Morten Arngren, Industrial PhD. Student - ma@imm.dtu.dk

Informatik og Matematisk Modellering

Classification of Grain Samples.

FOSS

INTRODUCTION

For the production of high quality consumer food products it is essential to have high quality raw materials.

This study will classify grain based on their compounds to identify infected or low quality grain. The analysis will be based on multispectral image processing using 2 layers of classification including both linear and quadratic discrimination.

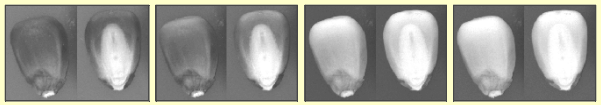
Normal healthy grain

Low quality grain

MULTISPECTRAL GRAIN DATA

A collection of 8 grain samples are used to generate a single multispectral image of 17 bands (Videometer system), where both sides of the grain are used.

Multispectral image sample of 4 grain, front and backside.



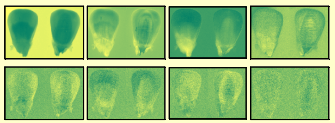
Pseudo RGB image of all the grain

Trainingset
The top 6 grains are used as the training set.

Testset
The lowest 10 grains comprise the testset incl. 1 outlier grain (encircled).

Segmentation
The entire multispectral image is segmented by a scalar threshold applied to the first Principal Component image.

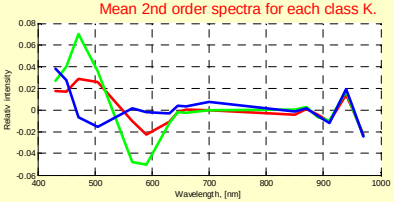
Principal Components, PC1-PC8



CLASSIFICATION OF CONSTITUENTS (1st level)

1. Features
From the 17 band spectra the *2nd order gradient* curve is extracted from each pixel and used as the feature for classification.

Mean 2nd order spectra for each class K.



2. Unsupervised Classification
A simple *K-Means algorithm* is applied to the training set conducting unsupervised classification. From the grain images 3 apparent classes appear and thus the amount of classes is set to K = 3.

The classification is further used as labels to train the supervised model (LDA and QDA).

Unsupervised classification of training data

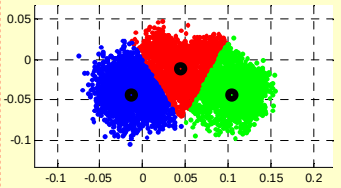
3. Pre-Processing, K-1 dim. reduction
Under the assumption of isotropic noise the data can be reduced to K-1 dimensions without loss of classification performance by projecting onto the hyperplane spanned by the class means.

In the initial pre-processing of the data is hence reduced to 2 *dimensions*.

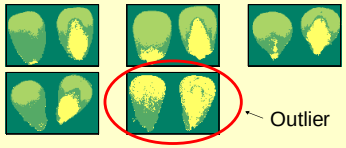
4. Supervised Classification
Two model types are used / tested

- *Reduced Rank LDA (RR-LDA)*
  Projects data using canonical directions to maximize class separation, while constraining to a common unity covariance.

Maximizing class separation

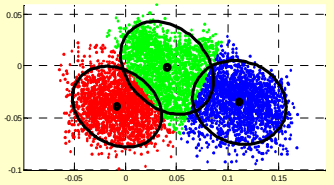


Classified pixels

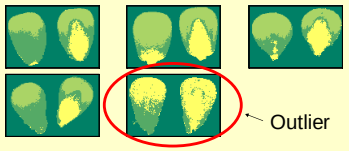


- *QDA*
  Models each class with a full rank Gaussian distribution.

Full rank Gaussian for each class.



Classified pixels

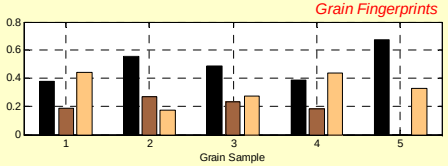


CLASSIFICATION OF GRAIN SAMPLES (2nd level)

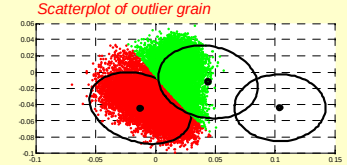
5. 2nd level Classification
From each grain the *distribution of class members (histogram)* learned from the supervised classification is used as individual *grain fingerprints*.

This allows for a 2nd level classification to identify potential low quality grain.

Grain Fingerprints



Scatterplot of outlier grain



6. Outlier Detection
The grain fingerprint can also be used to identify any potential outliers – e.g. by Euclidian distance to a predefined center marking high quality grains.

CONCLUSION

In the classification of multispectral images of grain we have shown how to extract a *fingerprint* feature for each sample based on grain constituents.

Further this fingerprint is used for individual classification and for potential outlier detection.

ACKNOWLEDGEMENTS

I would like to thank my supervisors for technical assistance and corrections & ideas.

- **Professor Rasmus Larsen (IMM/DTU)**
- **Ass. Professor Jan Larsen (IMM/DTU)**

DTU