Hyperspectral system trade-offs for illumination, hardware and analysis methods: a case study of seed mix ingredient discrimination

Carolina Blanch-Pérez del Notario,[a,b,*] Carlos López-Molina,[c] Andy Lambrechts[a] and Wouter Saeys[b]

[a]Imec, Kapeldreef 75, 3001, Leuven, Belgium

[b]KU Leuven, Department of Biosystems, MeBioS, Kasteelpark Arenberg 30, 3001, Leuven, Belgium

[c]Universidad Pública de Navarra, Pamplona, Spain

Contacts

Carolina Blanch-Pérez del Notario: carolina.blanch@imec.be

Carlos López-Molina: carlos.lopez@unavarra.es;

https://orcid.org/0000-0002-0904-9834

Andy Lambrechts: andy.lambrechts@imec.be

Wouter Saeys: wouter.saeys@kuleuven.be;

https://orcid.org/0000-0002-5849-4301

The discrimination power of a hyperspectral imaging system for image segmentation or object detection is determined by the illumination, the camera spatial–spectral resolution, and both the pre-processing and analysis methods used for image processing. In this study, we methodically reviewed the alternatives for each of those factors for a case study from the food industry to provide guidance in the construction and configuration of hyperspectral imaging systems in the visible near infrared range for food quality inspection. We investigated both halogen- and LED-based illuminations and considered cameras with different spatial_spectral resolution trade-offs. At the level of the data analysis, we evaluated the impact of binning, median filtering and bilateral filtering as pre- or post-processing and compared pixel-based classifiers with convolutional neural networks for a challenging application in the food industry, namely ingredient identification in a flour–seed mix. Starting from a basic configuration and by modifying the combination of system aspects we were able to increase the mean accuracy by at least 25%. In addition, different trade-offs in performance-complexity were identified for different combinations of system parameters, allowing adaptation to diverse application requirements.

Keywords: system parameters, hyperspectral, illumination, pre- and post-processing, classification accuracy, convolutional neural networks, spatial–spectral resolution

Introduction

Hyperspectral imaging (HSI), which combines spectroscopy and imaging, is increasingly investigated as a non-destructive, real-time tool for food quality analysis and control.[1] Key factors in hyperspectral systems and in

computer vision systems in general are the illumination system used,[2] the trade-offs in the camera characteristics [spatial resolution, spectral resolution, speed and signal-to-noise ration (SNR)] and the type of data analysis used,[3,4] which can be broadly categorised as either pixel-based (purely spectral analysis) or space-aware (jointly exploiting the spatial and spectral information available).

While many researchers have investigated trade-offs at the level of individual system components, very few have explored the full configuration of HSI systems, either collectively optimising all factors or exploring the potential synergies and trade-offs they involve. For example, Kerekes et al.[5] developed a model for an HSI system in remote sensing relating sensor settings and processing algorithms to a probability of detection. They concluded that the number of spectral bands is the most relevant parameter for a sub-pixel detection application. More recently, Gutierrez et al.[6] focused on system design aspects and trade-offs in a biomedical application, where mostly optical and image acquisition aspects were considered. They focused on time inefficiencies caused by API communications and the effect of Field of View selection on the elimination of spatial distortions. While they did not analyse illumination aspects, they acknowledged their importance as a future line of work.

Indeed, the illumination of an HSI system has a considerable impact on the system performance.[7,8] As a broad-spectrum illumination is desired for HSI, tungsten–halogen illumination is generally used together with hyperspectral systems. With the increasing availability of LEDs with different spectral characteristics, some research has also focused on exploring the suitability of LED illumination systems. To this end, Lawrence et al.[9] compared a traditional halogen system with a LED system for an application of faecal contamination detection obtaining a similar detection accuracy of 99% for both systems. Katrašnik et al.[10] developed a method to compare lighting systems based on spatial–intensity and spatial–spectral non-uniformity measures. Their focus was on avoiding specular reflections, shadows and shades affecting objects with geometry. However, they did not test this approach on a specific application to quantify the performance of these systems. Peter[7] developed several LED ring illumination systems for 680, 780 and 800 nm and tested these for a skin imaging application system, relating viewing angles to the number of LEDs required for uniform area illumination. Sawyer et al.[8] compared the uniformity of halogen and LED-based illumination

systems for a biomedical application but did not report their impact on the discrimination power. Carstensen[11] presented a LED-based system for a food control application in combination with a colour camera but did not benchmark it against halogen systems. The importance of the illumination is also highlighted by Shahrimie et al.[12] and Mishraa et al.[13] where the focus is on compensating the impact of illumination on the plant spectra in a close-range indoor hyperspectral setup, depending on the distance and angle from the light source.

In our previous study, the discrimination power of a halogen-based and a custom LED-based system is compared in a specific application case.[14] It was found that a considerable gain in discrimination power (up to 10% in mean classification accuracy) can be obtained by using a customised LED system, since a more balanced energy distribution is achieved.

Whereas the hardware/camera configuration is somehow limited to a list of parameters and devices, the options for data analysis strategies are very diverse.[3,4] Imaging analysis methodologies can present very different levels of complexity, normally featuring a trade-off between complexity, tractability and need for training data. Recently, there has been a shift in the image processing domain from simple, traceable methods based on classical image processing which require relatively little training data [such as Quadratic Discriminant Classifier (QDC)[15] or Support Vector Machines (SVM)[16]] to high-dimensional, untraceable machine learning methods with heavy training [such as Deep Learning and Convolutional Neural Networks (CNN)]. While the latter tend to outperform the former, at least when finely tuned, the former are still preferred in industrial applications, because they offer a more robust and reliable alternative, e.g. featuring a lower risk of overfitting and/or over-segmentation of the problem space.

CNNs, currently the most popular family of deep learning algorithms, are widely investigated for HSI processing. However, most of the work on the application of deep learning algorithms on hyperspectral data so far has focused on remote sensing applications.[17–19] Moreover, most studies were limited to the spectral (pixel-by-pixel) analysis, ignoring the spatial information in the images. Very recently, a few authors have introduced CNN approaches for joint spatio–spectral analysis in different types of close-range applications. For example, Al-Sarayreh et al.[20] used a 3D-CNN combining spatio–spectral features to detect meat adulteration and

found that it outperformed a pixel-based SVM classification. Wang *et al.*[21] used deep CNNs to detect internal mechanical damage in blueberries using hyperspectral transmittance data reaching classification accuracies between 85% and 90% on test data. In this respect, food control applications are well suited for CNNs thanks to the abundance of ground-truth data samples, which are much needed by CNNs in the training phase. Farooq *et al.*[22] used transfer learning to train CNNs to discriminate weeds from low-resolution hyperspectral images. Gruber *et al.*[23] compared deep learning schemes with more traditional classifiers for a black plastic recycling application based on fluorescence and hyperspectral imaging in the VNIR range. Finally, Gao *et al.*[24] achieved 98% discrimination accuracy on ripeness level of strawberries with CNNs in the VNIR range.

While CNNs can offer high classification performance, they also have some drawbacks in terms of their computational cost, lack of insight in the internal classification mechanism (once the CNN reaches two to three layers), robustness to noise and lack of coherence in results. The latter two problems can be solved by adaptively smoothing the original image prior to classification and/or the labelled image produced by the classifier. In the former case, adaptive smoothing is needed to cope with noise and signal irregularities, while in the latter the smoothing is carried out to remove isolated pixels or non-coherent neighbouring relationships. Some of the best known adaptive smoothing algorithms are Anisotropic Diffusion[25] and its extensions,[26] Bilateral Filtering[27] and Mean-Shift.[28] In the context of HSI, Liao *et al.*[29] combined Extended Morphological Profile (EMP) information with the hyperspectral image by bilateral filtering to improve discrimination in a remote sensing application. Kang *et al.*[30] presented an edge-preserving classification method based on bilateral filtering. They applied bilateral filtering to the classification map obtained from a pixel-based classifier such an SVM, resulting in an increased classification accuracy. An even simpler approach to combine (and simplify) spatial and spectral information jointly is the use of a Median Filter.[31] Median filters are widely used as a simple yet effective denoising method to correct pepper and salt noise in colour and hyperspectral images.[32] Similarly, median filtering can be used on a pixel-wise labelled image to correct locally incoherent pixel misclassifications.

As there is interaction between the different factors involved in optimising an HSI study, the aim of this study was to explore the joint impact of all involved factors on the resulting discrimination accuracy. This is evaluated for a challenging case study from the food industry, namely ingredient quantification in a seed–flour mix. Specifically, we study the use of the illumination systems presented in Blanch *et al.*[14] with cameras of different spatio–spectral resolutions and combine these with the use of different pre-processing and analysis methods. For the latter we compare a pixel-based classifier such as QDC[15] with the CNNs presented in Blanch *et al.*[33] We analyse the impact as well of pre- and post-processing methods such as spatial/spectral binning, bilateral filtering[29,30] and median filtering.[32,34] We also benchmark our work against colour imaging systems restricted to the human-visible spectrum. To our knowledge, this is the first study where the individual and joint impact of all these system aspects (illumination, analysis method, camera spatial–spectral resolution) is investigated for a specific HSI application.

With this study we pursue several goals. First, we want to assess what is the highest achievable performance of our hyperspectral system, starting from a basic system configuration and gradually enabling additional system parameters. Second, we want to study the relative impact of the different system parameters once they are jointly considered. Third, we intend to investigate the available system trade-offs to meet varied application requirements in terms of spatial resolution, classification performance, computing and memory cost, acquisition speed or hardware cost. Finally, we want to investigate the suitability of low spatial–spectral resolution cameras, which initially may seem insufficient to reach high discrimination accuracies for the application considered.

Materials and methods

This section presents all system parameters studied: the camera system used, the illumination systems and the different analysis methods investigated.

Materials

We imaged an identical seed-mix scene containing oat flakes, corn, millet, sesame, linseeds, sesame and pieces of soy grit and soy hull. The seed mix is to be later added to wheat flour for baking. This seed scene is imaged under both halogen and custom LED illumination to obtain the corresponding images for the same scene. This later allows us to force the selection of identical

training samples for all illumination and analysis methods under comparison. We created ground-truth masks for all these ingredients scanned by manually selecting all pixels corresponding to each ingredient with the help of the Gimp software.[35] An example of this ground-truth selection for a hyperspectral image is shown in Figure 1.

We use, in the spectral image, half of the image for the training set and the other half for validation. In the lower-resolution images, since the seeds are not so homogenously distributed, we make sure that training and testing pixels are selected from different groups of seeds in the image. In this sense, we avoid selecting random train/test pixels from the same individual seeds, which may create pseudo replicas. The created image ground truth is used to compute the classification accuracy.

Illumination systems

We used the two illumination systems already presented in our work:[14] a halogen-based system and a custom LED system. For the halogen system we use a basic configuration and a High Dynamic Range (HDR) mode that balances energy distribution by using a different exposure time per wavelength range. To perform HDR, the camera performs scans at several, in our case three, integration times (13, 26 and 39 ms). The highest integration time is used for the band range where less energy is provided by the halogen source, while the lowest integration time is used for the band range receiving the highest energy.

The advantage of the custom LED system with respect to the halogen system is that we can achieve a more balanced energy distribution on our sensor over the full visible and near infrared wavelength range. Moreover,

the illumination is also spatially homogeneously distributed. This translates into a reduced spectral variability, which translates into higher discrimination power for the different ingredient classes. In our previous study[14] we found that both a halogen system with the HDR method and our custom LED illumination system[36] outperformed a halogen system for the QDC classifier and a high-resolution hyperspectral image. In particular, the custom LED system resulted in a higher SNR spectra and noticeable improvement in class discrimination/classification accuracy, with up to 10 % higher mean pixel accuracy.

Figure 2 shows a sketch of the custom LED-prototype with four LED bars. The four identical LED bars are placed in a square formation to achieve an homogeneous spatial distribution of the light. Each LED bar consists of 3 repeated units and every unit contains 16 LEDs with a different peak wavelength between 451 nm and 940 nm. The relative intensity of each LED was tuned to obtain a balanced energy distribution for our sensor. The spectral improvement of the custom LED on some example seeds is shown as well.

In this work, we evaluate the impact of the custom LED system on the classification performance obtained with other two Snapshot cameras with reduced wavelength ranges. In addition, we now consider the use of additional pre-processing methods (binning, median filtering) and analysis methods which jointly exploit spatial and spectral information (bilateral filtering, convolutional networks).

Camera systems

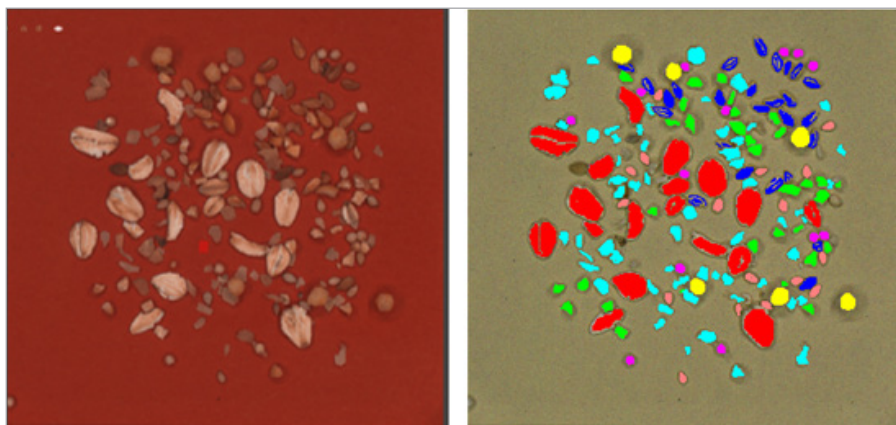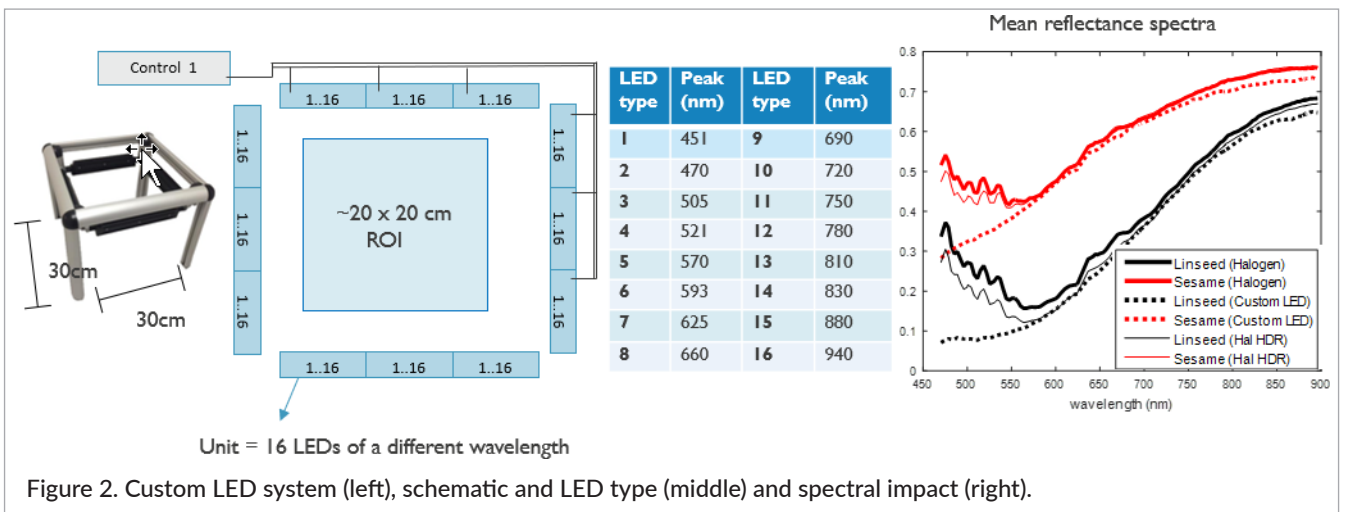We use the Imec Snapscan VNIR range camera system,[37] shown in Figure 3: a camera system concept that

Figure 1. Colour image of seed mix (left) and ground-truth mask selection for all ingredients (right).

Figure 2. Custom LED system (left), schematic and LED type (middle) and spectral impact (right).

combines the high spatial resolution and spectral resolution of line scan hyperspectral imaging technology.[34] It can acquire datasets for a static scene as easily as with a snapshot camera. There is no need for any external scanning movement: scanning is handled internally, using a miniaturised scanning stage.

Full hyperspectral images can be acquired within 10 s. The maximal RAW spatial resolution that can be reached is 3650 × 2048 pixels (7 Mpixels), with a spectral resolution of 150+ spectral bands within the 470–900 nm wavelength range. Some industrial applications, however, require portable and fast image acquisition even at the cost of a reduced spatial and spectral resolution. For this reason, we also test an Imec snapshot camera with a mosaic layout sensor where the full hyperspectral image is acquired for all bands simultaneously in one shot.[38] We have two types of the Mosaic camera. A Mosaic NIR camera with spatial resolution of 409 × 216 pixels, and 25 bands evenly spread in the 600–875 nm range and a

Mosaic VIS camera with spatial resolution of 512 × 256 and 16 bands in the 460–630 nm range. The use of a Mosaic NIR or VIS camera allows even faster image acquisition (potentially 180 frames per second) at the cost of a reduced spatial and spectral resolution.

For our high-resolution Snapscan images there are around 150,000 ground-truth pixels. Approximately 40 % of the pixels are randomly selected from the training half image. For our binned Snapscan or low-resolution Mosaic images there are 10,000–35,000 ground-truth pixels available and, therefore, up to 80–90 % of the pixels of the training set are used.

Analysis methods

The different analysis methods considered are summarised in the processing pipeline illustrated in Figure 4. The first pre-processing method that can be applied onto the input image is a denoising step, which can be implemented by either median filtering with a 3 × 3 or

Figure 3. Snapscan VNIR system (left) and Snapshot mosaic VIS sensor (right).
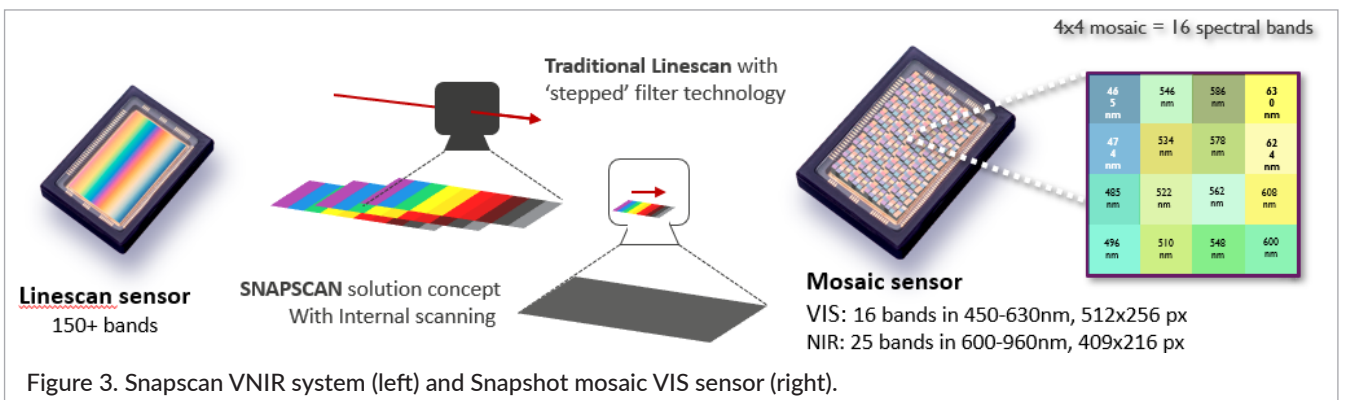
5 × 5 window or by spatial/spectral binning of the original image. Another type of pre-processing method that can be applied next is feature extraction to reduce the input dimensionality. In this study, LDA (Linear Discriminant Analysis) was selected for this purpose, based on our previous study.[39] This step is applied prior to the training of the classifier, for which we test two different classification methods: the first is the QDC,[15] a pixel-based classifier exploiting only spectral information per pixel. The second method is a CNN,[33] which is a more image-based method that jointly exploits spatial and spectral information. The final parameter or processing step is a post-processing method, which acts on the classified image with the purpose to correct some pixel misclassifications by smoothing or imposing spatial coherence on the obtained classified image. For this purpose, either bilateral filtering,[30] median filtering[31] or both were evaluated. Our performance metric is the pixel classification accuracy as percentage of correctly classified pixels. The mean pixel classification accuracy for all seeds as well as the pixel accuracy for the worst performing seed class are provided.

The feature selection and classification model building were performed in the PerClass software.[40] For the remaining analysis methods (median filtering, binning and bilateral filtering), we coded our own Matlab scripts or used the image processing toolbox in Matlab version R2015.[41] The remainder of this section explains all these methods in greater detail.

Pre-processing methods for denoising and feature selection

Median filtering

Median filters are a simple yet effective method for denoising colour and hyperspectral images by correcting, for instance, pepper and salt noise. In this sense, a median filter can be applied spatially to denoise the hyperspectral image, as well as on a classified image to correct for small pixel misclassifications. We analyse its impact for pre-processing when applied as a 5 × 5 filter on the Snapscan image or as a 3 × 3 filter in the Mosaic images. A lower filter size is selected on these lower spatial resolution images to minimise the image blurring.

Spatial and spectral binning

Another method to increase the signal-to-noise ratio in our hyperspectral image is to implement spatial or spectral binning since averaging over pixel spectral values reduces the noise in the spectra. However, binning samples also reduces the number of spatial or spectral samples available, this is, the effective spatial and spectral resolution available. This can have a noticeable impact when we are dealing with low spatial resolution images (e.g. Mosaic cameras) or with small objects (e.g. seeds of few pixels size) in a high-resolution image from the Snapscan camera. Therefore, binning may only be a valid alternative for denoising when the pixel size of the smallest objects is big enough.

For the first camera system, the Snapscan (1088 × 1048 pixels), we want to compare the different complexity–performance trade-offs for different spatial–spectral resolutions of our seed image. We create these different resolutions by binning the full-resolution image both spatially and spectrally by either a factor of two or a factor of four. For our Mosaic images (maximum of 512 × 256 pixels) we do not consider binning since the original spatial resolution is lower.

Linear Discriminant Analysis (LDA) or Linear Discriminant Classifier (LDC)

This method is used to reduce the feature dimension since it identifies the components (linear combination of the variables) with the highest information. LDA is a supervised method, assuming normal densities for
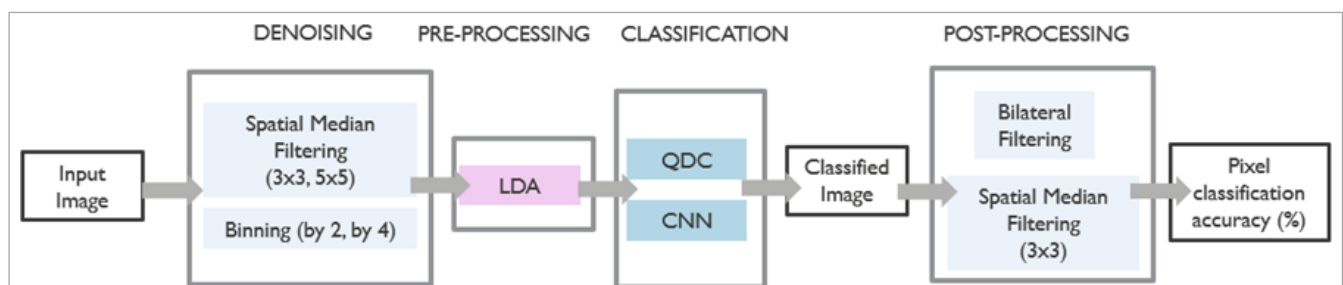

Figure 4. Processing pipeline.

the classes, in which the new features or components maximise the class variation. In our previous work on seed ingredient discrimination[39] we observed that for this application LDA provided better performance than Principal Component Analysis (PCA)[42] as the feature selection method. For this reason, we select LDA as the feature selection mechanism for both pixel-based and image-based (CNN) classifiers considered in this study. It is computed with PerClass[40] software by finding a projection that separates each class from all others. The number of features obtained in this implementation is equal to the number of classes we are considering, minus one. In our case, with 10 classes we compress the initial 147 bands to 9 new spectral features.

Classifiers can be broadly categorised into pixel-based classifiers, exploiting the spectral information per pixel, or image-based classifiers, jointly exploiting spatial and spectral information. To compare the impact of both types of classifiers as system parameters we use the following classifiers, as implemented in PerClass software. Both classifiers are applied after the LDA pre-processing step. **Quadratic Discriminant Classifier (QDC):** Also known as Quadratic Discriminant Analysis[15] assuming as well normal densities for our spectral classes. Our work[39] shows that for this application, QDC is the best performing classifier among the pixel-based ones and, therefore, we select it as the pixel-based classifier.

Convolutional Neural Networks: We use 3D CNNs, presented in our previous work[33] to extract simultaneously both spatial and spectral features from the hyperspectral image. We first reduce the hyperspectral image dimensionality by applying LDA. From the initial 147 bands we obtain 9 LDA bands that are input to the network. This reduces the input dimensions to the network and with it the network complexity and computational time while still retaining the most salient features.

The CNNs are then composed of two convolutional layers. Each of them followed by their corresponding batch normalisation layer and Recitified linear (Relu) units. A "Batch normalisation" layer uses statistics of individual batches to re-normalise outputs of the previous convolutional layer. It does not have any parameter and does not alter network geometry. A "Rectified linear" unit is a simple transfer function that turns all negative values to zero and lets all positive values pass through. It is known to significantly improve convergence speed. We then add two fully connected layers at the end, one with 50 hidden units and the final one with 10 (identical to the number of classes). While the inner convolutional layers are performing the so-called feature extraction from the input images, the last fully connected layers are performing the classification of the corresponding features into one of the output labels or classes.[43]

We base our CNN parameter selection in accordance with previous work[17,18,20] and opt for 16 convolutional filters of size 5×5, for an input block image of 7×7, a learning rate of 0.005, batch size of 100 and a maximum number of 40 iterations. We see experimentally that these parameters work well while still limiting the complexity of the network.

To choose the input block size we evaluate the impact of the selection of different image block sizes on the final classification accuracy. For this purpose, we test the use of image block sizes of 1×1, 3×3, 5×5, 7×7, 9×9 and 11×11 pixels on the full-resolution Snapscan image (1048×1048 pixels).

The size of the convolutional filters chosen for the CNN network for the different input image block sizes are given in Table 1. The larger the image input block size taken the larger convolutional filter we can use, the limit being the input block size. We describe the convolutional filters here as 2D, but they have a third dimension, given by the number of spectral bands, in this case the nine LDA bands. In the case of considering a 1×1 block size, a single pixel, we want to exploit only the spectral information per pixel and not the spatial information or any spatial features. This is interesting to evaluate the amount of information which is present in the spectra or present in the spatial features in the image. For any larger image block size, the CNN uses the information from a

Table 1. Input image block sizes versus convolutional filters used.

Block size	1×1	3×3	5×5	7×7	9×9	11×11
1st Conv Filter	1×1	3×3	3×3	5×5	5×5	7×7
2nd Conv Filter	1×1	1×1	3×3	3×3	5×5	5×5

pixel and its neighbouring pixels, therefore, the spatial information is exploited as well.

The larger the block size the more information we are giving as input to the network, in terms of input pixels. The network was trained with the same input pixels/training examples, which are the centre pixels of each input block. For instance, for a block size of 3×3, this corresponds to inputs of 9 pixels for every training example, resulting in a total of $3 \times 3 \times 9017 = 81153$-pixel data. For the comparison we use the same number of network iterations for all block sizes and the same network architecture.

Post-processing techniques

We consider two different techniques to smooth the labelled images produced by the classifiers. The first and most simple is median filtering, where median filters of 5×5 pixels are applied. The second technique is Bilateral Filtering. We used the Bilateral Filtering method as described in Kang *et al.*[30] In this method, first the probability maps from the classifier for every class label need to be extracted. This probability map is then bilaterally filtered based on a reference image. In Kang *et al.*[30] the reference image was created from the first three PCA components. In our case, we used as reference image the LDA bands, since in our case this shows better performance than PCA. This is in accordance as well with LDA performing better as a pre-processing technique for the QDC classifier than PCA.

The probability maps per pixel are then bilaterally filtered according to the spatial and spectral similarity of a pre-defined block size around the pixel in the reference image. This way, the joint bilateral filter is based on the widely used Gaussian filter, considering the distance in the spatial domain and the distance or similarity in the spectral domain. The spatial and spectral distances are defined using two Gaussian decreasing functions, as described for the Joint Bilateral Filtering process in Kang

et al.[30] The main parameters to tune the bilateral filtering are δ_s and δ_r, defining the decay of these Gaussian functions, or weight decrease with respect to spatial distance and spectral distance (in terms of intensity similarity) respectively. In this sense, δ_r defines how the pixel weight decreases with the intensity difference while δ_s defines the size of the local window used to filter a pixel: Block size $= (2\delta_s + 1) \times (2\delta_s + 1)$.

We explored the performance of these two parameters and set for the following parameter values according to the different spatial resolutions considered (Table 2).

As we can see in Table 2, different δ_s and block sizes are used for the different image resolutions considered. This can be related to the size of the seed spatial features in the different image resolutions. Table 3 shows how the approximate seed sizes for the smallest seeds relate to the image resolution used. We can see that the biggest seed sizes can be obtained for the Snapscan original and binned by 2, and the Mosaic VIS. The Mosaic NIR camera has only slightly lower spatial resolution than the Mosaic VIS. However, in our experiments the Mosaic NIR was placed slightly further from the scanned objects, and this causes it to image the seeds with roughly half the size of those in the Mosaic VIS.

We can see that the smallest seed size in pixels is related to a good performing value of block size where bilateral filtering around the central pixel is performed. This way, block sizes of 7×7, 49 pixels, are considered suitable for all image resolutions where the seed sizes are at least 100 pixels. For the Mosaic NIR and the Snapscan binned by 4, block sizes of 25 and 9 pixels, respectively, perform better since they correspond to a portion of an individual seed. In this sense, performing bilateral filtering over a too-big block size would not be advantageous since the pixel labels would differ strongly within the same block size when a block size comprises different adjacent seed types.

Table 2. Parameter selection per image size: input block size, and Gaussian decay functions δ_s (wrt spatial distance) and δ_r (wrt spectral distance).

Camera image	Image size	δ_s	δ_r	Block size
Snapscan original	1088×1048 pixels	3	4	7×7
Snapscan binned by 2	544×524 pixels	3	4	7×7
Snapscan binned by 4	272×256 pixels	1	0.4	3×3
Mosaic NIR	409×216 pixels	2	0.4	5×5
Mosaic VIS	512×256 pixels	3	4	7×7

Table 3. Approximate seed sizes in pixels for the considered image resolutions.

Image	Spatial resolution	Sesame seed size	Millet seed size	Linseed size	Block size
Snapscan original	1088 × 1048 pixels	~500 pixels	400 pixels	~800 pixels	7 × 7 = 49
Snapscan binned by 2	544 × 524 pixels	~125 pixels	100 pixels	~200 pixels	7 × 7 = 49
Snapscan binned by 4	272 × 256 pixels	~32 pixels	25 pixels	~50 pixels	3 × 3 = 9
Mosaic NIR	409 × 216 pixels	95 pixels	87 pixels	205 pixels	5 × 5 = 25
Mosaic VIS	512 × 256 pixels	~200 pixels	205 pixels	350 pixels	7 × 7 = 49

Moreover, next to previous parameters we need to define our distant metric from pixel i to neighbouring pixel j, $\|i - j\|$. For simplicity we choose it to be the Chebyshev metric.[44] In our image, if the points i and j have Cartesian coordinates (i_1, i_2) and (j_1, j_2), their Chebyshev distance is defined by Equation (1) as:

$$D_{Chebyshev}(i,j) = max\left(|i_1 - j_1|, |i_2 - j_2|\right) \qquad (1)$$

Benchmark RGB colour imaging

To benchmark HSI in the VNIR range (460–900 nm) against regular colour imaging cameras we create RGB images from our hyperspectral image. This way, we can compare both imaging systems under the same system conditions: illumination, lens and identical sample selection.

To obtain the corresponding RGB image from the hyperspectral image we use XYZ parameter computation[45] from the Snapscan hyperspectral image. The CIE XYZ colour space is a device-invariant representation of colour. From the given spectrum, S, the illuminant function I (here assumed to be the E illuminant standard of a theoretical equal energy radiator[45] and the CIE colour matching functions), we can obtain X, Y and Z colour values. We can then transform these X, Y, Z values to RGB colour components.[46]

To benchmark hyperspectral with respect to colour imaging we applied the same type of pre-processing, classification and post-processing techniques to the RGB images as we applied to the hyperspectral images. The only difference is that, since we only have three bands in RGB images, there was no need to apply feature reduction with LDA. Therefore, the CNN was directly applied on the three RGB channels. Since the spectral information was limited to these three broadbands, input blocks 11 × 11 were used to exploit, as much as possible, the spatial information. The size of the convolutional filters was then chosen as given in Table 1.

Results and discussions
Impact of illumination system

We compare the use of different illumination systems: halogen system, halogen with HDR and custom LED system on the pixel classification accuracy for the QDC classifier and the Snapscan camera image. No other pre-processing or post-processing methods were used at this stage. Both mean pixel accuracy and minimum pixel accuracy (for the most challenging ingredient) are increased by better balancing the energy of the illumination system. This way, using HDR functionality on top of a halogen system increases mean accuracy from 70.4 % to 73.2 % and up for the most challenging class from 45.9 % to 50.6 %. The custom LED system achieves the highest performance increase resulting in a 10 % improvement in mean accuracy (reaching 81 %) and up to 14 % increase for the most difficult ingredient (59.5 % versus 45.9 % of halogen). This happens thanks to its more homogeneous spectral and spatial distribution, which is seen in a factor of six reduction of the standard deviations of the spectral intensity over the white tile.

Impact of classifier and camera system

We analyse the effect of the image block size on the classification performance of the CNNs for the Snapscan camera system under halogen illumination. To obtain a fair comparison, the same number of network iterations and the same architecture were used for all block sizes. We observe that for a 1 × 1 block size the CNN performance increases the mean accuracy by 3 % with

respect to the QDC classifier, even though no spatial information is exploited yet. With an increased block size, we increase further the mean and minimum pixel accuracies. The largest performance increase, 11 % in mean accuracy, happens when going from a block size of 1 × 1 pixel to a 3 × 3 block size, since the network can then start to exploit the spatial/spectral information of surrounding pixels. From 3 × 3 block size up to a 11 × 11 block size, the performance increase is less noticeable, but there is still a total gain of around 5 % in mean accuracy and 8 % for the minimum accuracy (most challenging ingredient). From a block size of 7 × 7 on, the gain is more limited (less than 1 % in mean and 2 % for the minimum). In terms of convergence time we can see that for this block size a lower number of iterations required results in the minimum computing time required. A larger block size than seven increases both network complexity and convergence time, therefore, the 7 × 7 block size seems to be a good compromise and is used in all further experiments.

It should be noted that even the largest image block size considered, 11 × 11 = 121 pixels, is below the smallest seed sizes in our image, which are around 400 pixels for millet seeds. All other ingredients are bigger (e.g. linseeds around 1000 pixels and oat flakes from 3000 pixels up).

In this respect, the performance of the CNN classifier with a 7 × 7 block size is considerably higher than that of the QDC classifier, for all camera systems considered. The Snapscan system, with the highest spatial–spectral resolution, outperforms Mosaic camera systems, independently of the classifier: 70.4 % as mean and 45.9 % minimum for QDC, increasing to 89.1 % mean and 74.3 % minimum for the CNN. Both Mosaic systems, with a reduced spatio–spectral resolution and lower pixel SNR than the Snapscan camera, achieve poor performance with the QDC classifier (~45–50 % in mean and 2–5 % as minimum). However, their performance is considerably increased by the CNN, reaching 78.8 % as mean and 50.7 % as minimum for the VIS and 76.1 % mean and 51.1 % minimum for the NIR.

Impact of pre-processing steps

Pre-processing for denoising by median filtering

We assess here the impact of applying median filters as a pre-processing step to our classifiers for all camera systems with a halogen system. Identical training samples are used for the comparison of the different methods and accuracy is given over the whole image as the test set. Median filtering as a pre-processing step is a simple yet effective method to improve the classification performance. This is particularly the case for the QDC classifier with the Snapscan image. In this case, an increase up to 10 % in mean accuracy and 16 % in the minimum one can be achieved with a 5 × 5 median filter.

For the Mosaic VIS and Mosaic NIR images the median filtered applied is of 3 × 3, to avoid excessive blurring on these lower-resolution images. Median filtering increases mean accuracy by 5–7 % for the QDC classifier in both Mosaic systems. For the CNN there is also a 5 % increase for the Mosaic VIS. For the Mosaic NIR CNN the performance decreases, which could be due to excessive image blurring for this lower-resolution image.

Pre-processing for denoising by spectral/spatial binning

Applying spatial/spectral binning on an image can also help denoise the spectra. However, this is done at the cost of a reduced spatial or spectral resolution. We experiment with spatial and spectral binning of factors 2 and 4 on the Snapscan image (1048 × 1088 pixels). No binning is performed on the Mosaic images since these have lower spatial–spectral resolution. Table 4 shows the impact of binning and median filtering on the classification accuracy with the Snapscan image and halogen illumination. The use of binning increases classification accuracy for the QDC classifier up to 8.5 % in mean and up to 16 % for the minimum class accuracy. Therefore, it is another simple yet effective measure to increase performance. For the CNN classifier, with initial performance around 90 %, the impact of binning is marginal. The benefit of binning being in this case that the memory and computational requirements for the CNN are heavily reduced (by a factor of 16 when binning by 4).

As for the joint impact of binning and median filtering, we can observe that their individual performance increases do not add up when jointly applied. In fact, the highest performance increase is achieved by performing median filtering and not binning at all. The advantage of binning remains, nevertheless, that a lower amount of data needs to be processed. This can be interesting when using the computing-intensive CNNs. In this case, binning does not increase the performance of the CNN, but it heavily reduces computational and memory requirements while maintaining a similar performance.

Table 4. Joint impact of binning and median filtering as pre-processing (Snapscan, Halogen).

Classifier/ resolution	QDC		CNN	
	Mean	Min	Mean	Min
Original	70.4 %	45.9 %	89.1 %	74.3 %
Original + MF	76.0 %	46.6 %	89.5 %	74.1 %
Bin 2	72.2 %	29.2 %	90.5 %	68.0 %
Bin 2 + MF	74.4 %	21.9 %	90.6 %	67.2 %
Bin 4	76.4 %	44.4 %	87.9 %	61.4 %
Bin 4 + MF	77.6 %	33.1 %	87.6 %	58.2 %

Impact of post-processing steps: median filtering and bilateral filtering

We now evaluate the individual and joint impact of median filtering and bilateral filtering as post-processing steps applied on the labelled output image. Bilateral filtering results in a higher performance increase than median filtering (11 % versus 6 % mean increase for Snapscan with QDC) at the cost of higher complexity. The combination of both barely increases the performance. For the CNN there is also a modest performance increase by applying post-processing. The increase is less noticeable since the starting accuracy was high with lower potential for increase.

These post-processing techniques correct misclassified pixels by assuming most neighbouring pixels are correctly classified. Otherwise, spatial filtering cannot correct for this poor classification and may make it worse. This specially occurs with Mosaic systems and QDC where some ingredient accuracies are below 10 %. We can conclude that applying a pre-processing step to denoise the input spectra, is more effective than post-processing an already classified image.

Combined impact of pre-processing and post-processing steps

Table 5 shows the combined impact of the pre-processing and post-processing techniques with our QDC and CNN classifiers. In the case of QDC, median filtering of 5×5 provides best results while in combination with the CNN, median filtering of 3×3 as pre-processing step is slightly preferable.

From Table 5 we can make several observations. First, for the QDC classifier, median filtering is more effective as a pre-processing step than as a post-processing step. Second, a pre-processing step such as median filtering achieves a similar performance increase to bilateral filtering as the post-processing step (around 10 % for QDC). In addition, the joint use of median filtering for pre-processing and bilateral filtering provides further improvement even though

Table 5. Joint impact of pre- and post-processing techniques (Snapscan, Halogen system).

Pre- & post-processing + classifier	QDC		CNN	
	Mean	Min	Mean	Min
Classifier	70.4 %	45.9 %	89.1 %	74.3 %
Classifier + MF	76.0 %	46.6 %	89.5 %	74.1 %
Classifier + BF	81.5 %	55.4 %	90.3 %	74.1 %
MF + Classifier	80.8 %	61.9 %	88.8 %	75.0 %
Classifier + BF + MF	81.6 %	55.5 %	90.3 %	73.8 %
MF + Classifier + MF	82.7 %	62.4 %	89.1 %	74.8 %
MF + Classifier + BF	85.4 %	66.3 %	89.9 %	76.2 %
MF + Classifier + BF +MF	85.5 %	66.6 %	89.9 %	76.2 %

the contribution of each individual performance does not fully add up (10 % mean increase for each technique and 15 % for the combination of both). For CNN, the conclusions are different though, the use of median filtering for pre-processing can even slightly decrease the mean accuracy. The reason might be that the blurring effect of median filtering decreases the spatial information provided to the CNN. Only the use of some post-processing (as BF) can slightly increase the performance (by 1 %).

Combined impact of all system parameters

We now analyse the combination of all system parameters considered in previous sections: illumination system, pre-processing, analysis and post-processing parameters for our two classifier types and different camera systems. Table 6 shows a summary of the basic configurations (only classifier without pre- or post-processing), those with pre- and post-processing methods (excluding binning) and the different illumination systems proposed for the Snapscan camera. Configurations illustrated in the same colour correspond to a similar performance achieved. The last row in Table 6 shows the classification accuracy for the corresponding RGB image.

Figure 5 illustrates Table 6 showing which system parameters can be gradually added to increase the system performance. Starting from a basic configuration (QDC classifier, halogen system and no pre-post processing steps), we can considerably increase the performance, by up to 25 % mean accuracy, until the most complex configuration (CNN classifier, LED system and pre/post processing). Moreover, different parameter configurations can be used to achieve similar performance (showcased in the same colour) while meeting different application requirements.

In this way, from the basic configuration (halogen system and QDC classifier) we can increase the accuracy to a mean of over 80 % and around 60 % for the minimum by either:

■ Combining MF, QDC and BF in the halogen-based system
■ Using the basic classifier QDC without pre/post-processing but in the LED-based system

To further increase performance, to around 90 % for mean and over 70 % the minimum, we can either:

■ Use CNNs, in combination or not with pre/post-processing.
■ Use the LED-based system in combination with QDC classifier, pre- and post-processing (Median Filtering and Bilateral Filtering). This allows us to reach similar accuracy without resorting to CNNs.
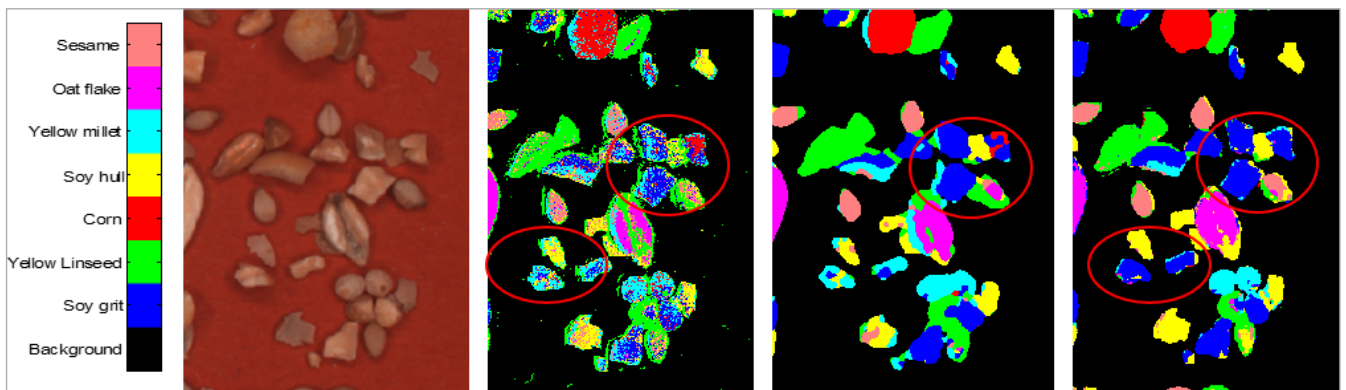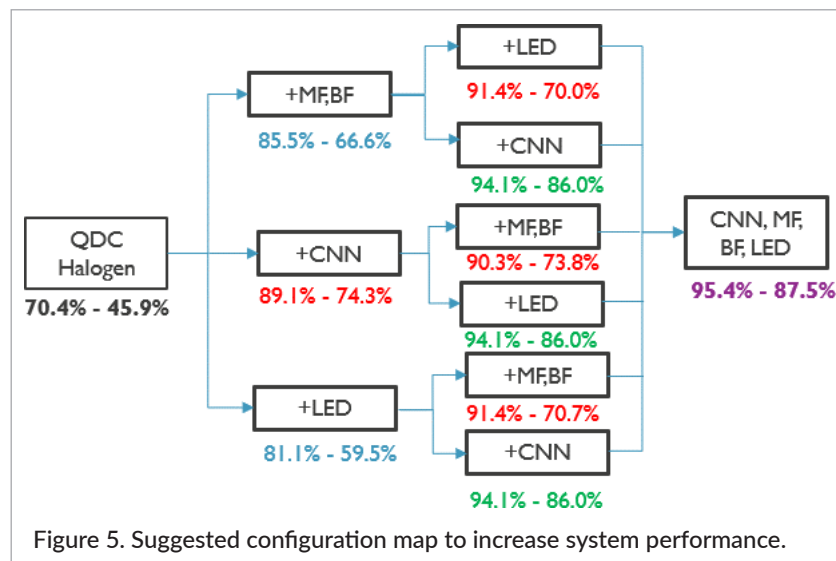
The highest reachable performance is achieved by combining the following system parameters: CNNs, LED system and pre- and post-processing methods.

Finally, we benchmark the results with respect to RGB imaging in Table 6. For this purpose, we take the best possible configuration for RGB imaging of CNNs in combination with post-processing. Even though the use of LED illumination also benefits RGB by around 6 % in mean accuracy, we can see that HSI outperforms RGB imaging even when comparing to the pixel-based approach of QDC, even by a 10 % in the minimum accuracy. When compared to the same analysis method, CNN, HSI outperforms by 20 % on the mean accuracy and around 40 % on the minimum accuracy.

To visualise the effect of these different system aspects on the classification accuracy we show the classified images for the Snapscan binned by 2 case. Figure 6 shows a crop of the original seed image with its corresponding classified images. For some seeds, the QDC has many pixel misclassifications (circled in red). This makes

Table 6. Joint impact of system illumination and analysis methods (Snapscan).

Configuration	Halogen		Halogen HDR		Custom LED	
	Accuracy		Accuracy		Accuracy	
	Mean	Min	Mean	Min	Mean	Min
QDC	70.4 %	45.9 %	73.2 %	50.6 %	81.1 %	59.5 %
MF+QDC+BF+MF	85.5 %	66.6 %	87.0 %	68.5 %	91.4 %	70.7 %
CNN	89.1 %	74.3 %	88.8 %	77.7 %	94.1 %	86.0 %
CNN+BF+MF	90.3 %	73.8 %	90.5 %	80.2 %	95.4 %	87.5 %
CNN+BF+MF (RGB)	67.6 %	34.1 %	—	—	73.5 %	35.6 %

Figure 5. Suggested configuration map to increase system performance.

Figure 6. Original crop and corresponding classified images, left to right with a) Halogen QDC, b) Halogen (MF+QDC+MF) and c) LED system and CNN.

the seeds unrecoverable even after post-processing for seed identification, which relies on the majority of pixel labels in a seed being correct. The use of median filtering prior to and after classification helps to increase the accuracy in some cases but fails in others. Indeed, the central part of the image shows one misclassified sesame seed, misclassified soy hull and soy grit wrongly detected as either millet or corn. This is corrected in the images acquired with the LED system and classified with the CCN, where the right type of seeds is correctly discriminated.

Table 7 shows the performance of selected system configurations for the Mosaic VIS and the Mosaic NIR. In both cases the basic configuration with a pixel-based classifier (LDA+QDC) obtains poor performance. However, by modifying some system parameters that either reduce the spectral noise (MF, LED) or exploit more spatial information (CNNs), the performance can be dramatically

increased (35–45% mean accuracy increase and up to 70% increase in minimum accuracy), then achieving very high performance with both Mosaic VIS and NIR.

It is important to note that LED illumination can help increase performance significantly in either Mosaic VIS or NIR range. This is also the case for the CNN classifier, where over 10% accuracy increase is achieved for the most challenging ingredients. The use of pre- and post-processing techniques can also increase the performance by over 10% in mean accuracy. However, for this application to reach a mean classification accuracy around 80% with the mosaic cameras, CNNs are required, exploiting both spectral and spatial information. Once a CNN is enabled, a more equivalent performance can be obtained from either adding LED illumination or pre-/post-processing. For a mean classification accuracy closer to 90%, we would need to resort to the optimal configuration combining LED illumination, pre-processing through

Table 7. Joint impact of system illumination and analysis methods for Mosaic cameras.

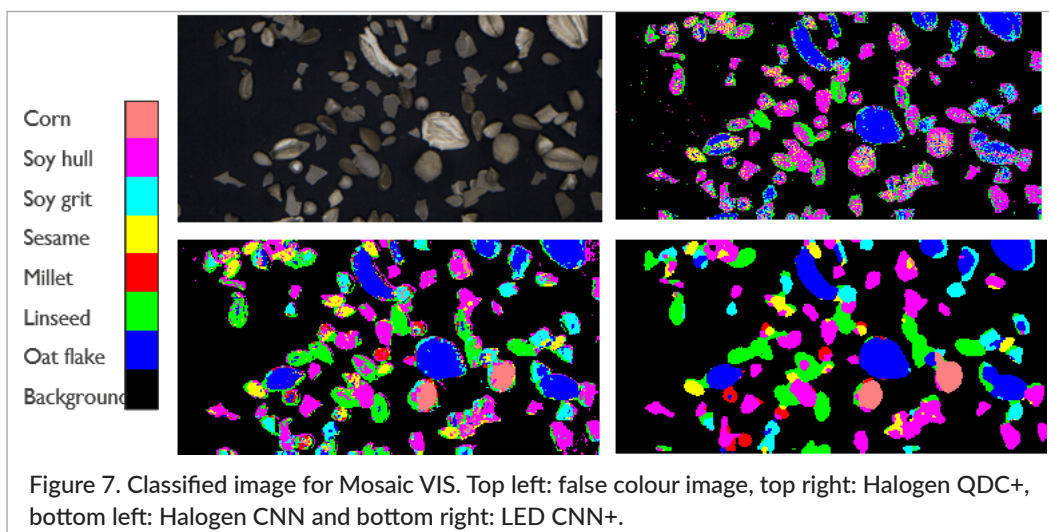| Camera	Mosaic VIS				Mosaic NIR			
	Halogen		Custom LED		Halogen		Custom LED	
Configuration	Mean	Min	Mean	Min	Mean	Min	Mean	Min
QDC	45.5%	2.3%	52.5%	9.3%	50.2%	5.6%	49.4%	6.1%
MF+QDC+BF+MF	56.8%	0.7%	65.0%	7.9%	60.1%	32.1%	62.6%	43.2%
CNN	78.8%	50.7%	80.5%	59.3%	76.2%	51.1%	83.1%	63.8%
(MF)+CNN+BF+MF	87.6%	75.6%	90.3%	75.1%	77.8%	52.5%	84.9%	66.0%

median filtering, a CNN classifier and post-processing. The most performing configuration under all illumination systems and cameras is achieved by using CNN, median filtering as pre- and post-processing and bilateral filtering. Only for Mosaic NIR images with CNN, due to the reduced spatial resolution, the optimal configuration does not use median filtering as pre-processing step.

Figure 7 shows an example of a classified image for the Mosaic VIS under different system settings. We can see how QDC poorly discriminates most seeds, while using CNN (bottom left) enables correct discrimination of the different ingredients with a mean accuracy close to 80%. The use of LED lights and extra pre- and post-processing on CNN (named as CNN+) further improves the classified image reaching up to 90% pixel classification accuracy.

Conclusions

In this paper, we proposed a methodology to find the optimal combination of several system parameters that allows us to increase our hyperspectral system performance, measured as pixel classification accuracy. We compared the impact of different parameters of a HSI system, namely illumination, camera and analysis methods. From our experiments we observe that illumination is a key aspect of a hyperspectral system, since a customised LED system can increase accuracy over a traditional halogen system in the order of 10% in mean accuracy and close to 15% for the minimum accuracy. In addition, the use of pre-processing and post-processing methods (such as median and bilateral filtering) can also increase the accuracy significantly: over 10% in mean, and around 20% for the minimum accuracy for a pixel-based classifier, but also around 1–3% the minimum accuracy in the CNN case. The impact of using an image-based versus a pixel-based classifier is even more noticeable for our application, where an increase in 15–20% mean accuracy can be obtained with respect to our pixel-based classifier. In total, the mean pixel classification accuracy for the Snapscan camera increased from 70% to 95% by replacing the basic combination (halogen

Figure 7. Classified image for Mosaic VIS. Top left: false colour image, top right: Halogen QDC+, bottom left: Halogen CNN and bottom right: LED CNN+.

illumination with quadratic discriminant classification and no pre-/post-processing) by the most advanced one: custom LED illumination with a CNNs and pre/post-processing methods applied. For more portable low-resolution Mosaic cameras, enabling all system parameters in this advanced configuration can noticeably increase our application mean pixel accuracy from 50% to over 90%.

Our system analysis has shown that similar performance can be achieved by enabling a different set of system parameters. For instance, we can reach the same performance as a CNN under halogen lights by using a QDC classifier and enabling custom LED illumination and the use of pre- and post-processing techniques. This specific system trade-off is valid for the application considered. A basic system with QDC and halogen lights might suffice to reach the highest accuracy in another application. Similarly, a custom LED illumination providing more balanced spectral and spatial distribution is beneficial, but its advantage could be more limited for a different sensor or camera system. Another important aspect is the degree of spatial and spectral information available in our application. The more information available in the spatial dimension, the more beneficial a joint spatial–spectral analysis with CNN will be with respect to a purely spectral pixel-based analysis such as QDC. This is application dependent. The existing performance–complexity trade-offs between high-resolution and low-resolution camera systems often hold for a generic hyperspectral system. This said, for some applications a low-resolution system may suffice to reach the application target performance. A general conclusion that can be drawn is the need for a joint study of all system parameters. By performing this joint analysis we can enable the system parameters that better fit the application requirements in terms of performance, spatial resolution, computation cost or speed of acquisition. Therefore, system-wide analysis can lead to better decisions when adapting to the specific application requirements.

Acknowledgement

References

1. J.M. Amigo, I. Martin and A. Gowen, "Hyperspectral imaging and chemometrics: a perfect combination for the analysis of food structure, composition and quality", in *Chemometrics in Food Chemistry*, Ed by F. Marini. Elsevier (2013). https://doi.org/10.1016/B978-0-444-59528-7.00009-0

2. T. Brosnan and D.-W. Sun, "Improving quality inspection of food products by computer vision – a review", *J. Food Eng.* **61,** 3–16 (2004). https://doi.org/10.1016/S0260-8774(03)00183-3

3. A. Signoroni, M. Savardi, A. Baronio and S. Benin, "Deep learning meets hyperspectral image analysis: a multidisciplinary review", *J. Imaging* **5,** 52 (2019). https://doi.org/10.3390/jimaging5050052

4. W. Lv and X. Wang, "Overview of hyperspectral image classification", *J. Sensors* **2020,** 4817234 (2020). https://doi.org/10.1155/2020/4817234

5. J.P. Kerekes and J.E. Baum, "Hyperspectral imaging system modelling", *Lincoln Lab. J.* **14(1),** 117–130 (2003).

6. A. Gutierrez-Gutierrez, A. Pardo, E. Real, J.M. Lopez-Higuera and O.M. Conde, "Custom scanning hyperspectral imaging system for biomedical applications: modeling, benchmarking, and specifications", *Sensors* **19(7),** 1692 (2019). https://doi.org/10.3390/s19071692

7. A.M. Peter, *Design and Evaluation of LED Illumination Source for Multispectral Imaging Applications*. Thesis for Master of Technology (2015).

8. T.W. Sawyer, A.S. Luthman and S.E. Bohndiek, "Evaluation of illumination system uniformity for wide-field biomedical hyperspectral imaging", *J. Optics* **19(4),** 045301 (2017). https://doi.org/10.1088/2040-8986/aa6176

9. K.C. Lawrence, B. Park, G. Heitschmidt, W. Windham and Ch. Thai, "Evaluation of LED and tungsten-halogen lighting for fecal contaminant detection", *Appl. Eng. Agric.* **23(6),** 811–818 (2007). https://doi.org/10.13031/2013.24048

10. J. Katrašnik, F. Pernus and B. Likar, "A method for characterizing illumination systems for hyperspectral

imaging", *Opt. Express* **21,** 4841–4853 (2013). https://doi.org/10.1364/OE.21.004841

11. J.M. Carstensen, "LED spectral imaging with food and agricultural applications", *Proc. SPIE* **10656,** 1065604 (2018). https://doi.org/10.1117/12.2304698

12. M. Shahrimie, P. Mishra, P. Scheunders, S. Mertens, N. Wuyts and S. Dhondt, "Modeling effects of illumination and plant geometry on leaf reflectance spectra in close-range hyperspectral imaging", *8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS),* Los Angeles, CA, pp. 1–4 (2016). https://doi.org/10.1109/WHISPERS.2016.8071753

13. P. Mishraa, S. Lohumic, H.-A. Khand and A. Nordon, "Close-range hyperspectral imaging of whole plants for digital phenotyping: Recent applications and illumination correction approaches", *Comput. Electron. Agric.* **178,** 105780 (2020). https://doi.org/10.1016/j.compag.2020.105780

14. C. Blanch, W. Saeys and A. Lambrechts, "Optimized custom LED illumination system for hyperspectral imaging application", *19th International Conference on Near Infrared Spectroscopy,* Gold Coast, Australia (2019).

15. T. Naes, T. Isaksson, T. Fearn and T. Davies, *A User-Friendly Guide to Multivariate Calibration and Classification.* NIR Publications (2004).

16. L. Pan, Y. Sun, H. Xiao, X. Gu, P. Hu, Y. Wei and K. Tu, "Hyperspectral imaging with different illumination patterns for the hollowness classification of white radish", *Postharvest Biol. Technol.* **126,** 40–49 (2017). https://doi.org/10.1016/j.postharvbio.2016.12.006

17. M.E. Paoletti, J.M. Haut, J. Plaza and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification", *ISPRS J. Photogramm. Remote Sens.* **145A,** 120–147 (2018). https://doi.org/10.1016/j.isprsjprs.2017.11.021

18. Y. Chen, H. Jiang, Ch. Li, X. Jia and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks", *IEEE Trans. Geosci. Remote Sens.* **54(10),** 6232–6251 (2016). https://doi.org/10.1109/TGRS.2016.2584107

19. F. Deng, S. Pu, X. Chen, Y. Shi, T. Yuan and S. Pu, "Hyperspectral image classification with capsule network using limited training samples", *Sensors* **18(9),** 3153 (2018). https://doi.org/10.3390/s18093153

20. M. Al-Sarayreh, M. Reis, W.Q. Yan and R. Klette, "Detection of red-meat adulteration by deep spectral-spatial features in hyperspectral images", *J. Imaging* **4(5),** 63 (2018). https://doi.org/10.3390/jimaging4050063

21. Z. Wang, M.-H. Hu and G. Zhai, "Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data", *Sensors* **18(4),** 1126 (2018). https://doi.org/10.3390/s18041126

22. A. Farooq, X. Jia, J. Hu and J. Zhou, "Knowledge transfer via convolutional neural network for multiresolution lawn weed classification", *10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS),* Amsterdam, Netherlands (2019). https://doi.org/10.1109/WHISPERS.2019.8920832

23. F. Gruber, W. Grählert, P. Wollmann and S. Kaskel, "Classification of black plastics waste using fluorescence imaging and machine learning", *Recycling* **4,** 40 (2019). https://doi.org/10.3390/recycling4040040

24. Z. Gao, Y. Shao, G. Xuan, Y. Wang, Y. Liu and X. Han, "Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning", *Artif. Intel. Agric.* **4,** 31–38 (2020). https://doi.org/10.1016/j.aiia.2020.04.003

25. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Trans. Pattern Anal.* **12(7),** 629–639 (1990). https://doi.org/10.1109/34.56205

26. J. Weickert, *Anisotropic Diffusion in Image Processing.* Teubner-Verlag (1998).

27. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images", *Sixth International Conference on Computer Vision* (IEEE Cat. No.98CH36271), Bombay, India, pp. 839–846 (1998). https://doi.org/10.1109/ICCV.1998.710815

28. D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis", *IEEE Trans. Pattern Anal.* **24(5),** 603–619 (2002). https://doi.org/10.1109/34.1000236

29. W. Liao, D.E.O. Donoso, F. Van Coillie, J. Li, C. Qi, S. Gautama and W. Philips, "Spectral-spatial classification for hyperspectral image by bilateral filtering

and morphological features", *8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Los Angeles, CA (2016). https://doi.org/10.1109/WHISPERS.2016.8071680

30. X. Kang, Z. Li and J.-A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering", *IEEE Trans. Geosci. Remote Sens.* **52(5),** 2666–2677 (2014). https://doi.org/10.1109/TGRS.2013.2264508

31. R.C. Gonzalez and R.E. Woods, *Digital Image Processing*. Prentice-Hall (2002).

32. V. Thirilogasundari, V. Suresh babu and S.A. Janet, "Fuzzy based salt and pepper noise removal using adaptive switching median filter", *Procedia Eng.* **38,** 2858–2865 (2012). https://doi.org/10.1016/j.proeng.2012.06.334

33. C. Blanch, W. Saeys and A. Lambrechts, "Convolutional neural networks for heterogeneous ingredient discrimination with hyperspectral imaging", *10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Amsterdam, Netherlands (2019). https://doi.org/10.1109/WHISPERS.2019.8921395

34. P. Gonzalez, K. Tack, B. Geelen, B. Masschelein, W. Charle, B. Vereecke and A. Lambrechts, "A novel CMOS-compatible, monolithically integrated line-scan hyperspectral imager covering the VIS-NIR range", *Proc. SPIE* **9855,** 98550N (2016). https://doi.org/10.1117/12.2230726

35. The GIMP Team, *GNU Image Manipulation Program, GIMP 2.0, 1997-2020* [retrieved 31/07/2014]. http://gimp.org

36. CVRL Ltd. http://cvrlighting.com

37. J. Pichette, W. Charle and A. Lambrechts, "Fast and compact internal scanning CMOS-based hyperspectral camera: the Snapscan", *Proc. SPIE* **10110,** 1011014 (2017). https://doi.org/10.1117/12.2253614

38. B. Geelen, C. Blanch, P. Gonzalez, N. Tack and A. Lambrechts, "A tiny VIS-NIR snapshot multispectral camera", *Proc. SPIE* **9374,** 937414 (2015). https://doi.org/10.1117/12.2077583

39. C. Blanch, W. Saeys and A. Lambrechts, "Fast ingredient quantification in multigrain flour mixes using hyperspectral imaging", *J. Food Control* **118,** 107366 (2020). https://doi.org/10.1016/j.foodcont.2020.107366

40. PerClass BV (2008–2019). http://perclass.com/perclass-toolbox/product

41. *Matlab Release 2015b*. The MathWorks, Inc., Natick, MA, USA (2015). https://www.mathworks.com

42. L. Smith, *A Tutorial on Principal Component Analysis* (2002). http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

43. X. Yang, Y. Ye, X. Li, R.Y.K. Lau, X. Zhang and X. Huang, "Hyperspectral image classification with deep-learning models", *IEEE Trans. Geosci. Remote Sens.* **56(9),** 5408–5423 (2018). https://doi.org/10.1109/TGRS.2018.2815613

44. C.D. Cantrell, *Modern Mathematical Methods for Physicists and Engineers*. Cambridge University Press (2000). ISBN: 0-521-59827-3

45. T.A. Goodman, "International standards for colour", *Colour Design*, *Theories and Applications*, 2nd Edn. Woodhead Publishing, pp. 417–452 (2012). https://doi.org/10.1016/B978-0-08-101270-3.00018-7

46. International Color Consortium, *sRGB Doc, IEC 61966-2-1.* (1999). http://www.color.org/chardata/rgb/srgb.xalter