

# **TSB-funded project ‘TADD’ – Trainable vision-based anomaly detection and diagnosis**

Ran Song, Hossein Malekmohamadi and Tom Duckett  
Agri-Food Technology Research Group, Lincoln Centre for  
Autonomous Systems Research, School of Computer Science,  
University of Lincoln UK

*Technical Quarterly Report*  
Nov. 2014 to Jan. 2015

## Abstract

This technical quarterly report is a summary of the work we have done at the University of Lincoln from November 2014 to January 2015 inclusive for the project Trainable Vision-based Anomaly Detection and Diagnosis (TADD).

## Summary of Work Completed

During this period, the two post-doctoral Research Fellows, Dr Ran Song and Dr Hossein Malekmohamadi completed their employments at the University of Lincoln. Their respective contributions to the project during this Quarter were based on the following dates:

- Ran Song – completed employment on Tuesday 25th November.
- Hossein Malekmohamadi – completed employment on Friday 5<sup>th</sup> December.

The major part of the work done by the two researchers in this period (from 1<sup>st</sup> October 2014 up to the respective leaving dates) was allocated to complete the planned developments at the University of Lincoln on the QA-TADD software and hardware set-up. The motivation here was to make sure that all planned work on QA-TADD at the University could be completed and handed over to the partners at Sutton Bridge Crop Storage Research Ltd., so that the two new replacement researchers to be recruited could be allocated to work on the Seal-TADD and 3D-TADD work packages, as shown in the Gantt chart in Fig. 1 (from the agreed Second Level Project Plan).

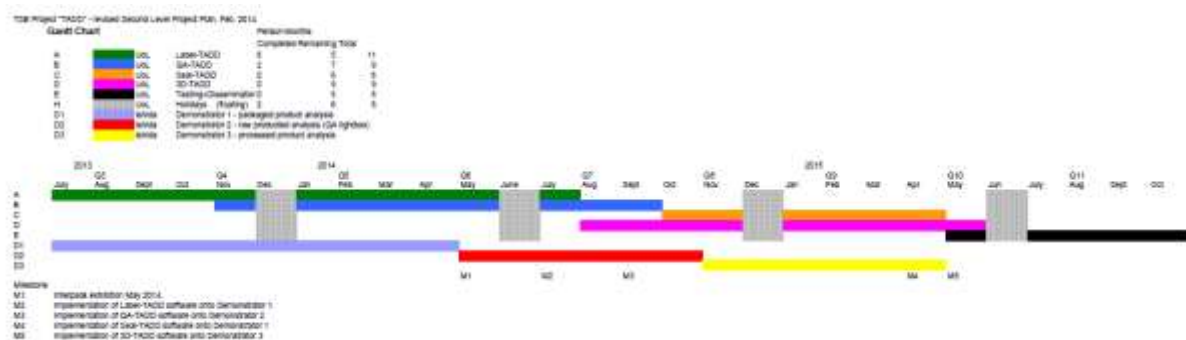


Figure 1: Second Level Project Plan - Gantt Chart for TADD Project

In addition, Chris Waltham, a Senior Technician from the School of Computer Science at the University of Lincoln, was employed on the project to build the required hardware set-up, including design, construction, lighting and electronics, for the QA-TADD Demonstrator 2 platform (see Fig. 2).



Figure 2: QA-TADD "Demonstrator 2" hardware set-up

The major achievements in this period included (i) completing the required evaluation of the QA-TADD software, (ii) writing up the developments on QA-TADD for submission to a journal in food engineering (please see Appendix A), (iii) deployment of the QA-TADD Demonstrator 2, comprising the completed QA-TADD hardware and software at Sutton Bridge Crop Storage Research (SBCSR). Hossein Malekmohamadi visited SBCSR from 25-27 November 2015 to carry out the deployment, training of SBCSR staff to use the system, and final testing of the system performance.

## Summary of Financial Claim for Quarter 8

The University of Lincoln's claim for this period comprises the following elements:

- Directly incurred staff costs (£5940.22) plus associated Estates and Indirect costs (see below) for Dr Ran Song and Dr Hossein Malekmohamadi.
- Travel and subsistence expenses (£969.94): this includes some items from the previous Quarter, due to the time taken for expenses to pass through the University's e-expenses systems. A major part of this was for Ran Song's attendance at the Siggraph 2014 Conference, as was reported in the report for Quarter 7. Additionally, this included some travel and subsistence for Hossein Malekmohamadi's stay at Sutton Bridge from 25-27

November 2014 to complete the QA-TADD deployment, and the Review Meeting conducted by the project partners at the University of Lincoln on 16 December 2014.

- Directly allocated Investigators (£2410.19) – similar costs as usual, to cover the respective contributions of Prof Tom Duckett and Mike Dudbridge to the project.
- Technicians (Directly Allocated Other cost) (£1642.40) – incurred to build hardware set up (see above).
- Estates costs (£1897.47) + Indirect costs (£6327.11) - reduction from previous quarters to reflect the direct link with the combined effort of Researchers and Investigators.

## Appendix A: Vision-Based Potato Blemish Detection via Interactive Image Segmentation

# Vision-Based Potato Blemish Detection via Interactive Image Segmentation

Ran Song\* · Hossein Malekmohamadi\* · Jamie Hutton · Glyn Happer · Tom Duckett

Received: date / Accepted: date

**Abstract** We present a vision-based system for detecting potato blemishes. It is a feasible solution to many different quality control tasks because it adopts a novel scheme of interactive image segmentation. In general, our system incorporates the interactive acquisition of image-based training data into the automatic classification of natural food products using adaptive boosting. To reduce the manual effort required for interactively collecting the online training data, the input images are efficiently processed through GPU-based over-segmentation. We also developed a user-friendly graphical user interface which allows users to edit (select, de-select and reselect) the training data. It also allows the users to add more training data to improve the classifier if necessary. In contrast to the state-of-the-art methods which typically rely on offline training using a large col-

lection of data, our system relies only on a tiny amount of training data offered by the users in an interactive manner. The experiments demonstrate that our system efficiently and effectively detects potato blemishes and can be potentially used for many other food processing applications.

**Keywords** Potato · Interactive Image Segmentation · Classification · Adaboost · Over-segmentation

## 1 Introduction

It has been more than two decades since computer vision techniques were applied to automatic inspection of food products for quality assurance [6]. The essential motivation is to use such techniques to replace or augment human inspectors for the evaluation of a variety of quality attributes of raw and prepared foods [7]. Benefiting from the rapid development of commercial computing and sensing technologies, computer vision systems designed for food quality assurance have been able to provide a high level of flexibility and repeatability at relatively low cost. In sharp contrast, labour cost has been widely increased in the food industry where manufactures typically have a low profit margin. Also, consumers tend to buy the food products which look good since in most cases, it is not possible to directly taste the food (e.g., potato) when buying it. Hence, many companies in the food industry are willing to invest in machine vision techniques to reduce the cost of manual inspection as well as to improve quality and boost sales.

Potato is one of the most important foods in human history and currently accounts for 70 – 80% of the carbohydrate consumed in the UK. The development of a vision-based system for detecting potato blemishes is

---

\* denotes equal contribution to this paper.

Ran Song  
School of Computer Science, University of Lincoln, Lincoln, UK  
E-mail: sora1998@hotmail.com

Hossein Malekmohamadi  
School of Computer Science, University of Lincoln, Lincoln, UK  
E-mail: hmalek@lincoln.ac.uk

Jamie Hutton  
School of Computer Science, University of Lincoln, Lincoln, UK

Glyn Happer  
AHDB Potato Council, Sutton Bridge Crop Storage Research, Spalding, UK  
E-mail: Glyn.Harper@potato.ahdb.org.uk

Tom Duckett  
School of Computer Science, University of Lincoln, Lincoln, UK  
E-mail: tduckett@lincoln.ac.uk

not just economically significant, but also has a great potential to benefit quality assurance for other types of food due to the high visual variation of potatoes. This paper presents a prototype system for detecting potato blemishes using state-of-the-art computer vision techniques. It is composed of a computer and a light box as shown in Fig. 1 where a digital SLR camera is fixed inside. The novelty of the proposed system is summarised as below:

1. The system relies merely on a tiny amount of training data in comparison with existing methods.
2. It enables a quantitative evaluation and analysis for the quality control of potatoes which is unavailable under the current working mode in industry.
3. The system is much more efficient than existing methods mainly due to the GPU-based implementation and the novel image processing pipeline based on classification of superpixels rather than pixels.
4. To the best of our knowledge, for the first time, a manual ‘correction mechanism’ is incorporated into a vision-based system designed for the quality control of natural food products. This is achieved by the novel interactive training which is essentially a “human-in-the-loop” scheme.

### 1.1 Related work

Barni *et al.* [3] presented an intelligent vision-based system for detecting defects on chicken meat. Defects were detected based on the analysis of the chromatic content of chicken images. Possible defect areas were first extracted by means of morphological image reconstruction and then classified according to a predefined list of defects. Russo *et al.* [18] a system for monitoring the production of fast food using a single static camera. By analysing the video with some computer vision techniques, a description of the fast food is produced for the ingredients and their physical arrangement to make sure that the food was produced correctly. Lee *et al.* [11] used colour information for monitoring food processing applications involving live products and biomaterials. Their system detected features based on colours and then employed principal component analysis to characterise target features from a set of training data. More recently, Joutou *et al.* [10] also investigated colour features where the food image recognition system also used other types of features such as texture and Scale invariant Feature Transform (SIFT) [12]. Then, the multiple kernel learning method was employed to estimate the optimal weights to combine these features for each category of food. Yang *et al.* [23] presented another food



**Fig. 1** The light box of the proposed prototype system for vision-based potato blemish detection. It can be seen that a digital SLR camera is fixed at the top of the light box and connected with the computer next to it. To reduce reflection, we laid the array of LED lights in a diffusion box where its front panel has been taken out.

recognition system which utilised a novel representation for food items that calculates pairwise statistics between local features computed over a soft pixel-wise segmentation of the image into eight ingredient types. The classification was based on the multi-dimensional histogram of these statistics. Boiman and Irani [4] proposed a method to detect irregularities for fruit inspection by comparing a test image to a reference image representing the ‘good’ case (expected or normal patterns). It generalised from the reference images and analysed the test images to find regions with low composition likelihood, which were considered as defects. To detect multiple food items from one food image, Matsuda and Yanai [14] proposed a method based on co-occurrence statistics and manifold ranking.

In particular, computer vision techniques have been used for inspecting potatoes for a long time. Tao *et al.* [21] used Fourier analysis to develop a classification method which achieved a high accuracy. Zhou *et al.* [25] used green levels to detect greening and sprouting defects in potatoes. They also classified potatoes in terms of shape by comparison to an ellipse template, and in size and weight by measuring the minor axis and area, respectively. Recently, a method [2] which used image

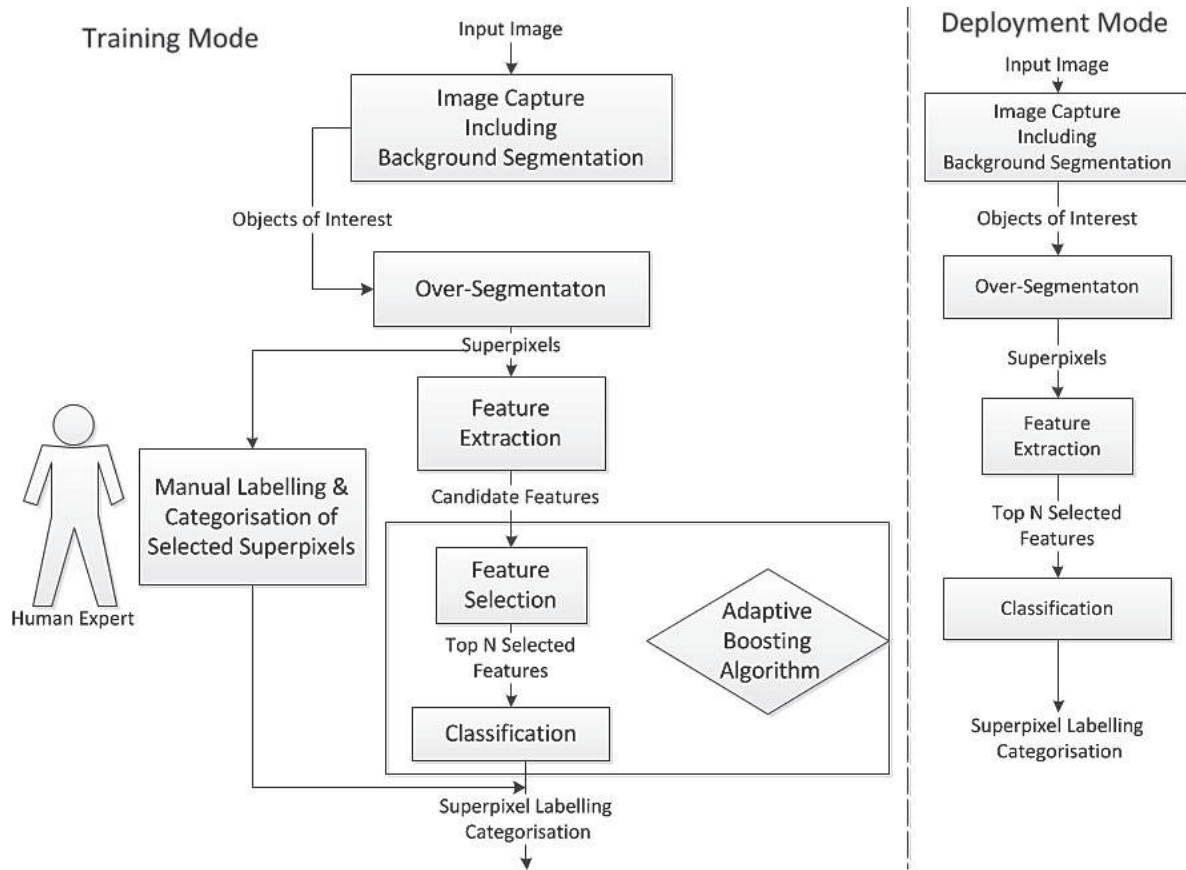


Fig. 2 Overview of our system

segmentation algorithms to detect potatoes blemishes has been proposed. This method is based on training a pixel-wise classifier using features extracted from the image. However, the method is very slow although its classification achieves a high success rate.

Many other related works where computer vision technology is used for food quality inspection can be found in the review papers [8, 5, 24, 22, 9].

## 1.2 Our system

It can be seen that there are several common issues existing in the vision-based techniques for food quality inspection.

Firstly, the computational efficiency of the technique is vital since food products are typically fast-moving consumer goods. It is common that a huge amount of food products (e.g., potatoes) need to be handled every day in food processing and packaging factories.

Secondly, due to the variety and the variation of food, most of the systems require a large dataset for training purpose in order to cover each type of defects, which could be extremely occasional, but do exist. However, it is not always easy to collect sufficient

data for training because a food product could change rapidly in terms of colour, intensity, shape and volume, and these features are important in most of the aforementioned techniques. Typically, such a data collection could take a couple of months (e.g., the Supermarket Produce dataset was the result of 5 months of on-site collecting in a local fruit and vegetable distribution centre [17]).

Thirdly, since usually a food company does not only process one type of food product (especially if the food is seasonal), the systems need to be periodically re-trained to adapt to different food products. This means that the time-consuming offline training, as well as the extra effort needed to prepare the new training data (e.g., manually labelling the images of another product) could be very inconvenient for practical applications.

Finally and most importantly, to the best of our knowledge, there is no manual ‘correction mechanism’ involved in any of the existing vision-based systems for food quality assurance. This is a waste of the expertise of the very people using the systems. They are typically quality control inspectors who possess a significant amount of knowledge about the possible defects. Moreover, safety standards on food products are quite



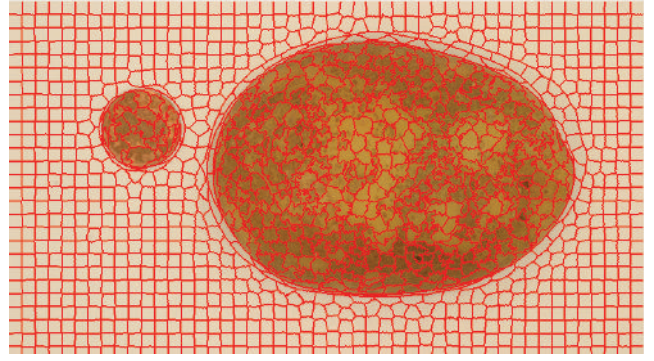
strict since a bad product could cause serious troubles and the producer might face huge fines and compensations. Therefore, from a practical point of view, it is very important to introduce some manual correction mechanism as there is no guarantee that the automatic system can always deliver successful quality inspection. When the automatic system cannot recognise a certain defect, the human inspector should be allowed to step in and offer the particular help to the system to correct the error.

These four issues form the motivation of developing our novel system designed specifically for detecting potato blemishes but with potential application to many other food products besides (e.g., fruits, beans, etc.). Fig. 1 shows the hardware components of the prototype system. With regard to the software, its most novel component is the idea of using real-time interactive image segmentation. To realise this idea, we develop a novel mechanism to help users to conveniently select the training data that they think necessary in an interactive manner. An adaptive boosting (AdaBoost) classifier [19] which enables feature selection is then trained using the manually selected data. And the test image can be assessed by the trained classifier in real-time since only the selected features need to be calculated during the classification. In the next section, we give the technical details of the user-friendly mechanism which enables the interactive training, as well as the classification based on a set of image features selected using AdaBoost.

## 2 Interactive image classification via Adaboost classification

Normally, to train the classifiers and test their performance, the images are marked up by hand to provide the ground truth information indicating the correct class of each pixel. Most existing publicly available databases which benchmark image segmentation techniques work in this way. However, due to the variety of visual appearance of potatoes, it is not realistic to apply such a strategy to the real tasks of potato quality inspection. Note that one specificity of potato is that its appearance changes rapidly and significantly over time and in many cases, time is the very reason that some defects occur. This is quite different from the objects in the image databases (e.g., the MSRC image database<sup>1</sup>) for benchmarking general image segmentation or object classification methods where the objects are generally invariant within a certain period of time. Such a speci-

<sup>1</sup> <http://research.microsoft.com/en-us/projects/objectclassrecognition/>



**Fig. 3** Visualisation of the over-segmentation for an image containing a coin and a potato using SLIC.

ficity makes the conventional offline training based on a large set of previously constructed ground truth data unreliable. This is why it is so important to implement the online training using the currently available data. To achieve this, we developed the system illustrated in Fig. 2.

In Fig. 2, the training mode is designed for iteratively teaching the adaptive boosting algorithm to distinguish between different categories such as healthy potato versus blemished skin, based on the visual appearance. This stage requires the user to select and label a set of blemishes, which is used later for training the system. The user is then able to receive feedback by visually inspecting the accuracy of the classification. If mistakes are being made the user simply retrain the system. Once trained the system can be used in deployment mode, which means no more training is required to classify and diagnose blemishes. The user is able to swap back to training mode at any time to retrain after any classification mistakes. This procedure means there is no limitation on how many times the system can be trained, relying on experts to teach and guide the machine learning algorithm.

### 2.1 Background segmentation

The proposed system use a simple thresholding method for background segmentation. After the background segmentation, the user is not allowed to select any region in the background as the training data.

### 2.2 Over-segmentation

Image over-segmentation is used in our system to enhance the interactive mark-up. The process involves over-segmenting the captured image into local clusters (i.e., superpixels) which follow the contour of edges by colour or intensity differences. This approach aims to

reduce redundancy in the image. Superpixels also allow an expert to select a relatively large area with only a single mouse click, which significantly mitigates the manual effort needed in the interactive scheme and makes it much more efficient. We employed the SLIC algorithm [1] for over-segmentation where to further improve the efficiency, we apply the parallelised GPU (graphics processing unit) version of SLIC [16] in our system. It works by loading the cluster's information into local shared memory. In each iteration, after all pixels have been assigned a label, one thread per cluster is used to update the centres. As an example, the result of image over-segmentation is shown in Fig. 3.

### 2.3 Feature extraction

In our system, a large collection of candidate features are required for each superpixel, from which the AdaBoost algorithm can select to build the classifiers. As suggested by [2], we use 7 colour channels: raw RGB, normalised RGB and the intensity channel. In [2], the image regions used for feature extraction were squares of size  $33 \times 33$ ,  $65 \times 65$ ,  $97 \times 97$ ,  $129 \times 129$  and  $161 \times 161$  pixels. Our system differs by using only superpixel regions, classifying each cluster of pixels based on their region characteristics. Superpixels reduce computation requirements in two aspects. Firstly, an image sized  $1280 \times 720$  is represented as around 9200 clusters requiring only 1% of the computational resources when compared to pixel-wise classification. Secondly, the superpixel regions adapt to cluster together similar pixels, compensating for the requirement of using multiple region sizes.

We then consider the first three statistical moments for each superpixel: average (Eq. (1)), variance (Eq. (2)) and skewness (Eq. (3)).

$$\bar{x} = \frac{1}{N} \sum_{j=1}^N x_j; \quad (1)$$

$$Var(x_1 \dots x_N) = \frac{1}{N-1} \sum_{j=1}^N (x_j - \bar{x}); \quad (2)$$

$$Skew(x_1 \dots x_N) = \frac{1}{N} \sum_{j=1}^N \left( \frac{x_j - \bar{x}}{\sqrt{Var(x_1 \dots x_N)}} \right)^3. \quad (3)$$

The first statistical moment is difficult to parallelize as this requires the implementation of a prefix-sum for the most optimal implementation, adding complexity with the use of shared memory. We use the large amount of segments produced by the gSLIC algorithm and sum each region on a separate thread using cluster centres. Once the first statistical movement has been

calculated both variance and skewness are simple to implement.

We calculate the three moments in each channel, resulting in 21 features. We also use a Sobel edge detector and a range filter in the intensity channel respectively. The aim of the edge detector is to determine the rate change in pixel intensities, which is a useful candidate feature for defects with a high rate of change. The range filter calculates the minimum, maximum and difference between the maximum and minimum values within a  $5 \times 5$  neighbourhood. It indicates the roughness of a texture, aiding in the identification of defects. Then we compute the 3 moments on each of the resultant images, which produces 6 more features. We also record the longest Sobel edge length in each superpixel as a feature, which has been demonstrated experimentally useful for improving classification of some defects. Other features such as Gabor feature [13] and Local Binary Features (LBP) [15] were investigated experimentally as well. However, we found that they did not significantly improve the classification in our case while slowing down the processing. Finally, we incorporate 28 features in total into the classification.

### 2.4 Classification via AdaBoost

We employ the Adaptive Boosting (AdaBoost) algorithm proposed in [2] to automatically select the best features for discriminating between defects and normal regions. The reason that we apply this approach in our system is that it can handle the natural variation in fresh produce due to different potato varieties, seasons, etc. [2]. Also, the method is able to build 'minimalist' classifiers that optimise detection performance at low computational cost, which is very important for a real-time interactive system.

This approach involves two stages: the first stage selects a feature set that will be used to train an AdaBoost classifier in the second stage. It also includes an additional step in order to limit the number of unique features used in the final classifier to a smaller number than the total number of weak classifiers allowed.

We further developed a graphical user interface (GUI) as shown in Fig. 4. With such a GUI and the aforementioned techniques combined, our system allows users to select a handful of superpixels as the training data and the system can show the classification results based on the AdaBoost classifier trained by them in less than 1 second including visualisation. The users can then interactively improve the results by selecting new superpixels (or deselecting existing superpixels). We also

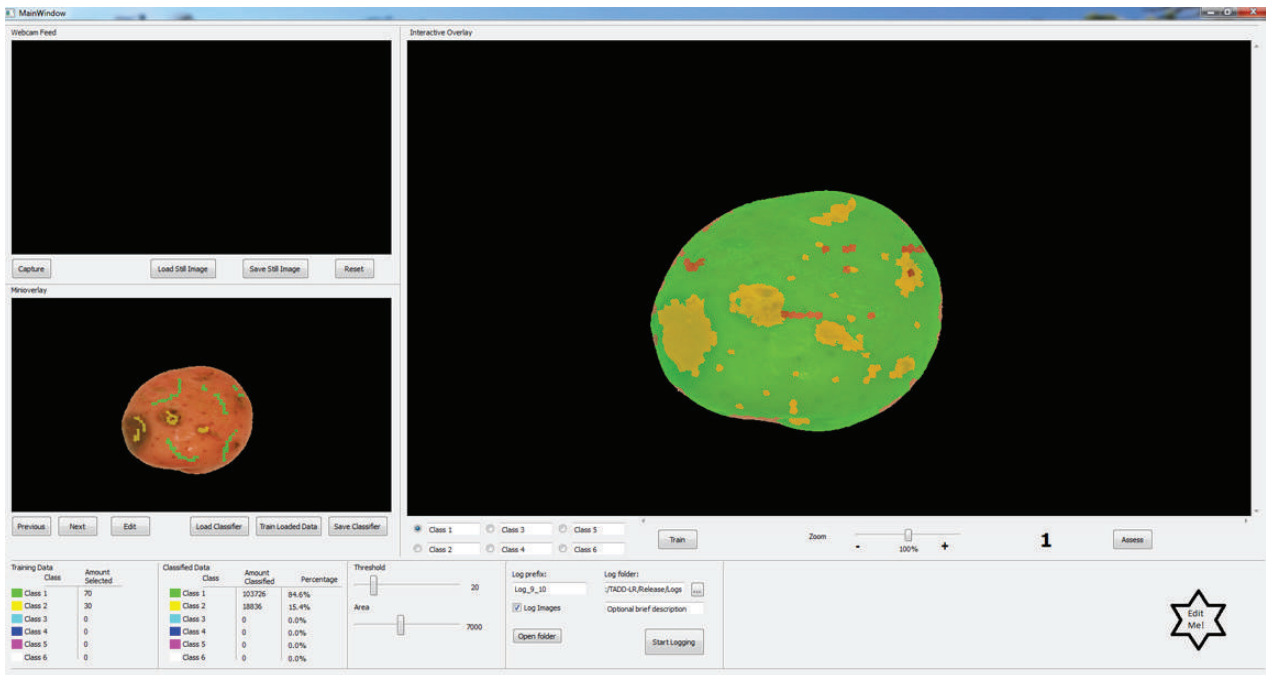


Fig. 4 The user-friendly GUI of the software

upload a demonstration video to YouTube<sup>2</sup> (1080p HD is available) to show how our system works.

### 3 Experimental results and evaluation

We test the proposed system in two ways. The first one is to evaluate its accuracy for two-label image segmentation. Two-label image segmentation can be used for grading tasks where we partition a potato into blemished and unblemished regions. Then the grading is simply based on the percentage of the size of blemished or unblemished regions. The second one is more complicated. It is designed to evaluate its performance in the tasks of detecting several common potato blemishes through multi-label image segmentation.

#### 3.1 Two-label image segmentation

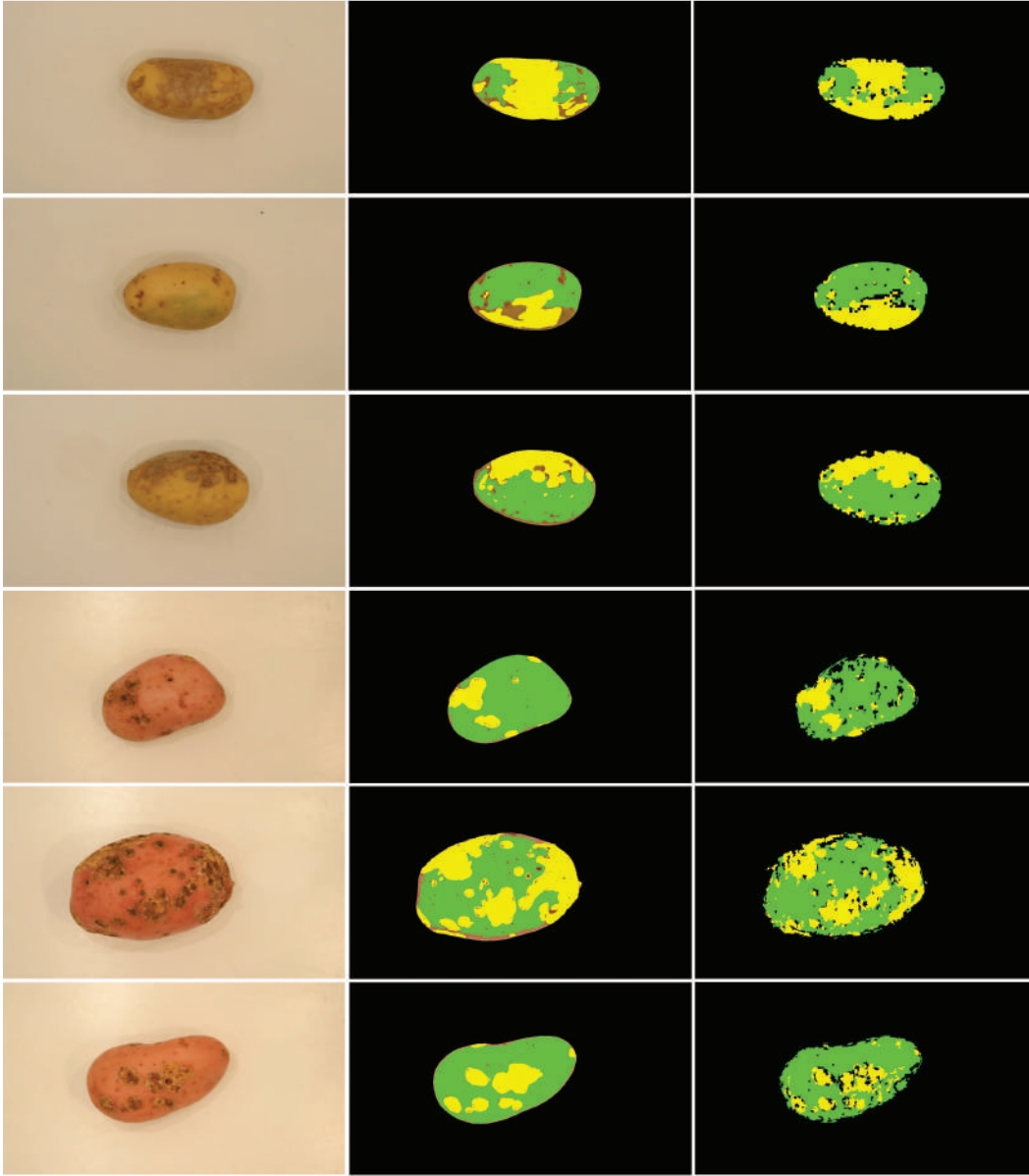
In the experiments, two datasets were used. One is a collection of images of 25 white potatoes and the other is that of 41 red potatoes. In each dataset, we only use the ground truth images of 5 potatoes for training. The reason that we only select a small number of images for training is that in real applications, users often tend to label a small number of superpixels in an interactive manner, which requires that the classifier must work well based on a limited amount of training data.

Some of the results are shown in Fig. 5 in association with the corresponding original images and the pixel-wise ground truth images marked by an expert in potato pathology. Note that some areas of high uncertainty or ambiguity were left unmarked, which is inevitable since even under a microscope, the expert could still be unsure about whether a particular region is good or not. These unmarked pixels are ignored during the training of the classifier. Also, our system relies on superpixel-based over-segmentation for the reasons of efficiency and convenient selection of training data. Some holes thus occur, particularly in the regions where the labels change as shown in Fig. 5 because one superpixel can be labelled into two different classes equally (equiprobable classes) so this leads to an unclassified superpixel. An alternative solution would be to choose one of the classes randomly.

To compare the results of the discussed prototype, the state-of-the-art method ‘ilastik’ [20] is used for training and testing with the same potato dataset. The ilastik system allows the user to perform segmentation and classification. Users need to mark up locations and provide training data pixel by pixel. This is typically a very tedious task. Classification is based on random forests. This system utilises up to three spatial and one spectral dimension in the feature calculation. It has capabilities like saving classifiers for later batch processing.

To perform the evaluation, we use the ground truth images of 5 potatoes in each dataset for training. This is actually impossible for the method in [20] in practice

<sup>2</sup> <http://youtu.be/0XKyxHy4Vkg>



**Fig. 5** Some of the results produced by the proposed system where we use green and yellow to denote good regions and blemishes respectively. Left column: original images; Middle column: ground truth images; Right column: our results.

because human users cannot mark up 5 images with a resolution of  $1280 \times 720$  pixel by pixel (but possible for our system based on superpixels which significantly mitigates the manual effort of mark-up). Even so, comparison results of the system in [20] to our system are shown in Fig. 6.

In these figures, we use the following equations to compute the accuracy in detecting the good (blemished) and the bad (unblemished) areas respectively.

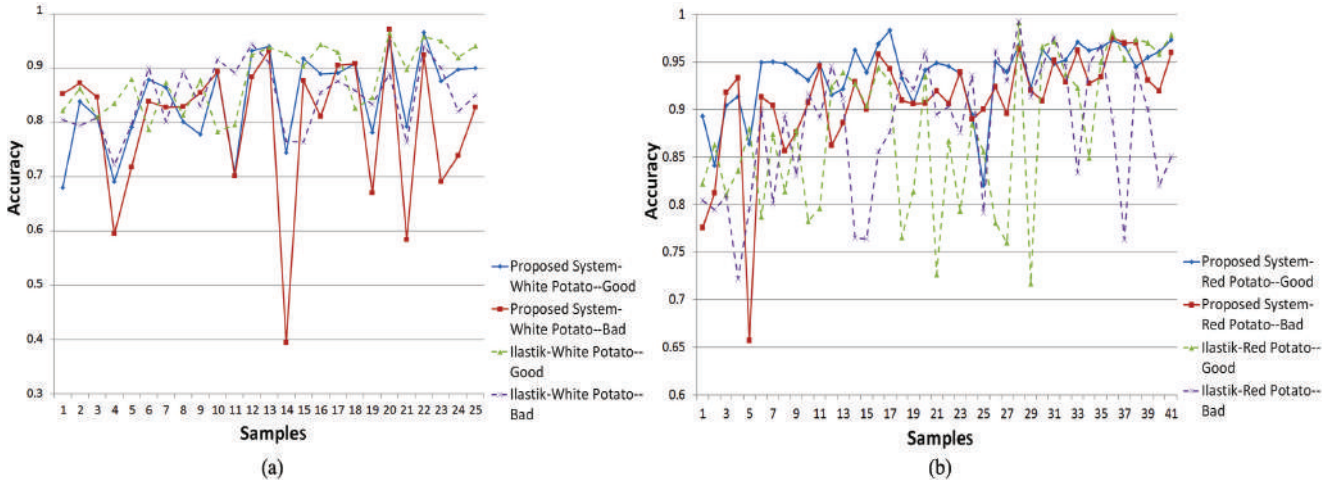
$$Acc_{good} = \frac{|G_r \cap G_g|}{|G_r \cup G_g|}, \quad Acc_{bad} = \frac{|B_r \cap B_g|}{|B_r \cup B_g|} \quad (4)$$

where  $G_g$  and  $B_g$  denote the good and the bad areas (calculated by the number of pixels) in the ground truth

images while  $G_r$  and  $B_r$  denote the good and the bad areas in the resultant images.

On average, the accuracy of the system in [20] in detecting the good areas for the white potatoes is 0.8807 and for the bad areas is 0.8455, while the average accuracy in detecting the good areas for the red potatoes is 0.8785 and for the bad areas is 0.8801. In comparison, the average accuracy of our system in detecting the good areas for the white potatoes is 0.8451 and for the bad areas it is 0.7983, while the average accuracy in detecting good areas for the red potatoes is 0.9372 and for the bad areas it is 0.9099. It can be seen that in terms of accuracy, our system is comparable with the





**Fig. 6** Comparative results of two-label image segmentation. (a) Results for 25 white potatoes trained and tested with the proposed system and the system in [20]; (b) Results for 41 red potatoes trained and tested with the proposed system and the system in [20]

Type	Unblemished	Black Dot	Sliver Scurf	Scab	Greening	Other Blemishes
Colour	Green	Blue	White	Cyan	Yellow	Magenta

**Table 1** The colour code for the visualisation of segmentation results of both white and red potatoes

state-of-the-art method. The results of the prototype system introduced in this paper, especially in detecting bad areas, can be improved by using some other features like frequency domain features. However, the prototype system provides a trade-off between accuracy and speed that makes it suitable for training in quality control tasks. Also, note that our method is based on labelling superpixel regions while in one superpixel there could be different types of potato regions (good or bad) but the competing method is pixel-wise. When we use the ground truth images of 5 potatoes to train our system, we always need to first convert the pixel-wise ground truth images into superpixel-wise ones. Such a conversion inevitably introduces errors. Furthermore, it is worth pointing out that in order to facilitate the evaluation, all of the experiments were implemented based on a ‘one-off’ operation, which means that we actually disable the powerful function of the system that allows the user to modify the training data interactively. In general, the segmentation output by our method can be easily improved by interactively modifying the training data. This can be demonstrated by our video on YouTube.

### 3.2 Multi-label image segmentation

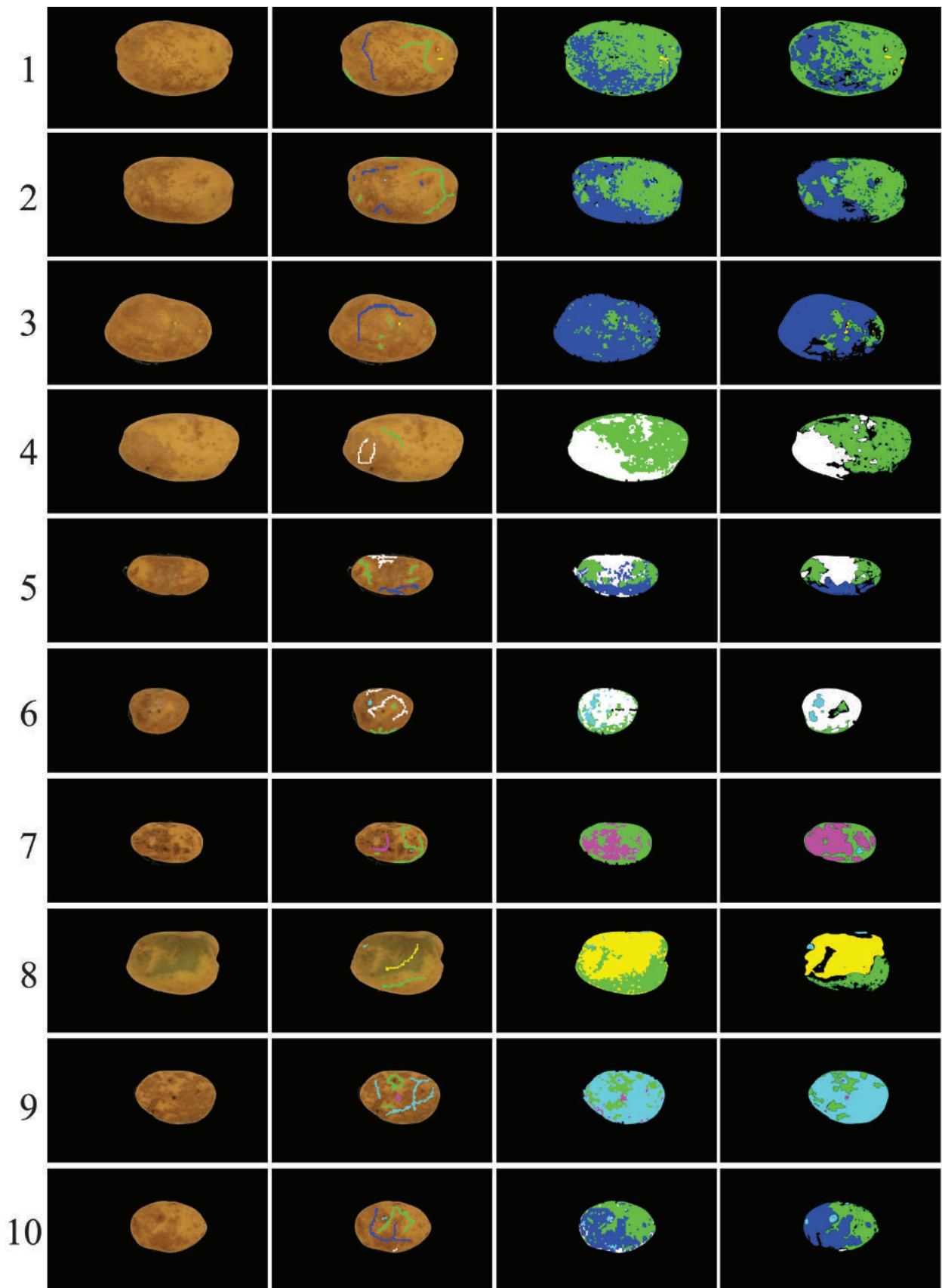
We also tested our system using 102 images of white potatoes and 41 images of red potatoes containing different types of blemishes. The results of 10 white potatoes are shown in Fig. 7. In Fig. 7, we use different

colours to represent different types of potato blemishes. The colour code of this multi-label image segmentation is shown in Table 1. It can be observed that our system utilises only a tiny amount of training data, typically several brush strokes composed of a couple of superpixels and output segmentations that are visually analogous to the corresponding ground truth images.

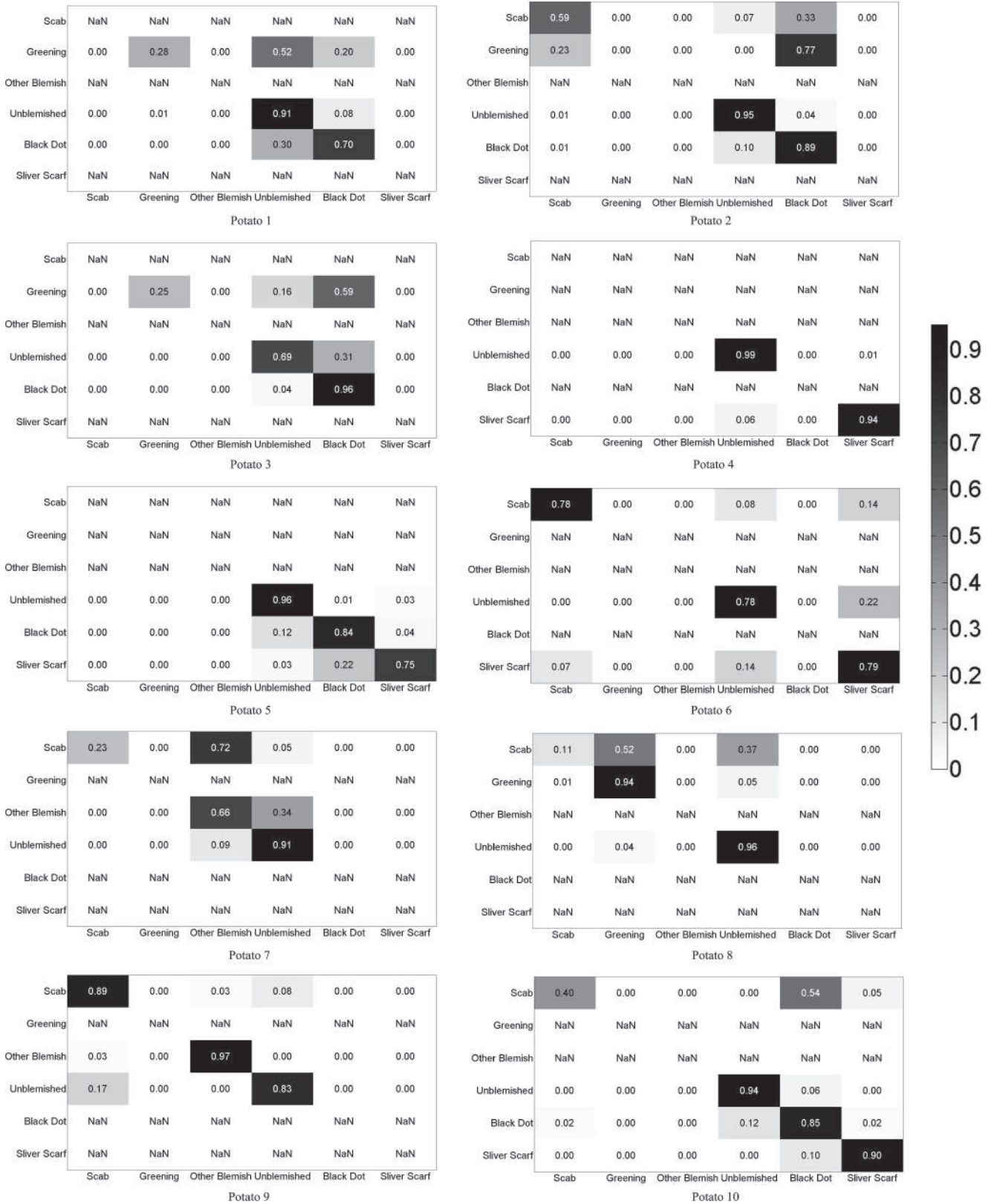
Based on the ground truth data, we managed to compute the confusion matrices for quantitatively evaluating the performance of our system. Fig. 8 shows the confusion matrices corresponding to the 10 white potatoes shown in Fig. 7.

For Potatoes 1 and 3, according to the confusion matrices shown in Fig. 8, the classification results for black dot and healthy areas are very good despite the fact that black dot is a type of blemish that can cover several pixels not a whole super-pixel.

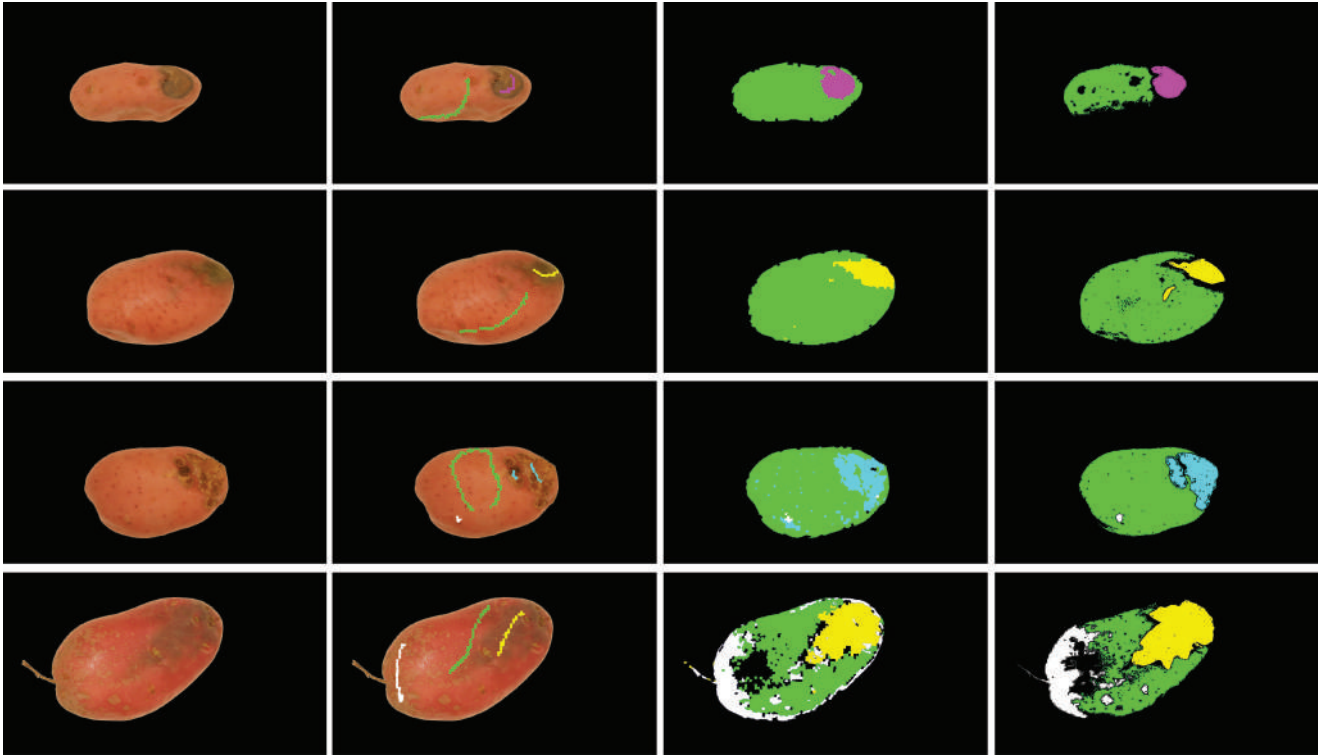
For Potato 2, it can be seen that the results for scab and greening are not good. The reason is that the region of greening is too tiny to be put in one or two superpixel(s) and it is hard to separate it for training and testing. In other words, the training data for greening is noisy. On the other hand, the colour information for the scab region contains some information from the neighbouring black dot region (for the superpixel algorithm it is very hard to segment these close neighbours) and this affects both training and testing. Overall, classification accuracy for unblemished areas and black dot is so high that this makes this system reliable in quality control tasks.



**Fig. 7** Segmentation results of 10 images of white potatoes. From left to right: Original images, training data, Segmentation results, ground truth images



**Fig. 8** Confusion matrices of the segmentation results of 10 white potatoes where 'NaN' denotes that there is no such type of blemish in the image.



**Fig. 9** Segmentation results of 4 images of red potatoes. From left to right: Original images, training data, Segmentation results, ground truth images

The segmentation result of Potato 4 demonstrates that the system is capable of distinguishing between silver scurf and healthy regions. This is due to the fact that our feature extraction covers colour information to recognise these areas.

The segmentation results for Potatoes 5 and 6 are very good. In the image of Potato 5, we have three classes which two of them are very hard to tell apart: black dot and silver scurf. For Potato 6, it can be seen that our system can distinguish scab and silver scurf reliably. The visual difference between scab and silver scurf is a key to have much less noisy data while training hence leading to a good segmentation.

Potato 7 contains three classes: unblemished area, scab and other blemish type. Colour information of other blemish type is similar to scab which makes it hard to distinguish. But, both blemish types are very different to the unblemished area by variations in colour information.

For Potato 8, the performance to distinguish between unblemished area and greening area is very good. But scattered scab regions are so tiny to be put in one or two whole superpixels. This makes training noisy.

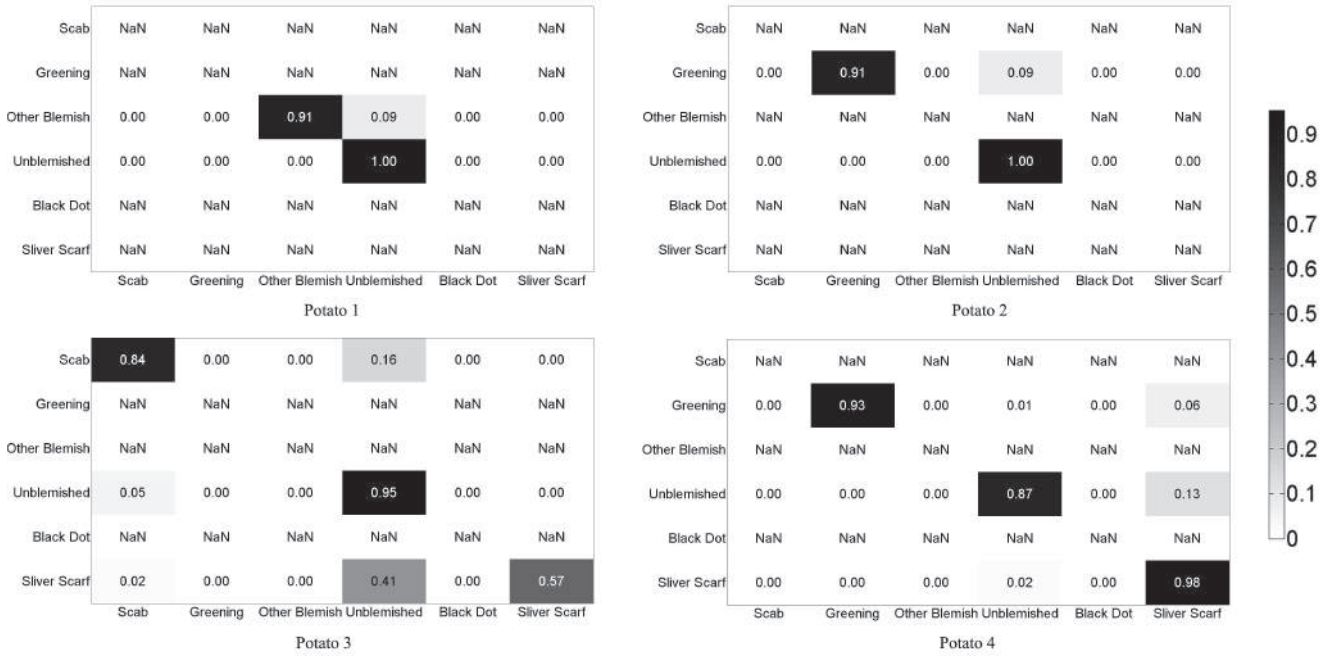
As a comparison to Potato 7, Potato 9 contains the same three blemish types. But since here the colour of other blemish type is quite different from that of the

scab, the system outputs a highly accurate segmentation result.

Potato 10 contains three blemishes: scab, black dot and silver scurf. According to the confusion matrix, the only downside is to detect scab region as they are mostly misclassified as black dot. Again this is due to the fact that superpixels can cover both regions in neighbouring parts which makes us a noisy training data.

Figs 9 and 10 show some of the results of the red potatoes. It can be seen that the performance of our system on the red potatoes are even better. This is mainly because the colour variation in the red potato images are generally larger than that in the white potatoes. In general, the experiments using both white and red potatoes demonstrate that the proposed system is quite effective to detect potato blemishes (or the defects of other types of foods such as fruits and beans) as long as the user can give correct training data in the interaction. Sometimes, even if the training data are slightly wrong or corrupted by noise, the system is still capable of producing good results.





**Fig. 10** Confusion matrices of the segmentation results of 4 red potatoes where ‘NaN’ denotes that there is no such type of blemish in the image.

### 3.3 Speed

Obviously the speed of the proposed system will vary, mainly depending on how many training data the user selects. And usually, the more selected training data, the more accurate the classification results. We thus demonstrate that our system can be implemented in real-time through our demonstration video on YouTube. The video recorded typical operations of the system using a PC with a dual core 2.13GHz CPU, 4GB RAM and a GeForce GTX 570 graphics card. It can be seen that the run time for classifying one image is usually 1 or 2 seconds, sometimes even less than 1 second.

## 4 Conclusions

The presented system based on interactive image segmentation shows that a “human-in-the-loop” scheme is crucial to deliver an effective and efficient solution to the real application of potato blemish detection. With the help of human expertise, a tiny amount of, but highly informative and discriminative training data are provided in an interactive manner. In comparison, traditional offline training requires a large amount of data, which inevitably slows down the processing but does not necessarily improve the segmentation results due to the redundancy within the training data. The experimental results also demonstrated that the introduction of superpixels for over-segmentation does not deterio-

rate the classification significantly but it accelerates the classification a lot.

There are several possible improvements to the image segmentation method used in our system. First, a pixel-wise refinement might be needed to further improve the accuracy of the classification. In particular, notably in Fig. 5 and Fig. 7, a number of disagreements between ground truth and classification results were located around the boundaries of ground truth blemishes. Albeit occasionally, some superpixels straddle two types of regions. In these cases, such superpixels have to be further segmented into the pixel level. Second, while the features are considered in the current classification are all texture features, the involvement of other types of features such as shape features could help to more accurately classify some specific potato blemishes which are distinctive mainly in terms of shape. Thirdly, the proposed system which estimates potato blemishes relies only on 2D information, while a real world scenario we would have to consider the entire surface area of a 3D tuber. 3D information is especially useful in the accurate computation of the areas of the blemishes for grading purpose since currently we actually compute the area of the 2D projection of a blemish region. This will cause errors, particularly in the boundary region of the potato shown in a 2D image. By conducting future research aiming at these improvements, we hope that we can further develop a more powerful and reliable system for real applications.

**Acknowledgements** This work is supported by Innovate UK. This support is gratefully acknowledged. Many thanks to Nigel Allinson and Chris Waltham for the design and build of the image capture set-up shown in Fig. 1. We also thank Graeme Stroud from Potato Council for preparing the ground truth images, and Michael Barnes for his earlier work on the project.

## References

1. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(11), 2274–2281 (2012)
2. Barnes, M., Duckett, T., Cielniak, G., Stroud, G., Harper, G.: Visual detection of blemishes in potatoes using minimalist boosted classifiers. *Journal of Food Engineering* **98**(3), 339–346 (2010)
3. Barni, M., Cappellini, V., Mecocci, A.: Colour-based detection of defects on chicken meat. *Image and Vision Computing* **15**(7), 549–556 (1997)
4. Boiman, O., Irani, M.: Detecting irregularities in images and in video. *International Journal of Computer Vision* **74**(1), 17–31 (2007)
5. Brosnan, T., Sun, D.W.: Improving quality inspection of food products by computer vision—a review. *Journal of Food Engineering* **61**(1), 3–16 (2004)
6. E.R., D.: *Machine Vision*. Academic Press Limited, London (1990)
7. Gunasekaran, S.: Computer vision technology for food quality assurance. *Trends in Food Science & Technology* **7**(8), 245–256 (1996)
8. Gunasekaran, S.: Computer vision technology for food quality assurance. *Trends in Food Science & Technology* **7**(8), 245–256 (1996)
9. Jackman, P., Sun, D.W.: Recent advances in image processing using image texture features for food quality assessment. *Trends in Food Science & Technology* **29**(1), 35–43 (2013)
10. Joutou, T., Yanai, K.: A food image recognition system with multiple kernel learning. In: *Proceedings of International Conference on Image Processing*, pp. 285–288 (2009)
11. Lee, K.M., Li, Q., Daley, W.: Effects of classification methods on color-based feature detection with food processing applications. *IEEE Transactions on Automation Science and Engineering* **4**(1), 40–51 (2007)
12. Lowe, D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60**(2), 91–110 (2004)
13. Manjunath, B.S., Ma, W.: Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI - Special issue on Digital Libraries)* **18**(8), 837–42 (1996). URL <http://vision.ece.ucsb.edu/publications/96PAMITrans.pdf>
14. Matsuda, Y., Yanai, K.: Multiple-food recognition considering co-occurrence employing manifold ranking. In: *Proceedings of International Conference on Pattern Recognition*, pp. 2017–2020 (2012)
15. Ojala, T., Pietikainen, M., Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(7), 971–987 (2002)
16. Ren, C.Y., Reid, I.: gSLIC: a real-time implementation of slic superpixel segmentation. Tech. rep., University of Oxford, Department of Engineering (2011)
17. Rocha, A., Hauagge, D.C., Wainer, J., Goldenstein, S.: Automatic fruit and vegetable classification from images. *Computers and electronics in agriculture* **70**(1), 96–104 (2010)
18. Russo, R., Shah, M., Lobo, N.: A computer vision system for monitoring production of fast food. In: *Proceedings of Asian Conference on Computer Vision*, pp. 769–775. Melbourne, Australia (2002)
19. Schapire, R.E., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. *Machine learning* **37**(3), 297–336 (1999)
20. Sommer, C., Straehle, C., Kothe, U., Hamprecht, F.A.: ilastik: Interactive learning and segmentation toolkit. In: *Proceedings of IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 230–233 (2011)
21. Tao, Y., Morrow, C., Heinemann, P., Sommer, H.: Fourier-based separation technique for shape grading of potatoes using machine vision. *Transactions of the ASAE* **38**(3), 949–957 (1995)
22. Wu, D., Sun, D.W.: Colour measurements by computer vision for food quality control—a review. *Trends in Food Science & Technology* **29**(1), 5–20 (2013)
23. Yang, S., Chen, M., Pomerleau, D., Sukthankar, R.: Food recognition using statistics of pairwise local features. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2249–2256 (2010)
24. Zheng, C., Sun, D.W., Zheng, L.: Recent developments and applications of image features for food quality evaluation and inspection—a review. *Trends in Food Science & Technology* **17**(12), 642–655 (2006)
25. Zhou, L., Chalana, V., Kim, Y.: Pc-based machine vision system for real-time computer-aided potato inspection. *International journal of imaging systems and technology* **9**(6), 423–433 (1998)