10.00

Given an image having many textured areas, determine the boundaries between the differently textured regions. The third issue concerns the use of the information from the second issue to perform a segmentation of the image. The segmentation of image sequences of agricultural fields has been investigated by Ma [10]. This paper describes the results of an investigation of the first and the second issue. The modeling step is filled in by selecting a number of methods from literature. The first issue is the subject of this paper.
Another, more practical, problem concerns the sequences of images to carry out experimental research. As part of MAIS a scaled model (fig. 1) of an agricultural scene has been constructed. This model simplifies the acquisition of sequences of images in a controlled environment. The selected methods are to be tested and compared by using these sequences. The comparison will be based on the performance of the methods and their sensitivity towards changing properties in the image sequences, such as altitude, flight direction, and so on.

1.3 Outline of this paper

Section 2 gives an overview of texture based classification methods. By looking at the special demands of this application a choice is made for three different texture based classification methods. The description of these methods is the subject of section 3. Knowing the three methods, section 4 discusses the approach to compare and evaluate the methods. The results of this comparison and evaluation can be found in section 5. Finally, section 6 presents conclusions. Based on the objectives and the conclusions this section closes with some recommendations for further research.

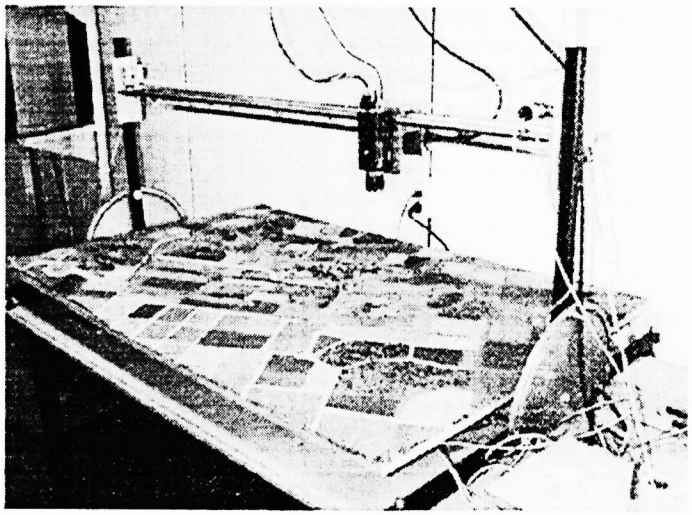


Figure 1: Scale model of an agricultural area and video camera system

In the next section an overview is given of relevant literature. Although this overview is not extensive, it provides a good impression of different methods used in the image analysis. This results in a discussion about texture, different approaches and general possibilities using these approaches. At the end conclusions are drawn from this overview, and which methods should be compared and evaluated.

2. TEXTURE

As stated in the previous section, texture is one of the principal methods used in the image analysis and pattern recognition. Lacking a precise definition only informal definitions are known. One of these informal definitions defines texture as: "*An organized area phenomenon, composed of a large number of more or less ordered similar patterns or primitives, giving rise to a perception of surface homogeneity* [6]" as already given in chapter 1.
Texture, being one of the principal methods in image analysis and pattern recognition, leads to the justifiable choice for the use of texture in this assignment. The existing literature on recognition of natural textures can be divided in three main categories, namely statistical, structural and a combination of both.
Simple shapes will correspond to a primitive spatial event. In this case the property measurement from this shape will provide enough information for a correct recognition of the shape. Classification comes down to comparing the vector of properties measured from the spatial event with the property vector from a prototype representative. This constitutes statistical pattern recognition. Complex shapes, on the other hand, will correspond to a set of primitive spatial events. In this case pattern recognition comes down to comparing each image property vector as well as the relation within the set of property vectors with the sets derived from prototype representatives. This constitutes structural pattern recognition.
On a lower level, texture-based image segmentation and classification methods can be subdivided in two main categories, feature-based and model-based. In feature-based methods, some characteristics of texture are chosen. Regions where these characteristics are relative constant are sought. Segmentation consists of finding the boundaries between these regions, whereas classification is based upon the value of characteristics themselves. Model-based methods hypothesize underlying processes for textures and use parameters of these models for segmentation and classification. In this view model-based methods can be considered as a sub-class of feature-based methods, using model parameters as texture features [6].

2.1 Feature-based methods

In literature feature-based methods are extensively discussed. The discussions include a number of comparative studies of texture algorithms used under different conditions. A number of frequently used algorithms have been discussed and are presented in combination with comparative results. The following algorithms are repeatedly found in literature: the *Spatial Gray Level Dependence Method (SGLDM)*, the *Gray Level Run Length Method (GLRLM)*, the *Gray Level Difference Method (GLDM)*, *Laws' Texture Measures*, the *Power Spectral Method (PSM) and Texture Units and the Texture Spectrum.*

2.2 Model-based methods

A number of different models are used in model-based texture classification. Two of these models, the *Markov Random Field model* and the *2-Dimensional Auto-Regressive model* have been widely used in image analysis. A 3rd model discussed has been developed by Chen, et al [4] at the University of Twente in connection with the MAIS project.

2.3 Discussion and conclusions

The selection of methods for the comparative study is based on the results and conclusions of a number of comparative studies [5 11 12 13] and on the requirements with respect to the expected sensitivity of the results for variations in flight direction, variations in altitude and luminance conditions.

Conners and Harlow [5] performed a theoretical comparison of texture algorithms, with the percentage of overall correct classification as the metric of comparison. In this comparison they included the SGLDM, GLRLM, GLDM and PSM. They concluded that SGLDM and GLDM both outperformed GLRLM and PSM.

Ojala et al. [11] evaluated the performance of the Gray-level difference method, Laws' texture measures, Center-symmetric covariance measures and Local Binary patterns, which are based on Texture Units. In their evaluation they used two image sets, Brodatz' images [2] and a set containing texture images created by Gaussian Markov Random fields. In accordance with the images used in the MAIS project, special attention is given to the conclusions drawn from the experiments with the natural texture images from Brodatz. Ojala et al. [11] found that the Local Binary Pattern outperformed the other methods with the Brodatz image set. The gray-scale invariance of the Local Binary Pattern method is a great advantage too, taking into account the requirements from the MAIS project. The Center-symmetric covariance measures and Laws' texture measures are also not selected, because of the larger sample sizes [11] required by these methods.

Ohanian and Dubes [12] also evaluated the performance of four classes of textural features. They evaluated the textural features by their ability to classify single texture images and did not consider the problem of segmenting an image containing several textured sub-images. Among the evaluated textures were SGLDM, Markov Random Fields and the PSM. Their findings confirm the results obtained in the other evaluations. They concluded the co-occurrence features to perform best. The conclusions drawn by Pietikainen and Ojala [13] just confirm the results from the comparative studies.

The results obtained from comparative studies as summarized in the first part of this paragraph, support our choice for the Local Binary Pattern method. The disadvantage of this method is its rotation-variant. A solution to this problem is discussed in the next chapter where the Local Binary Pattern method is discussed in more detail. In that chapter also the method based on the co-occurrence matrix and the CSAR method are discussed, which is already mentioned as the 3rd and final method to be evaluated. An additional advantage of the choice for these three methods is that these methods cover three different principles: a model-based approach, a 'common' feature-based method and a feature-based method using a spectrum.

3. TEXTURE MEASURES FOR CLASSIFICATION

In this section the three selected texture classification methods mentioned in section 2 are discussed. These three methods are the Co-occurrence Matrix method (COM), the Local Binary Pattern Method (LBP) and the Circular Symmetric Auto-Regressive method (CSAR). This section provides the theoretical background of the methods used in the performance evaluation.

3.1 COM method

The co-occurrence matrix uses estimated values of second-order statistics of the (sub)image. This is expressed by the second order distribution:

$$P(x_1, y_1, x_2, y_2, i, j) \tag{1}$$

This is the probability that pixel at (x_1,y_1) has brightness i and pixel (x_2,y_2) has brightness j. Within a region of uniform texture the distribution is assumed to be stationary. With this condition equation 3.1 simplifies to:

$$P(\delta x, \delta y, i, j) \tag{2}$$

which can be related to the co-occurrence matrix $C(i,j|d,\alpha)$ [7], containing an estimation of the probability $p(i,j|d,\alpha)$ of going from gray level i to gray level j, given the intersampling distance d and angle α. If d is small relative to the coarseness of the texture, the matrix values cluster near the main diagonal, while for large values of d the matrix is more spread out. To limit the computational effort the angles are often quantized to 45° intervals. This yields 4 different co-occurrence matrices. The interested reader is referred to Haralick et al [7] for more information. The co-occurrence matrix, being used in this investigation, is computed by adding the four co-occurrence matrices of the different angles α together and normalizing the result.

The co-occurrence matrix defined in this way has the following properties:

- The co-occurrence matrices are basically symmetric, i.e. $C(i,j|d,\alpha) = C(j,i|d,\alpha)$, except for effects occurring at the edge of the calculation area.
- The size of the co-occurrence matrix is determined by the number of gray levels in the image.
- The co-occurrence matrices are rotational-invariant
- The co-occurrence matrices are not gray-scale invariant.

The number of gray levels often used in images is 256. This results in a co-occurrence matrix of 256 by 256, which corresponds to 65536 elements. To have a statistical reliable estimation, this requires a large minimum region size. In the current application this is not desirable. A solution is found in reducing the number of gray levels in the image through preprocessing. On the other hand, the loss of textural information as effect of reducing the number of gray values must be kept minimal. In this paper the number of gray levels in (sub)images used by the co-occurrence method is reduced to 32. This choice is a good compromise, which is also supported by the investigations of Ma [10].

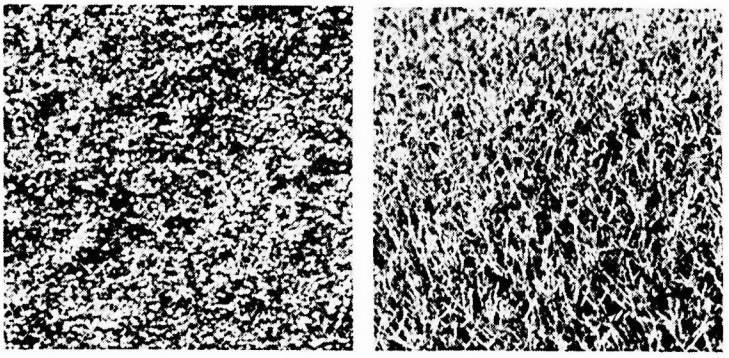
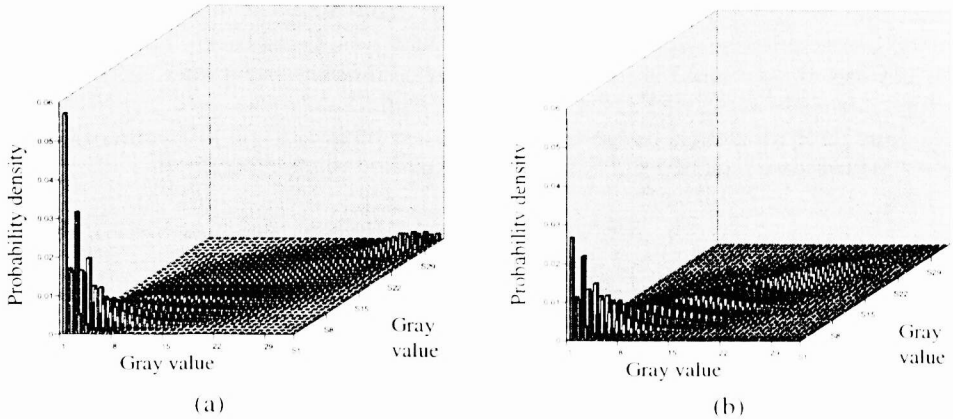


Figure 2: Brodatz images (leaves03 and leaves13)

To show the discriminating capabilities of the co-occurrence matrix the co-occurrence matrices of two different real textures from fig. 2 are computed. The resulting co-occurrence matrices are shown in fig. 3. These co-occurrence matrices show that the number of gray levels in the original images (256) as shown in fig. 3, has been reduced to 32 before calculating the co-occurrence matrices. It can be seen too that both images consist mostly out of dark areas. The most important conclusion however is the difference between both matrices, which makes the method qualified for extracting textural information. A simple difference measure e.g. is the summed quadratic differences at corresponding positions over the entire matrix.



(a) (b)

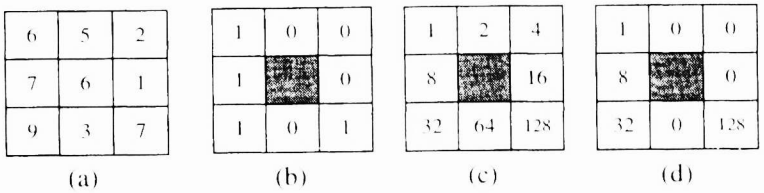
Figure 3: co-occurrence matrices calculated over Brodatz images (fig. 2)

3.2 LBP method

The Local Binary Pattern method was introduced by Wang and He [10]. This new method of texture analysis is based on the so-called texture unit, where a textured (sub)image can be characterised by its texture spectrum. A texture unit (TU) is represented by eight elements, each of which has one of three possible values (0, 1, 2), obtained from a neighbourhood of 3 x 3 pixels. This means that there are totally $3^8 = 6561$ possible texture units describing spatial three-level patterns in a 3 x 3 neighbourhood. The occurrence of texture units computed over a region is called the texture spectrum.

Ojala et al [11] proposed a two level version of this method (LBP), and showed that this performed just about as good as the original three level version. In this version there are only $2^8 = 256$ different texture units to describe a texture.

The principle of this method is shown in fig. 4a and b where a local neighborhood is transformed to a texture unit. The original neighborhood is in this version has been thresholded with the center pixel. The pixel values in the neighborhood are compared with the center pixel. Depending on whether it is smaller or equal/greater than the center pixel value, respectively a zero or one is assigned. The elements of the resulting matrix are weighted with the values of fig. 4c, which results, after multiplying the corresponding elements in the matrices (b) and (c), in fig. 4d. Adding the elements of the matrix (d), this method characterizes this local neighborhood by the number 169.

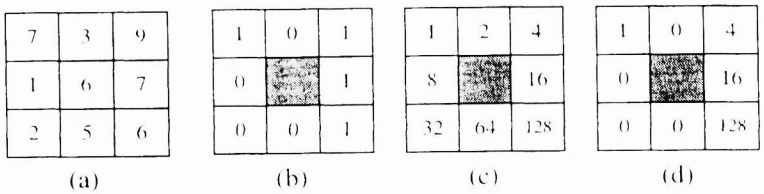


1+0+0+8+0+32+0+128 = 169

Figure 4: Two level version of the texture unit

The number thus obtained gives a description of the relative gray level relationship between the center pixel and its eight neighbors. A histogram containing the texture unit numbers of a (sub)image can be used to characterize a texture image.

The LBP method is gray-scale invariant and computational very simple, but it is not rotation-invariant. In this application rotation invariance is a desired property. In order to reduce the influence of possible rotations within texture a special variant of the LBP method is presented next.



1+0+0+8+0+32+0+128 = 169

Figure 5: One of the rotated neighborhoods for the rotation invariant version of the LBP illustrated for a rotation of 180°.

This variant implies rotating the 3x3 neighborhood three times 90 degrees. This results in four 3x3 neighborhoods, including the original neighborhood. Fig. 5a shows one of them. These neighborhoods are transformed to Texture Units the same way as before (b, c, and d). The resulting numbers are added to separate histograms, characterizing together the texture.

3.3 CSAR method

The third method used in the evaluation is described in detail in Chen et al [4]. Auto-regressive methods in texture analysis are based on the assumption that the gray-level of a pixel in an image depends linearly on the gray level of its neighbours. This idea requires a model to describe this relation and an estimator to estimate the value of the parameters of the model. The mean and the variance of these parameters, or their distribution, are used for classification.

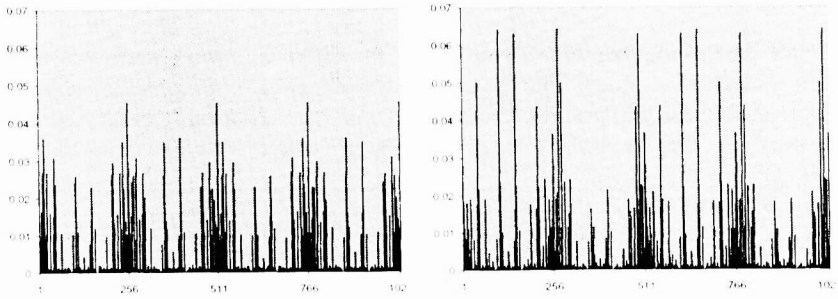


Figure 6: Rotation invariant LBP histograms of the Brodatz' images from figure 2

3.3.1 The CSAR model

Let $\{y(s), s=(s_1,s_2)\in\Omega\}$ be the set of gray values $y(s)$ of a pixel at (s_1,s_2) of an M×M digitized discrete image, with the set Ω defined as $\Omega=\{0\le s_1,s_2\le M-1\}$ and where s_1 and s_2 can only have integer values. It is assumed that $\{y(s), s\in\Omega\}$ is the realization of an underlying circular symmetric auto-regressive (CSAR) random field model described by (3.3), defined over an M×M torus lattice:

$$y(s)=\alpha\sum_{r\in Nc} y(s\oplus r)+\sqrt{\beta}v(s), s\in\Omega \tag{3}$$

where:

- N_c the so called "circular neighbor set", which contains 8 pairs of coordinates: N_c=[(0,1),(0,-1),(1,0),(-1,0),(√2/2,√2/2),(-√2/2,-√2/2),(√2/2,-√2/2),(-√2/2,√2/2)]
- r denotes the coordinates(r_1, r_2) of a the neighborhood pixel;
- v(s) is a correlated sequence with zero mean and unit variance;
- α, β are the coefficients of the CSAR model
- ⊕ is the operation that displaces position (i,j) to position in the neighborhood according the displacement *r*.

The intensity values of four of the neighbors in N_c are known, since their locations correspond to the grid points of the discrete digitized image. However, the values of the other four neighbors are not immediately known, since their coordinates are not integers and do not correspond to the conventional grid point locations. Therefore, their intensities are obtained by interpolation from the gray values of the neighboring pixels. The interested reader is referred to [4] for a more detailed description.

3.3.2 Parameter estimation of CSAR model

To determine the parameters α, β of the CSAR model for a given image, the Relative Least Squares (RLS) approach, which is based on minimizing the sum of squared relative deviations [3] and the Least Squares (LS) criterion are used for parameter estimation.

Leaves 03	Leaves 13
α_{mean} = 0.130358424	α_{mean} = 0.129399464
σ_α = 0.002289229	σ_α = 0.002352755
β_{mean} = 17.21447588	β_{mean} = 14.39087592
σ_β = 2.578733513	σ_β = 4.362283622

Table 1: Results CSAR for Brodatz images (fig. 2)

The same two Brodatz images (fig. 2) are used to show a typical results of the method. From these results (table 1) it can be seen, that there is indeed discriminating information contained in α_{mean} and β_{mean} , although in this case it should be noted that most of the information is situated in parameters β_{mean} and its standard deviation σ_β.

4. THE CLASSIFICATION SYSTEM

In the preceding chapters, we discussed the different texture measures and the three methods being selected for this comparative study. The set up of this comparative study is the subject of this section. The experimental objectives and the texture classification system are successively discussed, followed by the image sets used for the evaluation. The results of the experiments performed with this setup are shown in the next section.

4.1 Experimental objectives

The image sequences, to be analyzed by the selected texture methods, are acquired by a camera, assumed to be attached to an airplane. This configuration has as a consequence that a number of the acquisition parameters will not be constant. The altitude of the airplane, the flight course with respect to an agricultural field and the illumination of the scene may change from one image to another. To obtain good classification results, it is required that the used texture classification method should be, as much as possible, insensitive to these changing conditions. Another important characteristic of the texture classification methods is the effect of the region-size on the classification performance. The objective of the experiments is the comparison of the three methods on their sensitivity to these varying properties mentioned before.

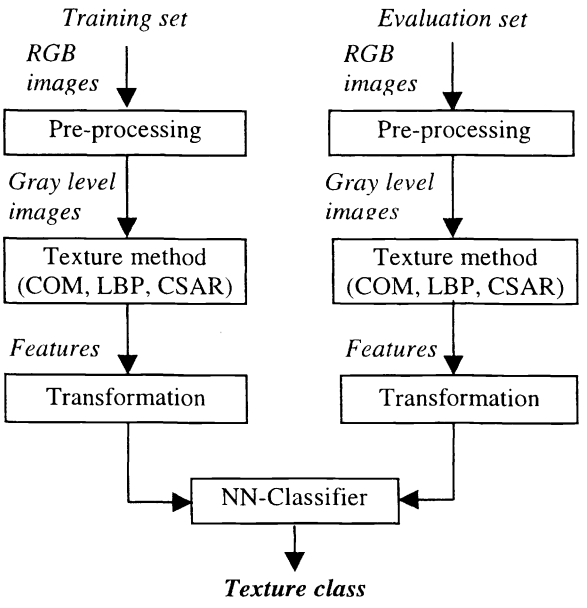


Figure 7: Texture classification system

4.2 The classification system

The diagram of the texture classification system is shown in figure 7. The input to this system is a set of digitized image sequences. Each image set (see 4.3 for the details) is subdivided into two parts, constituting the training set and a number of evaluation sets. All the images are passing a preprocessing stage, where the image format is adapted for the next stage. Images, having a size larger than 512x512, are reduced to 512x512 pixels by cutting out an image of this size. For the co-occurrence method, also the gray-scale is reduced to 32 gray-levels. In the next step, the preprocessed images form the input for one of the texture methods. The output consists of a 32x32 co-occurrence matrix (COM), a histogram (LBP) or α_{mean}, β_{mean} and their variances (CSAR). The following step is a method-dependent 'transformation' stage. The purpose of this step is to provide essential information for the classifier and to reduce at the same time the amount of data to be processed by the classifier.

The classification stage uses the information from the training-set as representatives of the respective classes. This training-set is necessary to adapt the classifier to each possible texture class before the evaluation set is applied to the classification system. The classifier itself is a Nearest Neighbor classifier. This kind of classifier is chosen, because it 'directly' uses the training set, without explicitly estimating probability densities, resulting in a simple, straightforward classifier. The texture classification system has been set up such that the way the rectangular areas (the samples) are taken from the images, is the same for the evaluation process as it is for the training process. This should ensure a good performance of the classifier.

4.3 Image sets

As already mentioned in section 1, because of the lack of real images of agricultural fields, the three methods should be tested on sequences of images acquired from a mock-up. A number of image sequences has been acquired from the mockup, which is equipped with an RGB CCD-camera being guided along a rail by computer control. One of these sequences is partly shown in figure 8. These image sequences are indicated as Image Set I. To be able to compare the three methods using images of natural textures, also a set of nine Brodatz' images is used, two of which were already shown in figure 2. These nine images together are indicated as Image Set II. An additional advantage of the use of Image Set II, will be the possibility of direct comparison of our results with results of other studies, as Brodatz' images are often used in these studies.

4.3.1 Image Set I

Taking into account the objectives of the experiments the following image sequences are acquired:

1. *Sequence of 25 images, which is the reference sequence (qua area, illuminance, altitude etc.)*
2. *Sequence of 25 images as the standard sequence with less illumination (not considered in this paper).*
3. *Sequence of 8 images as the standard sequence taken from another altitude.*
4. *Sequence of 8 images as the standard sequence from another area.*
5. *Sequence of 4 images as the standard sequence with a different flight course.*

For four sequences (2-5) one of the properties has been changed while keeping the others constant. This allows investigation of the effect of the separate parameters. Some of the parameters of these sequences are displayed in table 2. The orientation parameter denotes the difference in 'flight course' of the camera between the reference sequence 1 and sequence 5.

Figure 8: A selection of images from sequence 1

From these sequences one 'training-set' and several 'evaluation-sets' have been derived. The training-set consists of sets of pixels only from sequence 1. This enables the investigation of the experimental objectives by measuring the ability of the classifiers, trained with a 'normal' training set, to classify 'deviant' evaluation sets. In the images of the mockup, 5 different agricultural classes are to be distinguished. These classes are indicated by the numbers 0, 1, ..., 4.

Sequence	Illuminance [lux]	Altitude [m]	Orientation [°]
1	1700 - 1750	325	0
2	700 - 710	325	0
3	1700 -1750	285	0
4	1690-1720	325	0
5	1700 - 1750	325	54

Table 2: properties of image sequences

The images 1, 2, 3, 6, 9, 11 and 13 are used to constitute the training set. From these images fifteen rectangular non-overlapping areas of 64x64 pixels for each class are selected. The choice for this number and size of the areas is made, considering the relation between the total number of areas that can be distinguished in the sequence and the wish to have the areas as large as possible. By choosing a number of areas with a size of 64x64 pixels from the same agricultural field distributed over multiple images, there is accounted for effects varying with the position in the image and which may be caused by lens- and/or camera-errors.

Some exploratory investigations with Sequence 2 (different illumination) show dissimilar results when compared with the other sequences. Error-rates up to 100% occurred. The change in the illumination was so large, that this sequence didn't yield any information concerning the (in)sensitivity of the methods for variations in the illumination. Consequently, sequence 2 has not been considered anymore in the rest of the paper.

From the sequences 1, 3, 4 and 5 four different 'evaluation sets' are derived. The images from sequence 1, which were already used for the training-set, are not used for the evaluation set. One of the experimental objectives listed in paragraph 4.1 is the comparison of the three methods on the basis of the influence of the field-size. The fields vary strongly in size as is depicted in figure 8. Problems may arise, when fields are so small, that a method is not capable of extracting sufficient textural information to give a correct classification. To investigate the influence of the field-size on the performance of a method, the evaluation-sets consequently consist of non-overlapping rectangular areas of 64x64, 32x32 and 16x16 pixels respectively. The number of areas for every class within each sequence can be found in table 3.

Class	Sequence 1			Sequence 3			Sequence 4			Sequence 5		
size	64^2	32	16	64	32	16	64	32	16	64	32	16
0	22	88	352	24	96	384	45	180	720	20	80	320
1	10	40	160	18	72	288	10	40	160	2	8	32
2	30	120	480	26	104	416	17	68	272	17	68	272
3	23	92	368	23	92	368	16	64	256	x	x	x
4	46	184	736	26	104	416	x	x	x	x	x	x
Total	131	524	2096	117	468	1872	88	352	1408	39	156	624

Table 3: Number of areas per sequence and per class of the evaluation sets.

4.3.2 Image Set II

The image set II consists of nine Brodatz' images [7]. The used images are electronic versions of these images, obtained from the website[4] of the University of Bonn. Based on the objective of this investigation, images of the following classes are chosen: grass (2x), leaves (3x), water (3x) and wood-grain (1x). This set consists of natural scenes covering a broad range of scenes, but includes a number of closely related images too. This set opens the possibility to draw conclusions based on a broad range of scenes, but also to look at the discriminative power of the methods by testing the closely related images. The 512x512 images are subdivided into 64 areas of 64x64 pixels each. Sixteen areas, arbitrarily chosen from the set of 64 areas, constitute the training-set for a single class. The other 48 constitute the evaluation set for that class. The complete training set, containing all the nine different classes, consequently consists of 135 blocks of 64x64 pixels. Analogous to Image Set I, three evaluation sets are derived from Image Set II, respectively consisting of 432, 1728 and 6912 blocks, having rectangular sizes of respectively 64x64, 32x32 and 16x16 pixels.

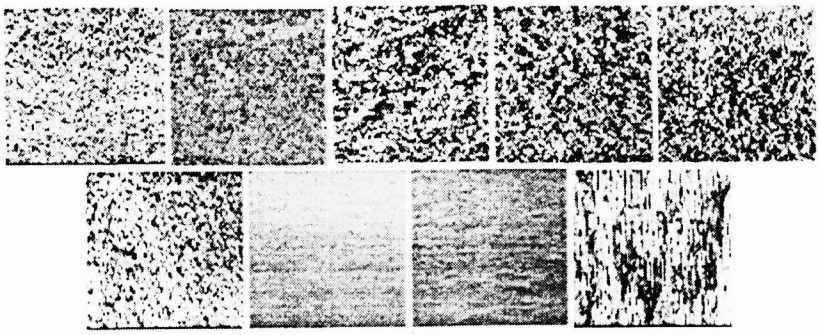


Figure 9: Image Set II. The classes are indicated as 0, 1, ... , 8.

4.4 Segmentation

To generate an evaluation set, consisting of an entire agricultural field, a segmentation algorithm has been applied. The segmentation focuses on an area of pixels within a single field, which should be as large as possible. Such an area should provide more textural information, than the rectangular areas of size 64x64, 32x32 or 16x16 pixels. The segmentation process uses the algorithm based on the co-occurrence matrix as described by Ma [10].

4.5 Classification

As pointed out in the preceding section, each texture method generates its own feature vectors. While not all the feature vectors have the same format, a 'transformation' stage is introduced, adapting the feature vectors to a format suited to the classifier (figure 7). The COM method generates a 32x32 matrix, representing the textural information of the input image. This matrix might be considered as a 1024-sized feature vector. The feature vector generated by the LBP method is also of size 1024. It is a common approach to derive a few features from the co-occurrence matrix or the LBP-histogram [5 8 11 13]. The reason to do so is twofold:

1) reduction of costs by reduction of the amount of data to be processed by the classifier, and
2) avoidance of a decrease of performance because of the ratio between the number of pixels in the evaluation set and the dimension of the feature space at the input of the classifier.

The problem of finding a moment derived from the co-occurrence matrix or the LBP-histogram containing all or satisfying discriminative elements is often mentioned in literature [5 11 12 13]. Using the co-occurrence matrix or the LBP-histogram directly, preventing the loss of any information in a transformation, leaves the dimensionality problem intact. In this investigation the dimensionality problem is coped with by carrying out *feature selection*, followed by *feature extraction*.

As can be seen in the figures 3 and 6, many entries in the feature vectors hardly contain discriminative information. Based on considerations in Van der Heijden [9] and taking into account the number of areas shown in table 3, it is expected that the dimension can be reduced to about 50 elements by using *feature selection*, without loosing substantial discriminative information. The benefit of using *feature selection* is that it's computational much cheaper than *feature extraction*, which makes it suitable to reduce the 1024 dimensions to about 50. The feature selection applied uses the intra- and interclass distances to select the features that produce the best classification results. The further dimension reduction towards 4 is performed by *feature extraction* [9].

The CSAR method reduces the textural information from an image to the parameters α_{mean} and β_{mean} and their standard deviations σ_α and σ_β. It will be obvious that for this method no dimension reduction is required. In this paper, the CSAR procedure is used as it has been developed by Chen, et al [4]. It is however possible to use the distributions of α and β as textural information, in the same way as it is done for the COM and the LBP method.

[4] http://www-dbv.informatik.uni-bonn.de/image/segmentation.html

Reduction of the feature vector size should be dealt with in the way as described above. For a more objective comparison, it would be better to do so.

The pre-processing of the image sets, the computation of the co-occurrence matrix, the LBP-histograms, the preparation of the feature selection for COM/LBP and the computation of the CSAR-parameters are programmed in C. This program yields the data-files for a KHOROS-program carrying out the feature-selection and the feature extraction, immediately followed by the NN-classifier and the computation of the confusion matrices.

5. EXPERIMENTAL RESULTS

In the preceding sections three methods for classification of sequences of images of agricultural areas and an experimental setup have been presented. In this chapter the results of the experiments described in chapter 4 are presented. The results are subdivided by the Image Sets. Following the results of each Image Set conclusions based on the experiments with the particular set are presented. In the next chapter, the general conclusions and recommendations based on the entire thesis can be found.

5.1 Image Set I

An overview of the error-rates in % for Image Set I is presented in table 4 . The error-rates in this table are obtained by taking the sum over all the 5 classes. This table shows that CSAR didn't perform well. The error rates of all the sequences are much higher than for the other two methods. This might be caused by the fact that only 2 features (α_{mean} and β_{mean}) were used for classification, instead 4 as was done for the other methods. A quick investigation of CSAR, applied to Image Set II indicates smaller error-rates (table 5) when σ_α and σ_β (computed from adjacent samples) are used. The results show that this information might be crucial for this method in order to get smaller error-rates. The way the features for CSAR are obtained, might be another reason for the lower performance. It should be investigated whether determining of histograms of α and β, as has been done for the LBP-method, and proceeding in the same way as for the other methods, leads to lower error-rates.

Size	64x64			32x32			16x16		
Seq.	COM	LBP	CSAR	COM	LBP	CSAR	COM	LBP	CSAR
1	0.76	0	12	1.9	2.1	16	7.5	22	27
3	0	0	33	0.43	2.8	37	7.1	22	41
4	17	9.1	23	19	11	24	22	25	27
5	2.6	0	33	3.2	3.9	31	9.5	24	30

Table 4: Overview of error-rates in % for Image Set I

The influence of the sample-size (size of the areas) is smaller for CSAR. This can be attributed to the fact that both COM and LBP profit substantially from the larger size of the areas, which makes the matrix respectively the histogram more dense and therefore more accurate. Another reason for increasing performances with increasing region-sizes is that a larger region will be less sensitive for small differences with respect to the reference sequence.

All the methods performed best on sequence 1 with low error-rates. This might be connected to fact that the training-sets for all the methods were obtained from this sequence. The classification errors made by COM on this sequence show the effect of the mean gray-levels on that method. The errors are mostly for account of misclassifications between the areas 3 and 4, and the areas 0 and 2 having comparative mean gray levels. This also proceeds from the COM scatter diagram [1], where these two classes are also very close. As can be seen in the same figure, absolute gray-levels play no role in LBP and CSAR. LBP classifies faultless with the 64x64 regions evaluation set, indicating its discriminating ability on the one hand and its dependence on the region-size when compared with 32x32 and 16x16 on the other hand. CSAR performs much worse than both other methods as it having great problems with the 'wood'-area 2. At least 80 percent of the misclassifications is on the account of area 2. This clearly shows the principle of the CSAR method as it uses mean values of the calculated properties instead of their distributions.

On sequence 3 (altitude) both COM and LBP perform excellent again, indicating that small altitude variations (10%) have no effect on these two methods. CSAR however decreases drastic in performance compared with sequence 1. This difference is caused by the almost complete misclassification of area class 3. A lower altitude results for class 3 in a lower β (another roughness), bringing the whole group in the scatter diagram (not pre-

sented; the reader is referred to Blacquiere [1]) next to group of class 1.
The results of sequence 4 (another area) show a remarkable decrease in performance for COM. This decrease is almost completely caused by the classification of samples from class 0 as being class 2. Slight differences in illuminance and color between the training field of class 0 and the fields of class 0 in this sequence, result in these erroneous classifications. With LBP, the misclassifications can be put on the account of the classification problems with class 1. CSAR performs poorly on the classification of class 1 and class 3 by mixing them up, resulting in a high error-rate.
Changing the flight course, evaluated with sequence 5, hardly influences the overall results of the three methods. CSAR falsely classifies most of fields of class 2 as belonging to class 4. The field of class 2 in this sequence is more evenly distributed than the training field of class 2. CSAR runs into major problems on these fields, but also COM and LBP perform about 50 percent of their misclassifications in relation with area 2.
LBP is the best method within the framework presented by these experiments. COM performs also good and is with smaller region-sizes preferable above LBP. With this training and evaluation sets CSAR is largely outperformed by both other methods. Losing the information of the distributions and the standard deviations between adjacent regions resulted in problems classifying the more spread out areas such as class 2.

5.2 Image Set II

The total error probabilities (summed over the classes 0 up to and including 8) of the classification (in %) are computed from the confusion matrices. The results are shown in table 5. This table shows that LBP performs considerably better than COM and CSAR for the 64x64 regions, where LBP and COM perform equally well for the smaller region sizes. The problems for COM with region size 64x64 are caused by image class 2. This is the most rough 'black and white'-type of image. The 'black and white' values are hardly present in the list of indices (not presented). This is because a number of images are very close related to each other, requiring most of the elements in the indices list[a] to discriminate between them. The error-rate of 64x64 LBP is mainly caused by wrong classifications of class 0 and class 5.

Method/size	64x64	32x32	16x16
COM	16	24	41
LBP	6.6	22	43
CSAR	53	54	59

Table 5: Overview error-rates in % for Image Set II

Based on the results of the 32x32 and 16x16 evaluation set, it is shown again that LBP requires larger regions for a good performance, of course in relation with the type of images used. The images in the set are closely related, as can also be seen from the statistics of the images. This is revealed by the small number of zeros in the confusion matrices of 16x16 as well as 32x32 of the three methods.

Subregion-size	Error-rate (in %)
64x64	53
32x32	49
16x16	43

Table 6: CSAR with standard deviations

Compared with experiments described in literature, the results for COM and LBP reasonably comply. The error-rates are a fraction higher than the results shown in literature [11 12 13], which might be a result of the choice of the Image Set. The closely related images in Image Set II, yield a more difficult classification problem.

The results of CSAR without the standard deviations are very poor, with error-rates above 50 percent. This is not caused by one or two classes being completely wrong classified, but in general by samples of all the classes being falsely classified into all the other classes. CSAR without the standard deviation in the feature vector is not suitable for these kind of closely related real texture images. The comparison between CSAR with the standard deviations and COM and LBP cannot be made directly, because with the standard deviations added, CSAR uses extra regional information.

Use of the standard deviations

To get an idea of the effects of using the standard deviations σ_α and σ_β in the feature vector of the CSAR model, a region of 64x64 subdivided into 16 regions of 16x16, 4 regions of 32x32, and 1 region of 64x64. Each subregions yields an α and β. The σ_α and σ_β are calculated as being the standard deviations of the respective 16 and 4 α's and β's. By definition no standard deviation can be computed from the single α_{mean} and β_{mean} computed from the single 64x64 region. The results are given in table 6. As can be seen from this table the error-rate decreases when the standard deviations are taken into consideration. The error-rate on the other hand stays well above the error-rate of COM and LBP, especially when taking into account that using the standard deviation actually implies the use of a 64x64 region.

[a] This list is the result of the feature selection, yielding the 50 most discriminative elements.

6. CONCLUSIONS AND RECOMMENDATIONS

This paper describes a comparative study performed on three methods for the classification and recognition of agricultural areas in image sequences. These three methods, COM, LBP and CSAR, were chosen based on their good results in literature and their appropriateness for the project.
The main conclusion is that the LBP method performed better than COM and CSAR with the sets of images used and given region-size of 64x64 or larger. COM performed considerably better than CSAR. With smaller regions, 32x32 or even 16x16, COM performs better than LBP (and CSAR) and should be preferrably used in these cases. The reason for the better performance of COM with smaller areas may lie in the fact that from each 3x3 window COM extracts 8 values as opposed to (standard) LBP only 1 and the LBP used in this paper 4 (correlated) values. CSAR without the standard deviations performs poor and should not be used for the classification and recognition in this way. Adding the standard deviations reasonably improves the classification result, but requires larger region sizes (64x64).
With respect to the experimental objectives, it can be concluded that both LBP and COM experience little problems with most of the changing properties. Changes in the area being considered, altitude of the airplane, or the flight-course had little effect. Investigations of the effect of changes in the illumination on the performance of LBP and CSAR should be carried out again. Although strictly speaking a comparison isn't allowed, but just to get an idea about the results, the performances of the methods with Image Set I are put side by side with those found in literature. The error-rates appear sometimes to be lower than those found in literature. This might be caused by the fact that the training-set is created from only one image sequence, which is in contrast with what usually is done in experiments in literature. The results of Image Set II correspond with those found in literature for comparable image-types.
With respect to the co-occurrence matrix size, a value of 32 is in relation with the required region-size a good choice. The use of a rotated LBP version results in the observed classification-rates with the rotated evaluation set and is therefore a useful expansion.

The most important recommendation concerns the image sequences. Instead of using images obtained from the mockup, real image sequences of agricultural scenes should be acquired and used. Based on the conclusions from this paper further research of the practical implementation on images with real agricultural areas should be performed. As already mentioned, using the distributions of the CSAR parameters instead of their first order statistics should improve the performance of the method. The method is than applied in the same way as the COM and LBP method in this thesis.

ACKNOWLEDGEMENTS

This work was supported by the Dutch Organization of Scientific Research (NWO) under project number projectnummer 612-321-001.

REFERENCES

1. Blacquière, Jan-Friso, Classification of agricultural fields in aerial sequences – a comparative study, May 1998, Report number: 010M98.
2. Brodatz, Phil. Textures. Dover Publications, Inc., 1966.
3. Chen, Haijun. A special least squares method for curve fitting. Proceedings of 1992 IEEE International Conference on Industrial Electronics, Control, Instrumentation and Automation (IECON'92), pp. 1358-1363, San Diego, USA (CA), 1992.
4. Chen, Haijun and Zweitze Houkes. Model Based Recognition and Classification for Surface Texture of Vegetation from Aerial Sequence of Images. August, 1997.
5. Conners, R.W. and C.A. Harlow. A theoretical comparison of texture algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2, No. 3, pp. 204-222, 1980.
6. Granlund, Gosta H. and Hans Knutsson. Signal Processing for Computer Vision, 1995. Kluwer Academic Publishers.
7. Haralick, Robert M. and Linda G. Shapiro. Computer and Robot Vision, Vol. I, 1992, Addison-Wesley Publishing Company, Inc.
8. He, D.C. and L. Wang. Texture features based on texture spectrum. Pattern Recognition, Vol. 24, No. 5, pp. 391-399, 1990.

9. Heijden, F. van der. Image Based Measurement Systems (object recognition and parameter estimation). 1994, John Wiley & Sons Ltd.
10. Ma, Ke Li. Topology preserving segmentation of aerial image sequences. May, 1997. Report number: 007M97.
11. Ojala, T., M. Pietikainen, and D. Harwood (1993). A comparative study of texture measures with classification based on feature distributions. Pattern Recognition, Vol. 29, No. 1, pp. 51-59.
12. Ohanian, P.P. and R.C. Dubes. Performance evaluation for four classes of textural features. Pattern Recognition, Vol. 25, pp. 819-833, 1992.
13. Pietikainen, M. and T. Ojala. Texture Analysis in Industrial Applications. Image Technology. Advances in Image Processing, Multimedia and Machine Vision, Jorge L.C. Sanz, pp. 337-359.
14. Spreeuwers, L.J. MAIS report. 11 September 1996.
15. Wang, L. and D.C. He. Texture Classification using texture spectrum. Pattern Recognition, Vol. 23, No. 8, pp. 905-910, 1990.