Feature extraction techniques to use in cereal classification

Ole Mathis Kruse, IMT

Department of Mathematical Sciences and Technology

2111
2005

Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Problem

- Is it possible to discriminate between different species- or varieties of cereal grains - using image analysis?

Barley

Oat

Wheat

Wheat - Mjølner

Wheat - Bjørke

Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

The images

- Images are taken with a high resolution (1280 x 1024) digital camera.
- Three images (rotated 120°) of each sample.

- Each image splitted in four
  - Larger data material.
  - Easy to implement cross-validation.
- 204 images for species analysis.
- 84 images for variety analysis.

Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Texture analysis

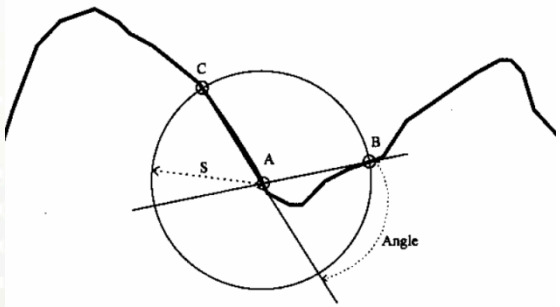
- How to discriminate the samples?
- ~~Single grain analysis~~
  - ~~Measure size, roundness, colour etc. for each grain.~~
  - ~~Extraction of singe grain is difficult.~~

- Texture analysis
  - Finds features for the texture of the image surface.
  - Easy to implement.
  - Several different features available.

Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

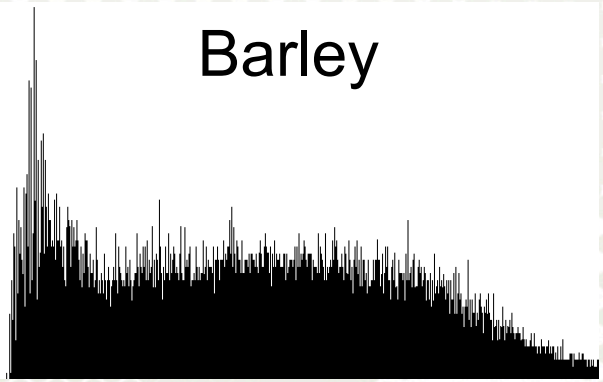
Feature detectors

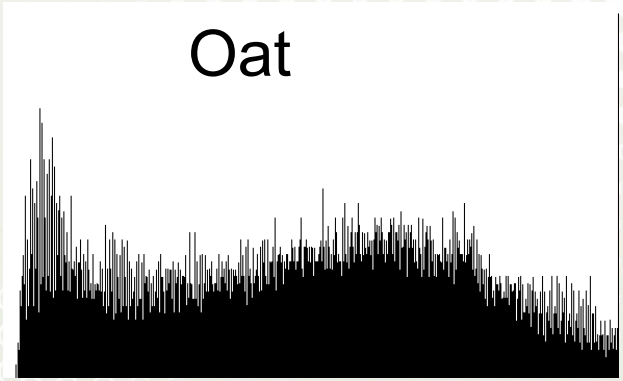
- Some examples of feature detectors
  - Angle Measure Technique (AMT)
    Images are folded out to an intensity vector.
    A circle is placed at random points on the vector, and the angle between the intersections are calculated. Parameters are calculated from these angles.



From: Esbensen, Hjelmen & Kvaal (1996)

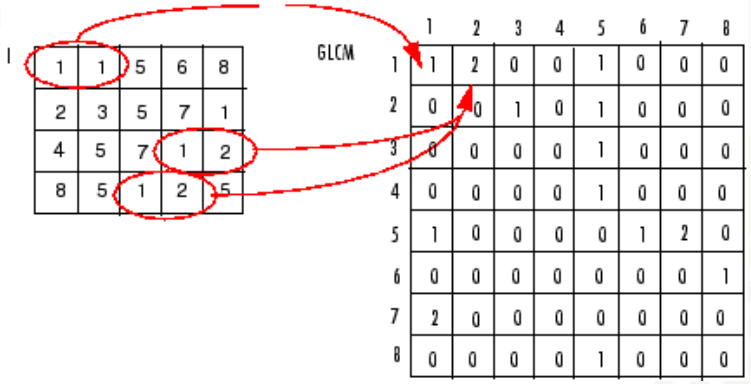
  - Histogram statistics
    Several (~10) statistical parameters are calculated for each image. Some of these parameters are based on the intensity distribution of the images.





Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Feature detectors continued



- Gray Level Co-occurence Matrix (GLCM)
  Builds a new matrix that counts the number of different neighbouring relations. Four statistical parameters are calculated from the GLCM matrix.
- Singular Value Decomposition (SVD)
  From Linear Algebra, a matrix M (e.g an image) can be factorized to the form M = USV*. U, S and V* are matrices which capture different characteristics of the image. We only use the S matrix, which has nonzero values only on the main diagonal. These nonzero values are the singular values of M and can be thought of as scalar "gain controls."
- There are lots of other methods, and also variants of the methods shown here.

Output from feature detectors

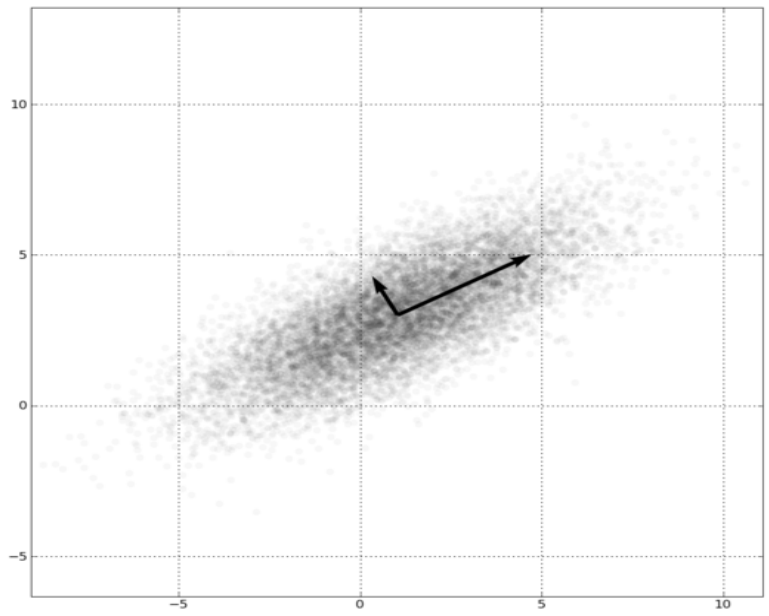
- All methods give a matrix that ranges from 4 to 1000's of columns with one row for each image

	V_1	V_2	V_3	V_4	V_5	V_6	....	....	....	....	....	....	V_k
Image 1													
Image 2													
Image 3													
....													
....													
Image n													

- A complicated dataset with many variables
- Difficult to analyse with univariate statistics (e.g. ANOVA)
- Must use multivariate techniques

Multivariate statistics

- Multivariate statistics analyses *all* variables at the same time and finds patterns that describe the variability
- Covariance between variables is taken into account
- Several different multivariate statistical methods
  - Principal Component Analysis (PCA)
  - Partial Least Square Discriminant Analysis (PLSDA)

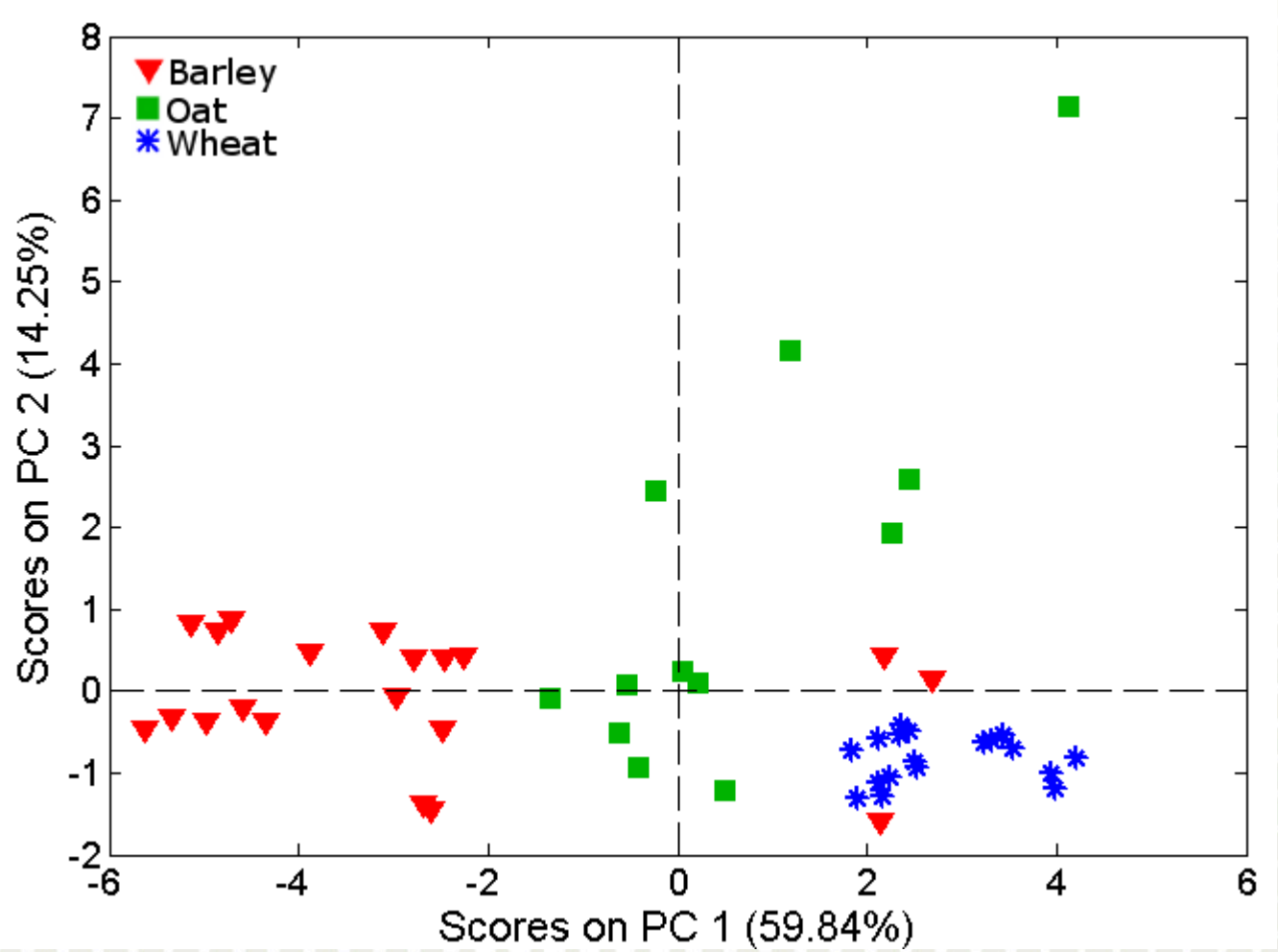


Fra: http://en.wikipedia.org/wiki/Principal_component_analysis

Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Principal Component Analysis (PCA)

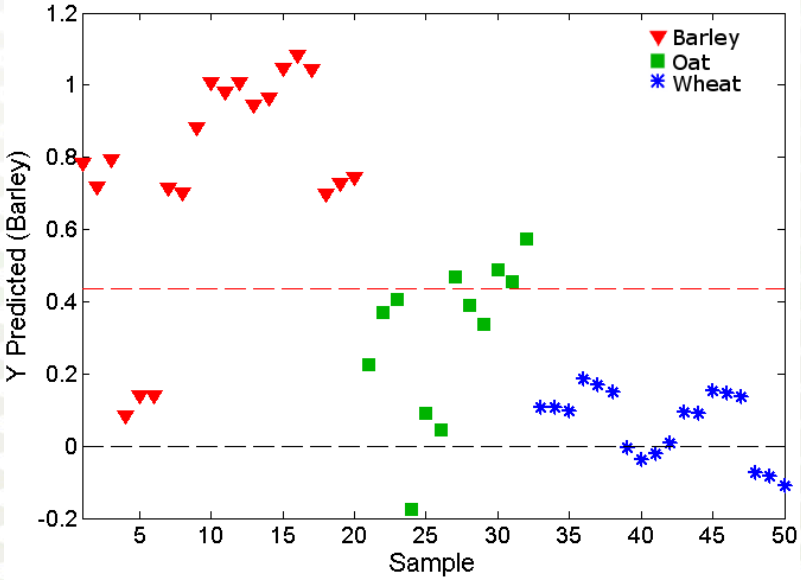
- Typical result from PCA analysis

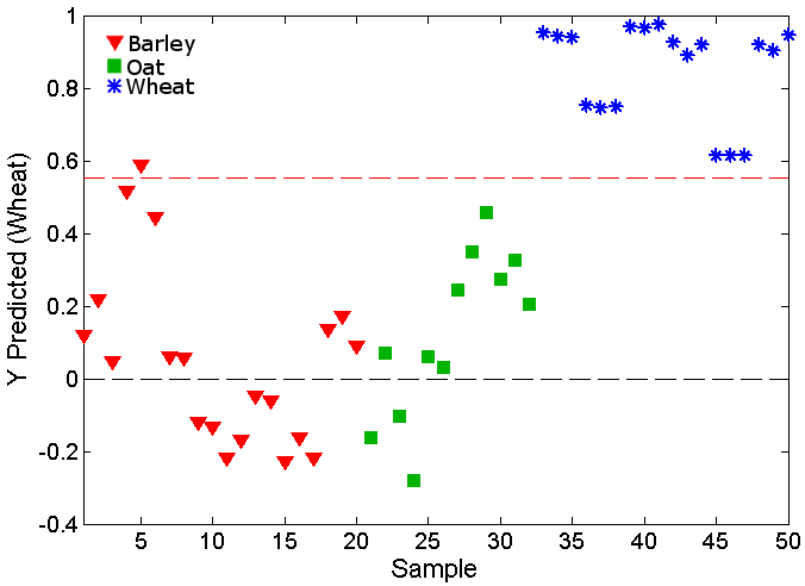


Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Partial Least Square Discriminant Analysis (PLSDA)

- PLS is a multivariate regression technique
- PLSDA allows discrimination/classification of the data





Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Classification results from different feature detectors

- Compare the PLSDA classification with the known species/ variety.
- Calculate % correct classification for each species/variety for the different feature detectors.

	Barley	Oat	Wheat
AMT	90	100	100
Histogram statistics	90	92	83
GLCM	85	50	100
SVD	100	100	100

	Bjørke	Magnifik	Mjølner	Olivin	Polka
AMT	71	25	58	33	83
Histogram statistics	100	0	54	67	50
GLCM	88	0	71	8	100
SVD	38	67	67	0	83

Department of Mathematical Sciences and Technology
NORWEGIAN UNIVERSITY OF LIFE SCIENCES

Summary

- Using 12 images of each sample gives better data quality.
- Feature detectors are used to extract texture information from the images.
- Multivariate statistics are used to analyse the feature data.

- The cereal grain species are classified quite well.
- SVD is the best detector.
- Classification of wheat is strongly dependent on the varieties and the different feature detectors.