Quantification of seed performance: non-invasive determination of internal traits using computed tomography

Joelle Claußen[1], Norbert Wörlein[1], Norman Uhlmann[1], Stefan Gerth[1]

[1]Development Center X-ray Technology EZRT, Fraunhofer Institute for Integrated Circuits IIS, Fuerth, Germany

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *The application of the 3D mean-shift filter to 3D Computed Tomography Data enables the segmentation of internal traits. Specifically in maize seeds this approach gives the opportunity to separate the internal structure, for example the volume of the embryo, the cavities and the low and high dense parts of the starch body. To evaluate the mean-shift filter, the results were compared to the usage of a median-smoothing filter. To show the relevance of the mean-shift extended image pipeline an automatic assessment of biological relevant samples was conducted. As data sets maize seeds of 16 different genotypes were used to segment the three different parts of the seed, embryo and the structure due to density within the starch body.*

Keywords. *plant phenotyping, maize, seeds, image processing, mean-shift.*

Introduction:

During the last years, X-ray technology has been applied for the non-destructive visualization of optical inaccessible structures in plants and seeds. Formerly, this technology was only used for medical imaging. Nowadays, it is used as a standard tool in industrial applications for material analysis. With X-ray computed tomography (CT) the 3D volume information of objects can be reconstructed using X-ray projections of the object from different points of view. A conical X-ray beam projects the plant on a 2D flat panel detector. The resulting spatial sampling frequency is determined by the geometrical magnification and the pixel size of the flat panel detector.

Due to the non-destructive nature of CT it is possible to extract morphologic features of seeds and analyze the inner structure, respectively. The extraction relies on 3D density variations and geometric specifications of the traits of interest. This leads to the possibility to quantify the volume, the density – resulting in a virtual biomass – the diameter and aspect ratio of different inner morphologic structures.

Typically, the seed segmentation is a combination of various filters for smoothing, morphological operations and the application of multiple thresholds (Otsu 1979; Kuan et al. 1985). For maize seeds, these techniques are often problematic due to various gray value gradients within the starch body of the seed. As shown in Fig. 1 there is no distinct separation between the embryo and the starch part or even between the internal structures of the starch body. Additionally, the computation of advanced digital traits relying on the morphological structure of the segmentation is not possible.

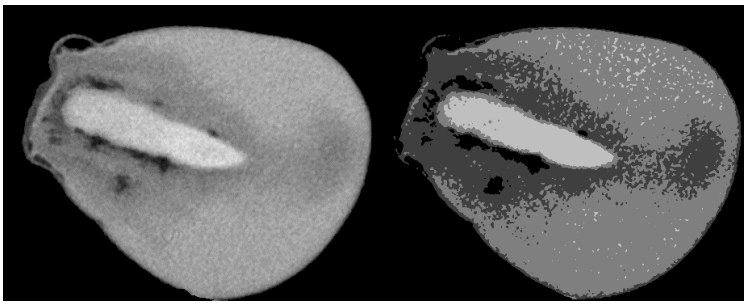

Fig. 1 B73 Gray scale segmentation of a maize seed

Typically, in 2D image analysis mean-shift algorithms are used for clustering different gray value regions while maintaining the morphologic structure. However, a reliable 3D mean-shift algorithm is needed for the application of this approach for CT-data.

Methods:

CT Setup:

CT setup is the CTportable25.50 system developed by the Fraunhofer EZRT in Fuerth. For the seed measurements, a X-ray tube with 40 kV acceleration voltage and 500 µA current was used. The detector has a field of view of 1024 x 1798 pixels and a pixel size of 48 µm.

As tradeoff of resolution and performance one scan consists of 800 images. This leads to a reconstructed volume with a resolution of cubic 39.25µm within 10 minutes for a measurement. To increase the throughput further, the measurements were done with a batch of 100 seeds in one container.

The complete imaging pipeline consists of several steps: (a) 3D-reconstruction, (b) the image separation, (c) the individual seed segmentation using the 3D mean-shift extension and (d) the statistical analysis.

(a) The 3D-reconstruction starts automatically while the measurement is running.
(b) Secondly, the separation extracts the individual seeds within the scanned container as individual 3D volumes.
(c) Subsequently, the image segmentation with the extension of the presented mean-shift algorithm is used to calculate the digital traits for each seed individually. The algorithm calculates 27 different digital traits out of the inner structure. For example, there are simple traits like the whole seed volume, the cavity volume, the embryo volume and advanced traits like the spherical ratio of the seed and the density in different specific areas.
(d) Afterwards, a principal component analysis (PCA) (Mackiewicz and Ratajczak 1993) is applied for the statistical analysis.

Within this publication, only the extension of the segmentation algorithm with 3D mean-shift is in the focus. This was shown to be the key in reliable digital trait quantification.

3D mean-shift extension:

This 3D extension to classical multiple gray value segmentation approaches is based on the mean-shift filter algorithm developed by Dorin and Meer (2002).

The kernel function K(x) is

$$\mathrm{K}(\mathrm{x}) = \mathrm{c}_{\mathrm{k,d}}\mathrm{k}(\|\mathrm{x}\|^2) \tag{1}$$

where $\mathrm{c}_{\mathrm{k,d}}$ is the normalization constant and $\mathrm{k}(\mathrm{x})$ is the profile kernel function.

For the profile k(x) of the kernel function a normalized Gaussian function was chosen within this series of measurements.

This leads to the kernel

$$\mathrm{K}(\mathrm{x}) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}(\frac{\mathrm{x}-\mu}{\sigma})^2} \tag{2}$$

with σ as standard deviation (StdDev) and expected value μ. According to Dorin and Meer (2002) the mean-shift for each position is given through the function

$$\mathrm{m}_{\mathrm{h,G}}(\mathrm{x}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x}_\mathrm{i}\mathrm{g}(\left\|\frac{\mathrm{x}-\mathrm{x}_\mathrm{i}}{\mathrm{h}}\right\|^2)}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{g}(\left\|\frac{\mathrm{x}-\mathrm{x}_\mathrm{i}}{\mathrm{h}}\right\|^2)} - \mathrm{x} \tag{3}$$

where $\mathbf{x}$ is the position in the 2D image, h is the bandwith and g is the kernel function.

Applying this idea to the reconstructed 3D-data, the mean-shift with equation **Error! Reference source not found.** and **Error! Reference source not found.** for each point is

$$\mathrm{m}_{\mathrm{h,K}}(\mathbf{x}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathbf{x}_\mathrm{i}\mathrm{K}\left(\left\|\frac{\mathbf{x}-\mathbf{x}_\mathrm{i}}{\mathrm{h}}\right\|^2\right)}{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{K}\left(\left\|\frac{\mathbf{x}-\mathbf{x}_\mathrm{i}}{\mathrm{h}}\right\|^2\right)} - \mathbf{x} \tag{4}$$

where $\mathbf{x}$ indicates the 3D kernel anchor position in the gray scale volume, I and h are the gray value bandwidths based on the actual gray value at position $\mathbf{x}$, and $\mathbf{x_i}$ are the positions within the kernel. The shift calculated for each point converged with the used kernel function. The convergence was proofed by Dorin and Meer (2002).

The mean-shift filter results were evaluated by comparing them with the results of a median-smoothing filter. The comparison was taken for mean value and the StdDev in a specific location for each part, the low dense part of the starch body (lds), the high dense part of the starch body (hds), the embryo and the transition area between and within these parts (Fig. 2).

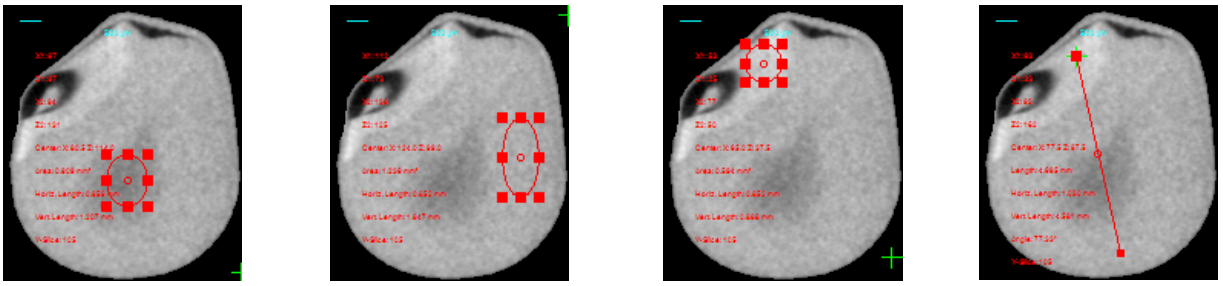

Fig. 2 Positions for comparing mean value and StdDev, lds, hds and embryo and the transition area between each part.

There are two opportunities to modify the result of the mean-shift filter. First the deviation of the Gaussian kernel and secondly the bandwidth of the gray values. For the median filter, different kernel sizes were chosen. Comparing the mean values, there is no big difference between the algorithms (Fig. 3). The mean values of each part can be separated from each other.

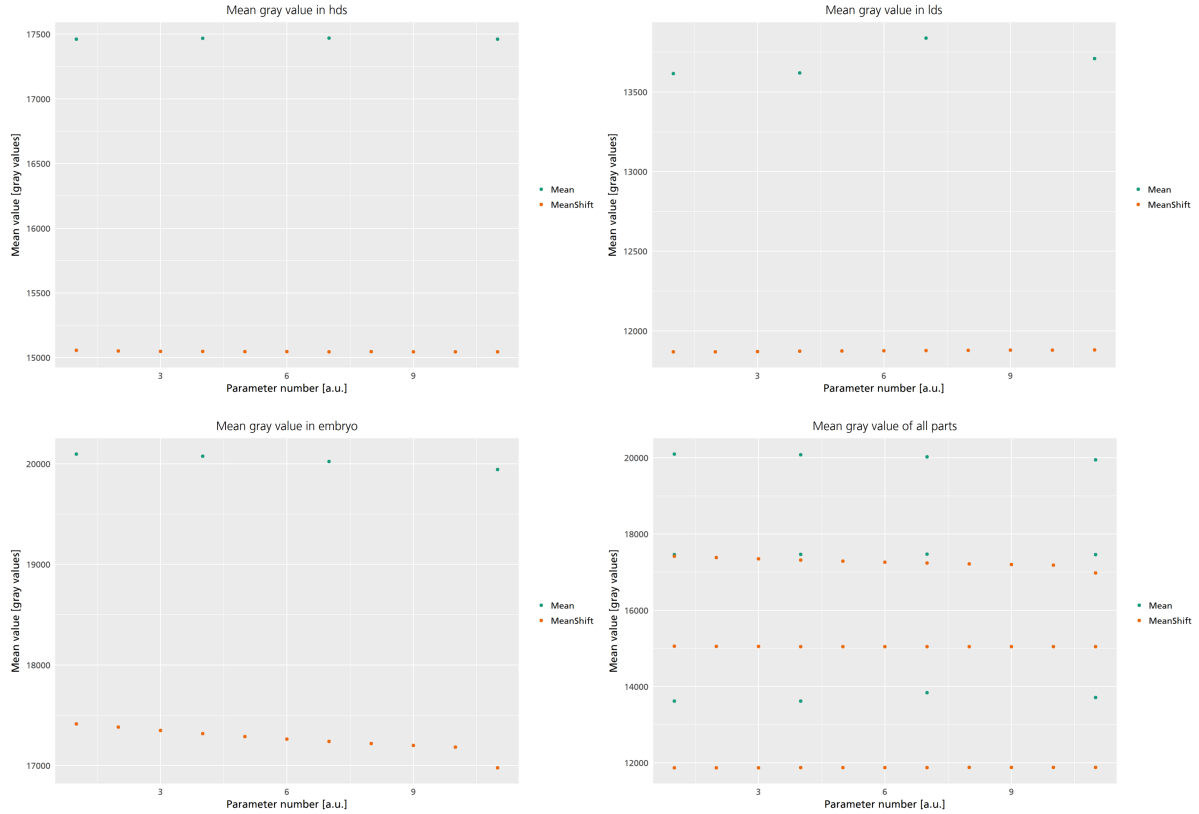

Fig. 3 Mean values of lsd, hds and embryo for different filter kernel and mean-shift parameter.

Comparing the StdDev the diagrams (Fig. 4) show, that the StdDev of the mean-shift filtered data

is much less than the StdDev of the median smoothed data.

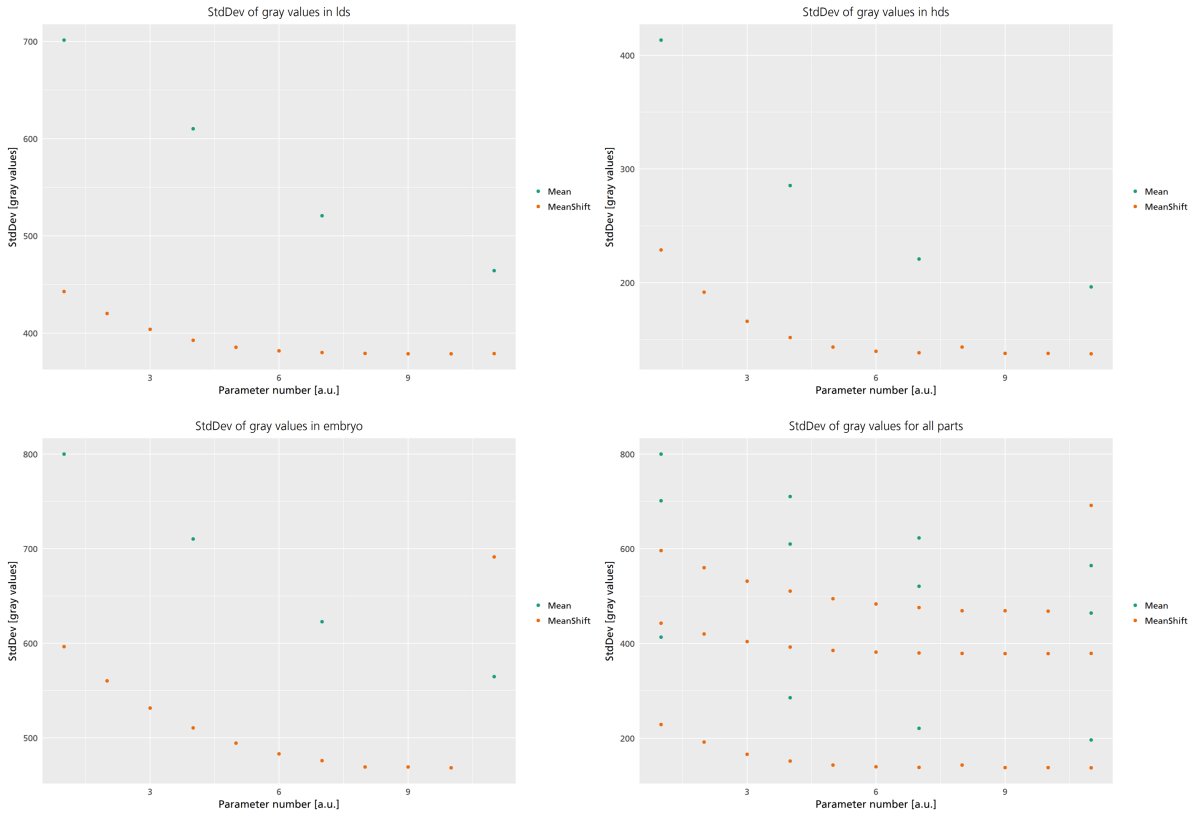

Fig. 4 StdDev of lds, hds and embryo for different filter kernel and mean-shift parameter.

Regarding the transition areas, it is clearly visible, that there is a clearer cut between each part with less noise for the mean-shift data compared to the median-smoothed data (Fig. 5).

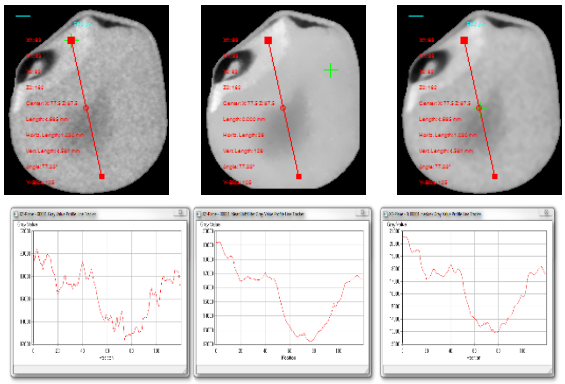

Fig. 5 Transition areas between each seed part for the original, mean-shift filtered and median filtered data with gray value profile.

It is possible to choose a large kernel for the median filter. With this, there is a more or less equal transition type possible (Fig. 6).

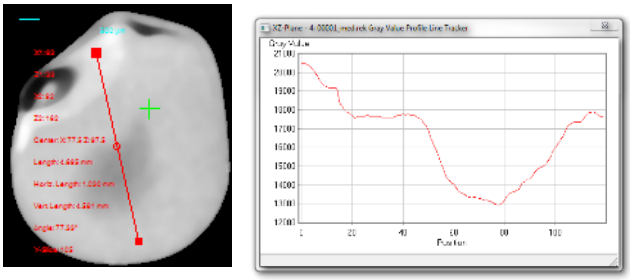

Fig. 6 9x9x9 kernel size for median smoothing with gray value profile in transition area.

The problem with this approach is, that this filter deforms the lds structure. Small excrescence vanishes. In contrast, the mean-shift filtered data structure preserves these parts (Fig. 7).

Fig. 7 Deformation of the lds on the left side with a binary volume of the 9x9x9 median-smoothed data. On the right side, a binary data set of the mean-shift filtered data. Marked deformation in both data sets.

By choosing a higher (greater than 9x9x9) median filter kernel size or false parameter for the mean-shift filter there is a deformation of the single seed parts. Thus, the calculation of advanced digital traits, which are highly dependent on the morphological structure, is not possible with a pure median filter approach.

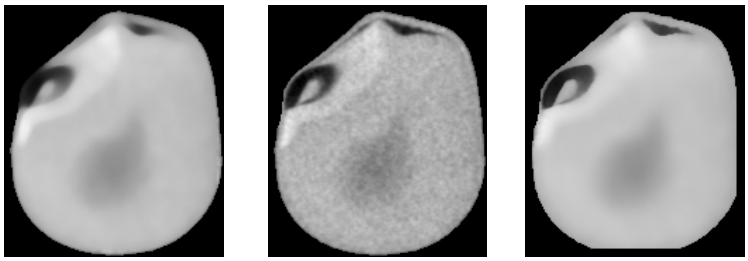

Fig. 8 Large median filter kernel on the left side, original data in the middle and false parameterization of mean-shift filter on the right side. No valid segmentation is possible.

Within this publication, the best parameterization for the mean-shift algorithm for the maize seeds was achieved with a bandwidth of 2000 gray values and a deviation of 10. These values are resulting in a StdDev of 392.5 in the lds, 151.6 in the hds and 510.5 for the embryo part areas. This leads to the filtered and segmented volume shown in Fig. 9.

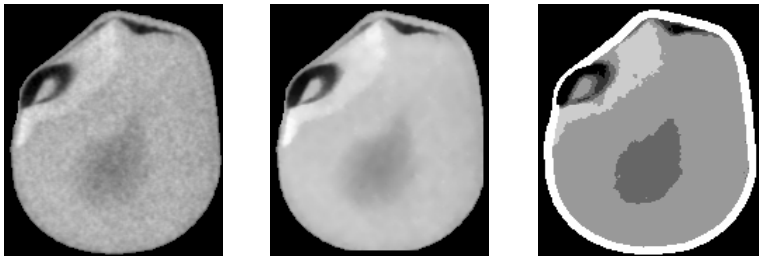

Fig. 9 Original volume with the filtered and segmented data.

Results:

In Fig. 10 the comparison of the segmentation with and without the mean-shift algorithm in the image pipeline is shown. In the new segmentation the structure within the starch body is much more clear and for this it is possible to calculate the digital traits. In the segmentation without the mean-shift algorithm it is impossible to calculate the different digital traits within the 3D-data.

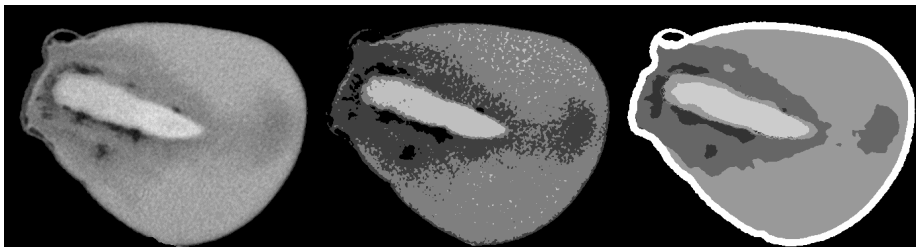

Fig. 10 Left: original seed, Middle: segmented seed based on gray values, Right: segmented seed by using the 3D mean-shift algorithm.

To show the relevance of the 3D mean-shift extended imaging pipeline, an automatic assessment of biological relevant samples was conducted. Within this, 16 different genotypes of maize where analyzed. Due to the possibility to measure one container and to calculate in the imaging pipeline the digital traits of every individual seed it is possible to visualize the variation within distinct genotypes.

Fig. 11 shows the cavity volume as distribution of individual seeds within a genotype. Each boxplot contains a genotype and represents the variation within the genotype. There are some statistical outlier but there are four genotypes which have a distinct higher percentage volume of cavities.

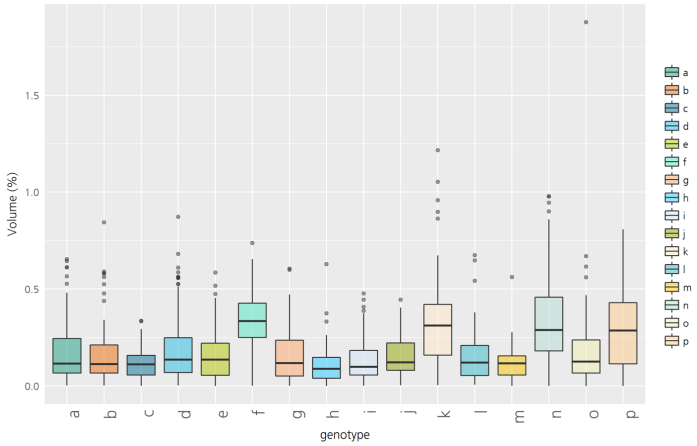

Fig. 11 Percentage of cavity volume for different genotypes.

There is also a significant variation in the lds volume in the seed as shown in Fig. 12

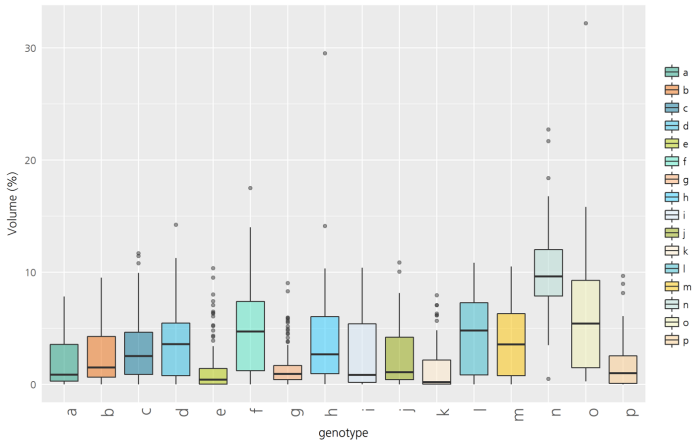

Fig. 12 Percentage of lds volume within the seeds for 16 different genotypes

To see which digital traits from the algorithm combined with the trait of interest clusters, a PCA was made. Two digital traits combined with the trait of interest divides the genotypes in the PCA as shown in Fig. 13.

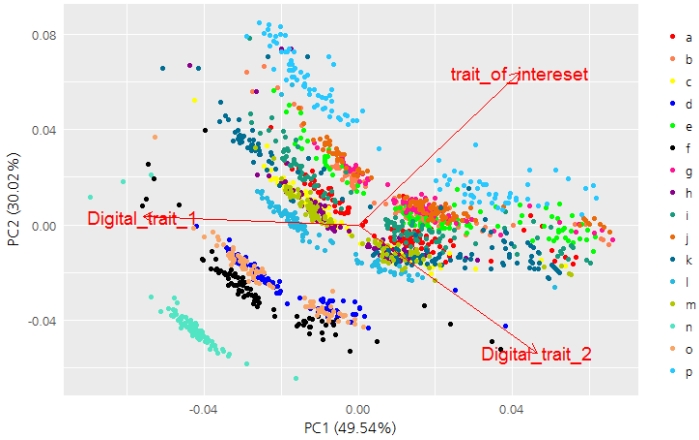

Fig. 13 PCA of two digital traits and the first trait of interest for 16 different genotypes

There is also a second trait of interest, which leads to a clustering of the different genotypes in the PCA as shown in Fig. 14.

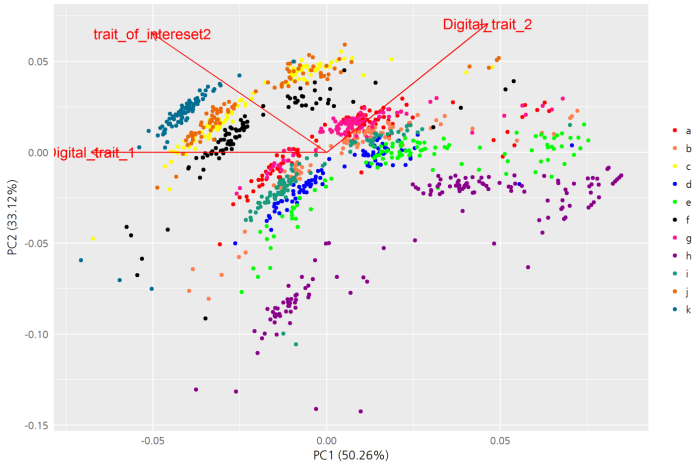

Fig. 14 PCA of two digital traits and the second trait of interest for 11 different genotypes

Conclusion

This leads to the conclusion, that the mean-shift algorithm gives a good preparation for the seed segmentation in comparison to the gray value based segmentations. With these segmented data, it is possible to give relevant answers for biological questions.

Acknowledgements

We thank AG Alumni for provide the seeds of different genotypes and the biological traits of interest.

References

Dorin, C., Meer, P. (2002). Mean Shift: A Robust Approach toward Feature Space Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 24(5), 603–619.

Kuan, D., Sawchuk, A., Strand, T., Chavel, P. (1985). Adaptive Noise Smoothing filter for images with signal-dependent noise. *IEEE Transactions on pattern analysis and machine intelligence,* Pami-1(2), 165-177.

Mackiewicz, A. & Ratajczak, W., (1993). Princopal Componets analysis (PCA). *Computers & Geoscience*, 19(3), 303-342.

Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, man, an Syvernetics*, SMC9(1), 62-66