

Visual Detection of Blemishes in Potatoes using Minimalist Boosted Classifiers

Michael Barnes¹, Tom Duckett¹, Grzegorz Cielniak¹, Graeme Stroud² and Glyn Harper²

¹*School of Computer Science, University of Lincoln, Lincoln LN6 7TS, UK*

²*Potato Council Ltd., Sutton Bridge Experimental Unit, Spalding PE12 9YD, UK.*

Abstract

This paper introduces novel methods for detecting blemishes in potatoes using machine vision. After segmentation of the potato from the background, a pixel-wise classifier is trained to detect blemishes using features extracted from the image. A very large set of candidate features, based on statistical information relating to the colour and texture of the region surrounding a given pixel, is first extracted. Then an adaptive boosting algorithm (Adaboost) is used to automatically select the best features for discriminating between blemishes and non-blemishes. With this approach, different features can be selected for different potato varieties, while also handling the natural variation in fresh produce due to different seasons, lighting conditions, etc. The results show that the method is able to build “minimalist” classifiers that optimise detection performance at low computational cost. In experiments, blemish detectors were trained for both white and red potato varieties, achieving 89.6% and 89.5% accuracy, respectively.

Key words:

1. Introduction

Potatoes (*Solanum tuberosum*), with an estimated worldwide production of over 300,000,000 tonnes in 2005 (Food and Agriculture Organisation, 2005), account for 70-80% of the carbohydrate consumed in the UK with millions of tons harvested worldwide. For the fresh market the main factor affecting consumer preference is physical appearance and, to maximise return, great effort is expended ensuring that the appearance best matches

a particular market. There are no current legislation standards for tuber blemishes but standards are driven by market forces, principally by the requirements of the larger supermarkets' customers. Most potatoes are still sorted by hand. Problems with manual sorting include the subjectivity, fatigue and high cost of human inspectors, while currently deployed artificial vision systems require manual calibration and have limited accuracy. There are a number of biological techniques used to identify potato diseases, including microscopy and methods based on the recognition of proteins or DNA of the pathogen, e.g. De Haan and van den Bovenkamp (2005). However, although the methods are accurate and may be sensitive enough to detect disease before symptoms develop, they are time-consuming and not suited to the individual assessment of potatoes on a commercial scale. Thus there is considerable motivation for automating visual inspection.

In typical machine vision systems for quality analysis of food products, there are several major steps: after pre-processing (e.g. to segment the object of interest from the background), image features are extracted that summarise important qualities of the object, then a pattern recognition system is used to categorise the input data. For example, Bolle et al. (1996) developed the VeggieVision system, using HSV-colour and texture histograms to classify different types of fruit and vegetables, with application to a supermarket check-out for automatic produce recognition. Unay and Gosselin (2006) developed methods to distinguish between blemishes in apples and healthy apples with visible stem or calyx. Images were recorded using special filters to restrict the observed light frequencies, then various features including statistical moments and shape features were used for pattern recognition. Jelinski et al. (2007) introduced visual inspection methods for pasteurised cheese. They also used thresholding to detect ingredients such as chives, and developed methods to measure the distribution and quantity of the detected ingredients. Munkevik et al. (2007) developed a machine vision system for automatic descriptive sensory evaluation of meals, where a neural network was trained to mimic the opinion of human experts in describing the sensory attributes of a prototypical meal. Jarimopas and Jaisin (2008) introduced a system for sorting sweet tamarind, by measuring the size and shape of tamarind pods as well as detecting defects in the form of broken pods. Thresholded intensity values were used to distinguish blemishes from non-blemishes.

In the area of machine vision for potatoes, Tao et al. (1995b) used Fourier harmonics to describe the shapes of potatoes, forming a metric based on the

first ten Fourier harmonics of the potato's outline to develop a classification method which agreed with human classification 89.2% of the time. Muir et al. (1999) used custom lighting equipment to project light at a variety of different wavelengths to demonstrate the different reflective properties of specific blemishes at each wavelength. This work was commissioned in part by R.J. Herbert Engineering for use in their Upgrader product line (R.J. Herbert Engineering Ltd, 2008). Tao et al. (1995a) describes the use of the HSI colour space for identifying greened potatoes as well as yellow and green apples. This was done by use of histograms produced from each of the HSI channels. It was noted that more bins in a histogram resulted in a higher performance. Heinemann et al. (1996) graded potatoes by size and shape to meet United States Department of Agriculture (USDA) standards. Size was measured by the longest distance between two points on the boundary, while the shape was determined using Fourier descriptors. The system achieved 97-98% accuracy when classifying stationary potatoes but dropped to between 77% and 88% when tested on moving potatoes. Zhou et al. (1998) developed a system using green levels to detect green defects (greening and sprouting) in individual 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. These were compared to USDA standards with an overall success rate of 86.5% with a false positive rate of 57.1%. Without the blemish detection the results improved to 90% overall success rate with 17% false positives. Guannan et al. (2009) detected misshapen potatoes by comparing the local rate of change of the radius of a potato. In addition they detected sprouting using a comparison of the green colour channel with the intensity. This value at each pixel was compared to the average value across the potato and if the difference was above a threshold the pixel was determined to be part of a sprout. We have also investigated the use of features that are summarised across the whole potato.

A limitation of typical machine vision systems is that the set of image features for pattern recognition has to be designed by the system engineer to work with a specific configuration of produce, imaging system and operating conditions. Such systems typically do not generalise well to other configurations, where the required image features may well differ from those used to design the original system. The novelty of the approach presented in this paper involves the use of an adaptive boosting algorithm called Adaboost (Freund and Schapire, 1999) to automatically select good features for a particular pattern recognition task. A minimal set of features is selected

from a very large set of *candidate features*, which measure statistical properties of the colour and texture distribution of the image region surrounding a given pixel. Thus the *selected features* used to build the final pattern recognition system are optimised for a particular application by learning from examples, and the system can be retrained to select a different set of features in order to accommodate different varieties of produce, seasonal variations, etc.

The objective of this research is to introduce an automatic method for detecting blemishes in digital images of potatoes. The system developed should be trainable, so that it can work with different varieties of potatoes and variations in seasons, lighting conditions, etc. A human expert is required to mark up areas of blemishes and non-blemishes in a set of training images. After training, the system should be able to classify individual pixels as blemishes or non-blemishes with high accuracy. A further objective, with eventual deployment in industrial settings in mind, is to enable real-time processing of images (possibly in rapid succession) by building “minimalist” classifiers that extract a minimal subset of all features that optimise detection performance at the lowest possible computational cost. Finally the feature selection mechanism developed should be perspicuous to human users, allowing operators to understand which features are important to distinguish blemishes from non-blemishes for different potato varieties. The contribution of the paper towards meeting these objectives is demonstrated in experiments by learning minimalist blemish detectors for both white and red potatoes, achieving 89.6% and 89.5% accuracy, respectively.

1.1. Potato Blemishes

There are a number of conditions affecting potato tubers that, although superficial and of no health consequence to humans, strongly and negatively influence consumer choice. These include black dot, silver scurf, powdery scab, common scab and skin spot. The fungal species of *Rhizoctonia solani* also causes significant skin blemish as black scurf and elephant hide. The causes of blemishes are known. However, customer preference may be for susceptible potato varieties, and different environmental and field conditions during cultivation favour different diseases. These inevitably lead to some crops being infected in a generally unpredictable fashion. Other forms of blemish include physical damage, e.g. growth cracks, mechanical damage and slug damage, as well as physiological effects, e.g. greening and sprouting. Figure 1 shows examples of several common blemishes. Potatoes and

their tubers are also susceptible to more significant diseases, in particular blight, and other fungal and bacterial rots that are outside the scope of the work presented here. Blemish conditions present a variety of different coloured, sized and textured symptoms on the skin surface. Such diverse visual information provides us with a rich source of indicators that can be used for training an automatic blemish detector. In this paper, our collected image data contained the following subset of these potato blemishes: black dot, silver scurf, common scab, slug damage, greening, powdery scab, elephant hide, sprouting and growth cracks. These were the blemishes present in the potato samples collected from the Spring 2008 harvest at the Sutton Bridge Experimental Unit of the Potato Council Limited, UK. In this article our research focus is on detection of blemish versus non-blemish.

2. Materials and methods

2.1. Image Acquisition

The experimental data for this system, consisting of images of potatoes, were acquired using a colour camera (Sony DSLR-A350K) fixed above the tubers, which in turn were placed on a white board. The camera was set to autofocus at a distance of 60cm from the camera objective to the base on which the subjects were placed, with a focal length of 70 mm and an aperture setting of F22. The resolution of the images was 1536×1024 pixels. In industrial settings a different camera might be used, but due to the use of machine learning classifiers in our approach, the system should work with any similar colour imaging device. A pixel in such an image covers an area of around 0.02 mm^2 . To reduce the effects of shadows and changing light conditions the potatoes were placed inside a white cylinder with daylight bulbs placed around the top. The equipment used to capture these images is shown in Figure 2.

2.2. Data Sets and Ground Truth Information

There were two sets of data collected for white and red potatoes, respectively, including potatoes affected by different blemishes. The white potato data set consisted of 102 images including 19 images containing a single blemish type, 39 images with two distinct blemish types, 38 images with three blemish types and 6 images containing more than three blemish types. The most common blemishes were black dot and silver scurf, appearing in 69 and 53 images, respectively, while the rarest were powdery scab, elephant hide

and growth cracks, with no more than 3 images of each. The red potato data set consisted of 22 images with the most common blemishes again being black dot and silver scurf, appearing in 13 and 6 images, respectively, as well as common scab, appearing in 12 cases. 10 red potato images had 2 different blemish types, 3 images had 3 types, 1 image had 4 types, and the remaining 8 images had only one blemish type.

To train the classifiers and test their performance, the images need to be marked up by hand to provide the “ground truth” information indicating the correct class of each pixel. The mark up process begins with a semi-automatic method for background removal, using the Magic Wand tool in Adobe Photoshop to label the image region surrounding the potato. The potato area is then hand labelled by an industry expert into regions corresponding to blemish and non-blemish. It is not necessary to label all pixels in an image: some areas of high uncertainty or ambiguity were left unmarked in our experiments, and these pixels are ignored during training of the classifier. Background pixels are also omitted from the subsequent calculations.

Some examples of the obtained ground truth images can be seen in the left column of Figure 7.

2.3. Feature extraction

The first step of the procedure is to extract different image features that should indicate the presence or absence of blemishes in a potato image. The features include statistical summaries of the whole potato and local regions centred on each pixel as well as the data of the pixel itself. The statistics used were the mean, variance and skew of various image properties listed below. Other systems have used only the mean of the region such as Tsai and Tsai (2002) or histograms as in Bolle et al. (1996). The proposed system uses the RGB colour space, the original colour format of the camera output. An alternative solution would be to use the HSI colour space but this would create an additional processing overhead including colour conversion and calculation of circular statistics. Other systems use more complex hardware set-ups such as customised lighting, as is an option for the Maf-Roda Agrobotic (Maf Roda Group, 2008) or using specific colour filters as in Unay and Gosselin (2006). The code for the system software was implemented in MATLAB.

The image regions used for feature extraction in our experiments were squares of size 33×33 , 65×65 , 97×97 , 129×129 , and 161×161 pixels, plus the whole potato, giving 6 regions in total. Our system uses seven

colour channels; raw RGB, normalised RGB and the intensity channel $I = \frac{1}{3}(R+G+B)$. From these channels we consider the following image properties:

Colour. Intensity is especially of relevance for dark blemishes, e.g. black scurf or skin spot, while the most obvious blemish to be detected by other colour channels would be greening. The three statistical moments collected from the seven colour channels represent the first 21 features for each region. Seven additional features describe the colour properties of the pixel itself.

Edges. An edge detector determines the rate of change of pixel values in a given neighbourhood in a specific direction. Some blemishes tend to coincide with high rates of change, such as powdery scab when the skin splits. The Sobel edge detector was used in this case with a standard 3×3 kernel size. The edge detector was run on the same seven colour channels listed above. These statistics provide 21 features for each region and seven for the pixel itself.

Range. The range filter determines the maximum difference between pixel values in a given neighbourhood indicating the roughness of the texture. Higher values tend to correspond to rougher, potentially damaged areas of the potato. The range filter was run on the same seven channels with a 5×5 neighbourhood. The three statistical moments collected from the resulting range information provide another 21 features for each region. Seven additional features describe range properties of the pixel itself.

In summary there are 7 colour channels \times 3 feature types \times 3 statistical moments, making 63 features for each region and $7 \times 3 = 21$ features for the pixel itself. All these features are used as the candidate feature set. Since there are 63 different features that can be extracted from each region, this gives $63 \times 6 = 378$ features which, with additional 21 features for the pixel itself, gives us $F_c = 399$ candidate features in total. These features were used as the training input to our classifier.

2.4. Classification

The AdaBoost algorithm (Freund and Schapire, 1999) is used to build a classifier, which combines results from so-called “weak” classifiers (each constructed using one of the candidate features) into one “strong” classifier that performs better than any of the weak classifiers alone. It has been used previously in the classification of apples to avoid falsely classifying apple features as blemishes (Unay and Gosselin, 2006). The high performance of

the final strong classifier is due to the emphasis put on the training examples which are most difficult to classify during the learning process. This method is called boosting. During training AdaBoost makes a number of passes, called rounds or iterations, through the training data. Each time it finds the next best feature to improve the number of correctly classified examples, prioritising those examples which were misclassified previously. In each pass one feature is selected and assigned a weight and a threshold to create a new weak classifier. The weak classifiers are then combined into a strong classifier wherein each weak classifier is given a weighted vote in the classification of a given example.

The Real AdaBoost algorithm proposed by Schapire and Singer (1998) is a generalisation of this algorithm that provides a lower error rate by allowing weak classifiers to vote by their individual degree of certainty instead of simply voting yes or no. It is the version used in our experiments, hereafter referred to simply as AdaBoost.

In our system the AdaBoost algorithm is used to classify individual pixels of potato images into two categories: blemished potato and good potato based on features defined in Section 2.3. The reduced set of selected weak classifiers allows for preprocessing only the most useful features, saving considerable computation time.

2.4.1. Minimalist classifier

The AdaBoost classifier selects a set of the most useful features from all candidate features. If the training data is not normally distributed, AdaBoost will often choose the same feature for more than one weak classifier. Therefore it is of interest to see how much the classification success rate would be affected by the original candidate feature set being restricted to a subset of features, selected by AdaBoost itself. By doing so the total number of unique features required to be extracted for classification will be reduced and therefore less computational time will be required by the feature extraction stage. We refer to this subset of features as the “selected features”.

Our approach involves two stages, both incorporating the AdaBoost algorithm: the first stage selects a feature set that will be used to train an AdaBoost classifier in the second stage. Algorithm 1 presents the Real AdaBoost algorithm as described in Huang et al. (2005), applied with our addition of step 4, 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. Using MATLAB we have extended the AdaBoost implementation

within the GML AdaBoost Toolbox (Vezhnevets, 2006), to build the minimalist classifier.

2.5. Evaluation metrics

The output of the classifier is a binary image with pixels indicating good potato or blemish. The performance of the system can be measured by comparing the output image to the ground truth information. The following statistics were collected for each output image:

- TP - true positive, number of pixels that were classified as blemish and matched ground truth;
- FP - false positive, number of pixels that were classified as blemish but did not match ground truth;
- TN - true negative, number of pixels that were classified as good potato and matched ground truth;
- FN - false negative, number of pixels that were classified as good potato but did not match ground truth.

From these statistics we could calculate the two following metrics:

- sensitivity = $\frac{TP}{TP+FN}$;
- specificity = $\frac{TN}{TN+FP}$.

We represent the performance of our binary classifier using the most common method based on ROC curves, which provide detailed information about the relationship between these two metrics with respect to different parameter settings of the system.

3. Results and discussion

3.1. Training and testing

When training, the minimalist classifier first chooses a number of “selected features” (F_s). To investigate the impact of this parameter on the classification rate we used $F_s = 1, 2, 5$ and 10. For comparison we also used $F_s = F_c$, the equivalent of a non-minimalist method. The number of AdaBoost rounds in the second stage was set to $T = 40$.

Tests were carried out using the training data on a hold-one-out basis whereby one of the images is removed from the training data and used as test data instead. The testing is carried out for every image in this manner with each image being tested by a classifier trained on every thousandth pixel from the other images. The success rates of the minimalist classifier for different potato colours and different values of F_s are presented in Table 1. In addition, the performance of the classifier using ROC curves is presented in Figure 3 for white potatoes and in Figure 4 for red potatoes.

To determine the importance of different feature categories (i.e. colour, edge and range), the tests were carried out for different subsets of these categories. The results are presented as ROC curves in Figure 5 for white potatoes and in Figure 6 for red potatoes.

Figure 7 shows the output of the classifier compared to the ground truth information. Some of the disparity between the classifier output and ground truth could be due to human inaccuracy at the markup stage which can be seen more clearly in Figure 8.

3.2. Preferred features

The top ten chosen features for white potatoes can be found in Table 3 and for red potatoes in Table 4. In the top 5 features chosen by the minimalist classifier for both white and red potatoes the most prominent features were based on the edge detector and range filter run on the red channel followed by features based on intensity. The preferred regions were the pixel itself and the smallest region of size 33×33 pixels.

3.3. Success rates

Results presented in Table 1 for different number of selected features indicate that using $F_s = 10$ features does not negatively impact the performance of the classifier, resulting in success rates of 89.6% and 89.5% for white and red potatoes, respectively. In our tests the minimalist classifier came very close to or even slightly outperformed the non-minimalist classifier. Further reducing F_s to 5 still gives satisfactory results but the performance drops noticeably below that number. The difference in performance between classifiers using different values of F_s can also be seen clearly in ROC curves (Figures 3 and 4).

The results of using different subsets of feature categories are shown in Table 2. Using only colour gives a result of 86.9% accuracy for white potatoes

and 82.2% for red potatoes. Adding edge features gives an increase of classification rate up to 87.5% and 88.2% for white and red potatoes, respectively. On the other hand range features give an increase of classification rate up to 90.2% and 88.0% for white and red potatoes, respectively. This indicates that range features provide more relevant information than edges. Including all features does not greatly affect the classification rates, resulting in 89.7% and 88.7% accuracy for white and red potatoes, respectively. The difference in performance between classifiers using different feature categories can also be seen clearly in the ROC curves (Figure 5 and Figure 6).

The reduced performance when using only colour for the red potatoes may be related to the importance of the red colour channel in white potatoes. Since red potatoes have a smaller dynamic range in this colour channel the classifier may need to rely on other features. To ensure this was not simply due to the smaller number of training examples, the test was re-run on white potatoes using only 21 images, every fifth image. This resulted in similar results to before, with colour-only classification rates only dropping from 86.9% to 86.3%, which would tend to confirm the above hypothesis.

4. Conclusions and Further Work

The presented results show that an AdaBoost based system is able to build minimalist classifiers that optimise detection performance at low computational cost. A minimalist classifier using only ten selected features achieves success rates of 89.6% for white potatoes and 89.5% for red potatoes. The use of AdaBoost in this minimalist form provides a comparable and sometimes slightly better result than simply providing the whole feature set. This may be related to AdaBoost’s vulnerability to overfitting (Ratsch et al., 1998), meaning that the learned statistical model describes random error or noise in the training data instead of the underlying relationship between the classifier inputs and outputs.

A number of disagreements between ground truth and classification results were located on the edges of ground-truthed blemishes, visible in Figure 7. Many of these disagreements may be due to human error in the markup stage, which can be seen as symptomatic of the problem which this research sets out to solve, that of human assessment of blemishes being subjective and prone to error. A machine vision system is likely to be more accurate than the human who produces the ground truth. This is especially noticeable in Figure 8 where a larger area has been marked in the ground truth as being

affected by black dot than detected by the classifier. The errors are due to the ground truth including an area marked as black dot, which is actually good potato skin speckled with black dot blemish. The classifier is able to detect the blemish pixel by pixel, so a portion of the reported error seems to be caused by inaccurate ground-truthing rather than misclassification.

The issue of human ground-truth inaccuracy might also be addressed by a semi-automatic method, e.g. by using an unsupervised method which would cluster similar potato features and allow the human to select clusters and specify what they represent.

There are a number of possible improvements to the image processing method used in our system. Some initial research has suggested it might be possible to replace the fixed square regions used in our experiments with regions from segmentation algorithms, e.g. normalised cuts (Mori, 2005). Also the use of textons (Varma and Zisserman, 2005) to provide additional texture information might improve results. However the processing time needed to follow the method used in Varma and Zisserman (2005) made it unappealing to pursue. Further tests have yet to confirm a possible gain in accuracy or speed of these approaches. Other features that might be involved would include shape features, if the boundaries of individual blemishes can be located, and Fourier harmonics as in Tao et al. (1995b).

The proposed system estimates blemished areas of potatoes in 2D images, while a real world scenario we would have to consider the entire surface area of a 3D tuber. The question has been already addressed in industry, for example, with the Herbert Upgrader (R.J. Herbert Engineering Ltd, 2008) which takes a selection of images of a rolling potato and averages the area of each. Alternatively, the Hiquip system (Noordam et al., 2000), developed in the Netherlands, uses a single camera to take photographs of potatoes passing a series of mirrors to ensure full coverage.

It has been shown that the smaller dataset for red potatoes does not cause a large reduction in the classifier’s performance compared to the 102 images used for white potatoes. Given that the minimalist classifier was intended to reduce the amount of data required to classify a potato, the question remains of how small a selection of white potatoes would be necessary to still achieve a satisfactory classification rate.

An overall aim of our work is to provide improved management and control strategies for the individual blemish diseases. Further research is directed at distinguishing between blemish types, which will provide a significant tool to investigate, for example, trends in individual diseases across time, geo-

graphic areas or weather.

5. Acknowledgements

This work was partly funded by the Potato Council Ltd. who also provided expert assistance in the field of potato science. Thanks especially to Adrian Cunningham. In addition thanks to Hongying Meng and William I. Tyne for their help in understanding Guannan et al. (2009) and De Haan and van den Bovenkamp (2005), respectively.

References

- Bolle, R., Connell, J., Haas, N., Mohan, R., Taubin, G., 1996. Veggievision: A produce recognition system. In: Proceedings of the 3rd IEEE Workshop on Applications of Computer Vision (WACV '96). IEEE Computer Society, Washington, DC, USA, p. 244.
- De Haan, E. G., van den Bovenkamp, G. W., 2005. Improved diagnosis of powdery scab (*spongospora subterranea* f.sp. *subterranea*) symptoms on potato tubers (*solanum tuberosum* l.). *Potato Research* 48 (1-2), 1–14.
- Food and Agriculture Organisation, 2005. FAO Statistics. Online, (<http://faostat.fao.org>).
- Freund, Y., Schapire, R., September 1999. A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence* 14 (5), 771 – 780.
- Guannan, Z., Yuzhi, T., Junxiong, Z., Wei, L., 2009. Automatic detecting and grading method of potatoes with computer vision. *Nongye Jixie Xuebao / Transactions of the Chinese Society of Agricultural Machinery* 40 (4), 166–168+1.
- Heinemann, P. H., Pathare, N. P., Morrow, C. T., 1996. An automated inspection station for machine-cision grading of potatoes. *Machine Vision and Applications* 9 (1), 14–19.
- Huang, C., Wu, B., Al, H., Lao, S., 2005. Omni-directional face detection based on Real AdaBoost. In: *International Conference of Computer Vision*.
- Jarimopas, B., Jaisin, N., 2008. An experimental machine vision system for sorting sweet taramind. *Journal of Food Engineering* 89, 291 – 297.

- Jelinski, T., Jin Du, C., Sun, D.-W., Fornal, J., 2007. Inspection of the distribution and amount of ingredients in pasteurized cheese by computer vision. *Journal of Food Engineering* 83, 3 – 9.
- Maf Roda Group, October 2008. Meeting with Maf-Roda personnel. (<http://www.maf-roda.com/>).
- Mori, G., 2005. Guiding model search using segmentation. In: *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*. Vol. 2. pp. 1417 – 1423.
- Muir, A. J., Ross, D. W., Dewar, C. J., Kennedy, D., 1999. Defect and disease detection in potato tubers. In: *Proceedings of SPIE - The International Society for Optical Engineering*. Vol. 3543. pp. 199–207.
- Munkevik, P., Hall, G., Duckett, T., 2007. A computer vision system for appearance-based descriptive sensory evaluation of meals. *Journal of Food Engineering* 78, 246 – 256.
- Noordam, J., Otten, G., Timmermans, T., van Zwol, B., March 2000. High-speed potato grading and quality inspection based on a color vision system. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*. Vol. 3966 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. pp. 206–217.
- Ratsch, G., Onoda, T., Muller, K. R., 1998. An improvement of AdaBoost to avoid overfitting. In: *Proceedings of the International Conference on Neural Information Processing*. pp. 506–509.
- R.J. Herbert Engineering Ltd, October 2008. Meeting with representatives of R.J. Herbert Engineering Ltd., Marshland St. James, Cambridgeshire, UK. (<http://www.rjherbert.co.uk/>).
- Schapire, R., Singer, Y., 1998. Improved boosting algorithms using confidence-rated predictions. In: *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*. pp. 80–91.
- Tao, Y., Heinemann, P. H., Varghese, Z., 1995a. Machine vision for color inspection of potatoes and apples. *Transactions of the American Society of Agricultural Engineers* 38 (5), 1555–1561.

- Tao, Y., Morrow, C. T., Heinemann, P. H., 1995b. Fourier-based separation technique for shape grading of potatoes using machine vision. *Transactions of the American Society of Agricultural Engineers* 38 (3), 949–957.
- Tsai, D., Tsai, Y., 2002. Rotation-invariant pattern matching with colour-ring projection. *Pattern Recognition* 35, 131–141.
- Unay, D., Gosselin, B., 2006. Stem and calyx recognition on 'jonagold' apples by pattern recognition. *Journal of Food Engineering* 78, 597 – 605.
- Varma, M., Zisserman, A., 2005. A statistical approach to texture classification from single images. *International Journal of Computer Vision* 62, 61–81.
- Vezhnevets, A., 2006. GML AdaBoost MATLAB Toolbox.
URL <http://research.graphicon.ru>
- Zhou, L., Chalana, V., Kim, Y., 1998. Pc-based machine vision system for real-time computer-aided potato inspection. *International Journal of Imaging Systems and Technology* 9 (6), 423–433.

6. Figures and tables

Algorithm 1 Our implementation of the Real AdaBoost learning algorithm.
Freund and Schapire (1999); Schapire and Singer (1998)

Given a dataset $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$ where $x_i \in X$ and $y_i \in \{-1, +1\}$, the weak classifier pool K , containing all possible weak classifiers from F_c candidate features, a specific number of weak classifiers to be chosen T and a maximum number of unique features to be used to choose these weak classifiers F_s .

Initialise the sample distribution $D_1(i) = 1/m$

For $t = 1, \dots, T$

1. For each weak classifier h in K do:
 - a. Partition X into several disjoint blocks X_1, \dots, X_n
 - b. Using the weights in distribution D_t calculate

$$W_t^j = P(x_i \in X_j, y_i = l) = \sum_{i: x_i \in X_j, y_i = l} D_t(i)$$

Where $l = \pm 1$

- c. Set the output of h on each X_j as

$$\forall x \in X_j, h(x) = \frac{1}{2} \ln \left(\frac{W_{+1}^j + \epsilon}{W_{-1}^j + \epsilon} \right)$$

- d. Calculate the normalisation factor

$$Z = 2 \sum_j \sqrt{W_{+1}^j W_{-1}^j}$$

2. Select the h_t Minimising Z i.e.

$$Z_t = \min_{h \in K} Z$$

$$h_t = \arg \min_{h \in K} Z$$

3. Update the sample distribution

$$D_{t+1}(i) = D_t(i) \exp[-y_i h_t(x_i)]$$

and normalise D_{t+1} to give a probability distribution function.

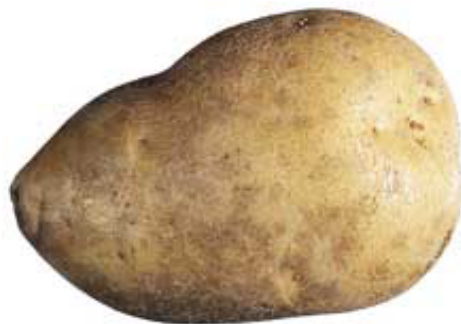
4. Count the number of unique features used by all weak classifiers. If the total equals F_s then update K to only contain weak classifiers pertaining to already selected features.

The final strong classifier H is

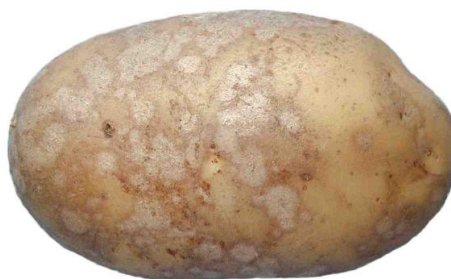
$$H(x) = \text{sign} \left[\sum_{t=1}^T h_t(x) - b \right]$$

The confidence of H is defined as

$$\text{Conf}_H(x) = \left[\sum_{t=1}^T h_t(x) - b \right]$$



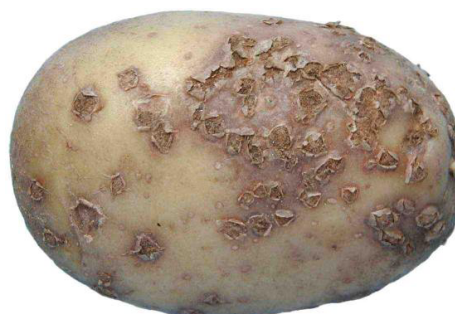
(a) A potato blemished by black dot.



(b) A potato blemished by silver scurf.



(c) Potatoes blemished by common scab.



(d) Potato blemished by powdery scab.



(e) Potato affected by skin spot.



(f) Physiological greening of the skin of a potato.

Figure 1: Examples of potato blemishes

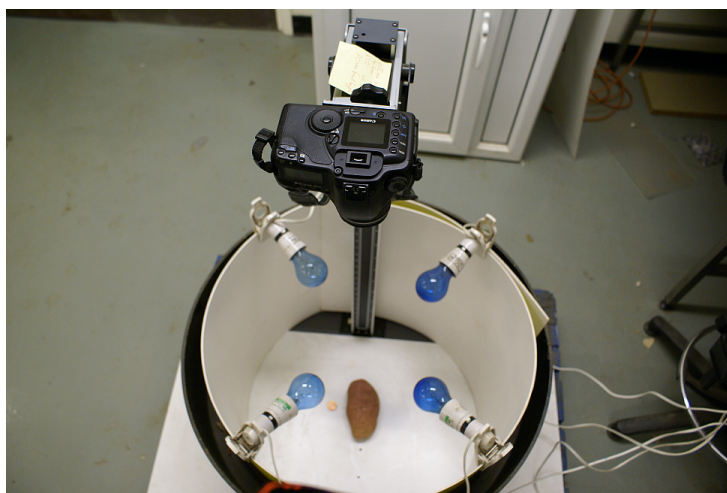


Figure 2: The camera setup for photographing the training data at a constant distance with all-around lighting.

Features	Success Rate	
	White Potato	Red Potato
$F_s = 1$	82.7%	84.8%
$F_s = 2$	87.6%	84.5%
$F_s = 5$	89.8%	88.7%
$F_s = 10$	89.6%	89.5%
$F_s = F_c$	89.7%	88.7%

Table 1: Success rates for different numbers of selected features F_s

Features	Success Rate	
	White Potato	Red Potato
colour only	86.9%	82.2%
colour and edges	87.5%	88.2%
colour and range	90.2%	88.0%
colour, range and edges	89.7%	88.7%

Table 2: Success rates for different subsets of feature categories ($F_s = F_c$)

Rank	Region	Feature Type	Statistical Moment
1	33×33	edge red	var
2	pixel	red	-
3	33×33	range normalised red	skew
4	pixel	range red	-
5	161×161	range red	skew
6	33×33	range blue	skew
7	97×97	range red	mean
8	33×33	range green	skew
9	whole	edge normalised blue	var
10	whole	edge normalised red	var

Table 3: The first ten selected features for a minimalist classifier using white potatoes.

Rank	Region	Feature Type	Statistical Moment
1	65×65	edge red	var
2	65×65	normalised blue	mean
3	33×33	range red	mean
4	33×33	range intensity	skew
5	whole	edge normalised green	var
6	pixel	red	-
7	161×161	edge normalised green	var
8	whole	normalised red	var
9	pixel	range red	-
10	129×129	range blue	skew

Table 4: The first ten selected features for a minimalist classifier using red potatoes.

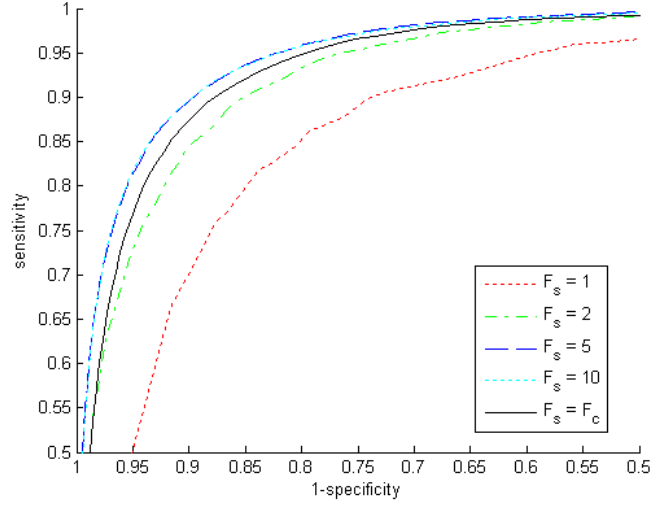


Figure 3: ROC curves for different numbers of selected features F_s tested on white potatoes.

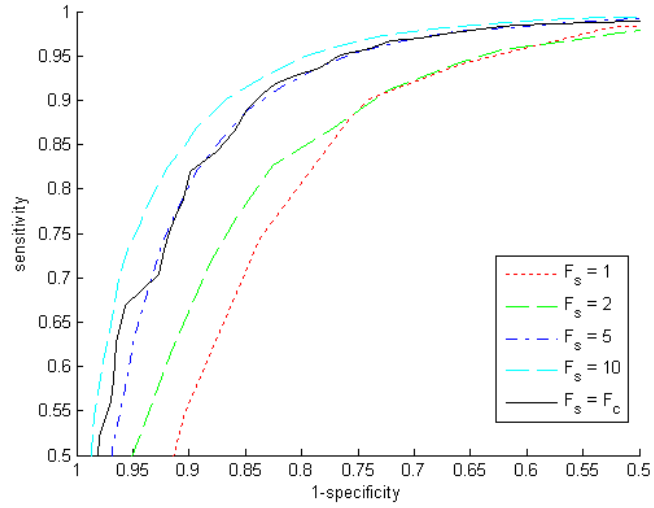


Figure 4: ROC curves for different numbers of selected features F_s tested on red potatoes.

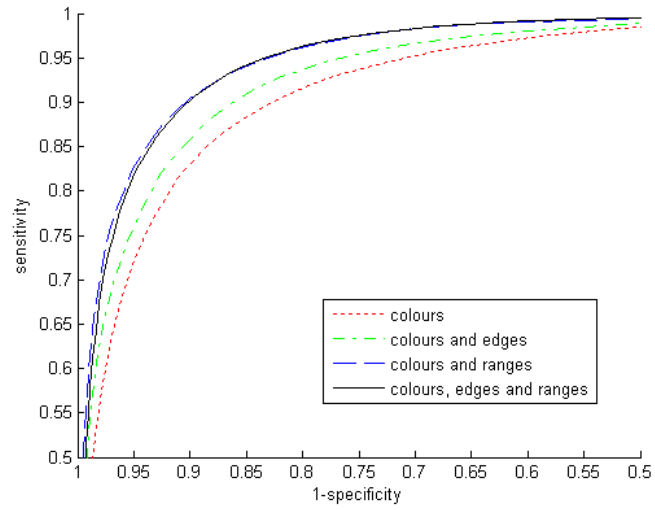


Figure 5: ROC curves for different subsets of feature categories tested on white potatoes.

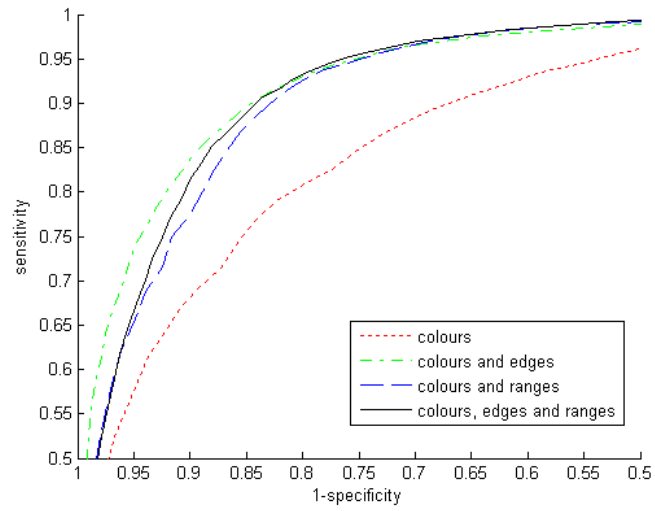


Figure 6: ROC curves for different subsets of feature categories tested on red potatoes.

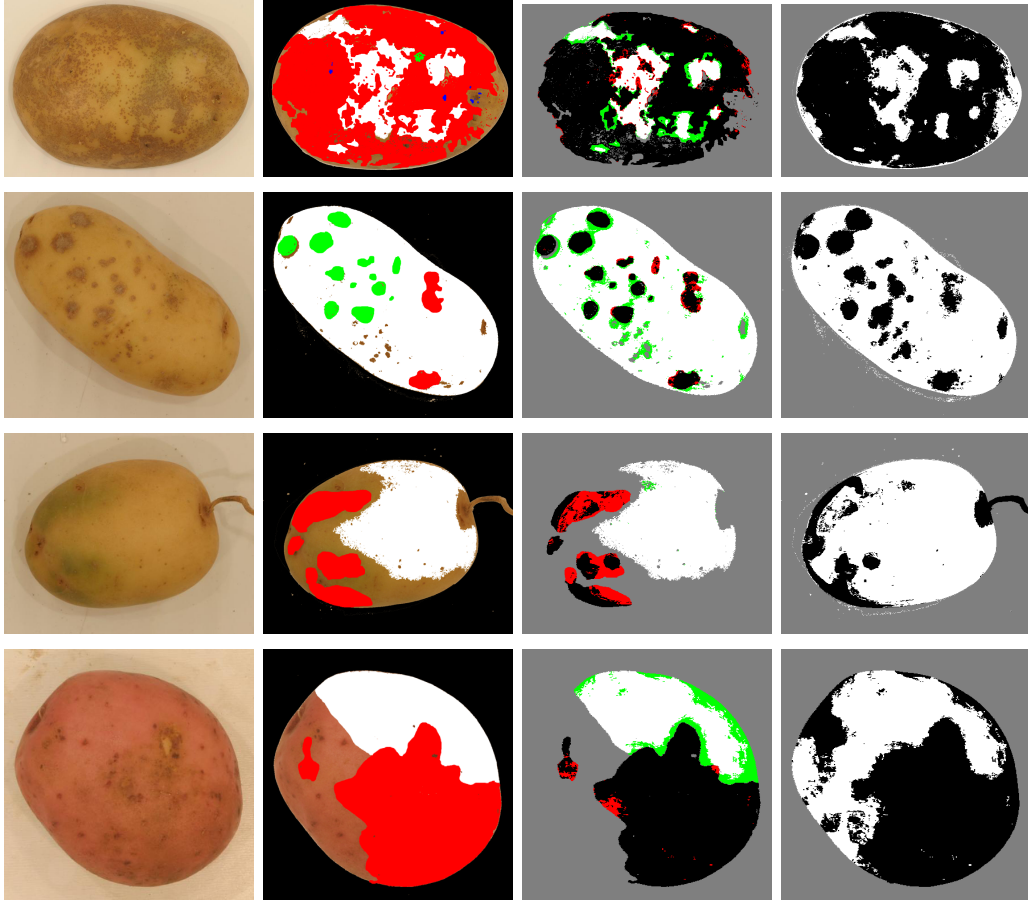


Figure 7: Example images (left to right): first an original photograph, then a ground truthed image, then an error image, showing false positive results in red and false negative results in green, for the detection of blemish. The final image is from the processing of the entire original image without reference to ground truth, blemishes in black and good potato in white. Most disagreements between ground truth and classification results can be seen to be around the edges of blemishes where ground-truthing is less accurate.

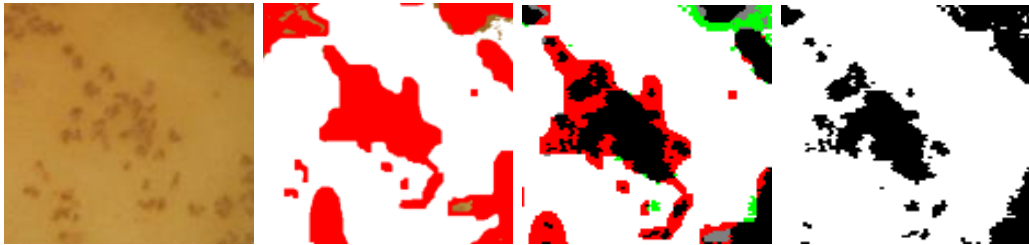


Figure 8: Zoomed-in view of the middle of the first image set in Figure 7 clearly showing that errors are being caused by imprecise markup.