

# Minimalist AdaBoost for blemish identification in potatoes

Michael Barnes, Grzegorz Cielniak, and Tom Duckett

University of Lincoln, Lincoln, UK

**Abstract.** We present a multi-class solution based on minimalist AdaBoost for identifying blemishes present in visual images of potatoes. Using training examples we use Real AdaBoost to first reduce the feature set by selecting five features for each class, then train binary classifiers for each class, classifying each testing example according to the binary classifier with the highest certainty. Against hand-drawn ground truth data we achieve a pixel match of 83% accuracy in white potatoes and 82% in red potatoes. For the task of identifying which blemishes are present in each potato within typical industry defined criteria (10% coverage) we achieve accuracy rates of 93% and 94%, respectively.

## 1 Introduction

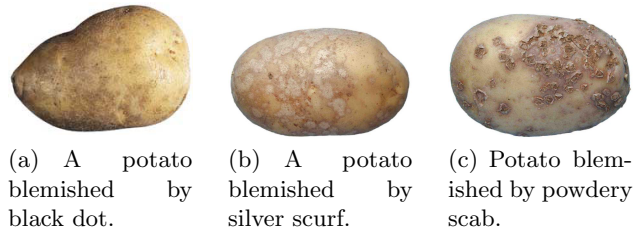
In this paper we present a multi-class minimalist AdaBoost solution to identifying blemish types in images of potatoes. Potatoes are an important part of the food supply in much of the world and the biggest factor in their market price is appearance. A typical human grader can make a quality classification for a small baking potato every 0.45 seconds, thus there is a need to consider both speed and accuracy in an automatic classifier. 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 by type in a set of training images. After training, the system should be able to classify individual pixels by their blemish type, and hence grade whole potatoes according to typical industry criteria, in this case 10% coverage.

### 1.1 Potato blemishes

There are a number of diseases affecting potato tubers that, although superficial and generally of little or no health consequence to humans, strongly and negatively influence consumer choice. These conditions present a variety of differently 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.

These blemishes include black dot, a mostly speckled blemish with a soft texture ranging from brown speckles to a silvery area with black speckles. Silver scurf is a fungal infection which manifests as a silvery area, often with a brown

perimeter and black speckles. Powdery scab and common scab both manifest primarily as areas where the skin splits open along straight lines presenting an often star-like shape. Other forms of blemish include physiological blemishes, e.g. greening and sprouting. See Fig. 1 for examples of these blemish classes.



**Fig. 1.** Examples of the potato blemish types used in these experiments

The identification of blemishes is important for a number of reasons. Supermarkets will usually be more tolerant of some blemishes than others, for instance greening, of the bright “field green” variety, is more of an issue than silver scurf, since the latter can be peeled off. Secondly, there is an interest in the possibility of using such a system for epidemiology purposes, to determine which farms are growing more potatoes with certain blemishes in order to coordinate efforts to provide treatments such as fungicides to the regions most afflicted.

## 1.2 Related work

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), features are extracted that summarise important qualities of the object, then a pattern recognition system is used to categorise the input data. For example, [1] developed methods to distinguish between blemishes in apples and healthy apples with visible stem or calyx. In [2], HSV-colour and texture histograms are used to classify different types of fruit and vegetables.

In the area of machine vision for potatoes, [3] used Fourier harmonics to describe the shapes of potatoes. Custom lighting equipment is used in [4] to project light at a variety of different wavelengths to demonstrate the different reflective properties of specific blemishes at each wavelength. [5] used the HSI colour space to identify green potatoes as well as yellow and green apples. [6] graded potatoes by size and shape. In [7] a system uses green levels to detect green defects (greening and sprouting) in potatoes. In [8] mis-shapen potatoes are detected by comparing the local rate of change in radius of a potato as well as detecting sprouting by comparing the green colour channel with the intensity.

A major limitation of typical systems is that the set of 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

features may well differ from those used to design the original system. In previous work, [9] we used a single class method using AdaBoost [10] to both select good features for a particular pattern recognition task and then to select weak classifiers based on the features chosen. In this paper we extend the approach to identify the different types of blemish described in Section 1.1.

## 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 at a distance of 60cm from the camera objective to the base on which the subjects were placed, with a focal length of 70mm and an aperture setting of F22. The resolution of the images was  $1536 \times 1024$  pixels. 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.

### 2.2 Ground truth

There were two sets of data collected for white and red potatoes respectively, including potatoes affected by different blemishes. For our classifier we combined greening and sprouting into one class as well as combining powdery and common scab into a single class, since the latter pair are very difficult to tell apart [11]. We are left with five classes; black dot, silver scurf, scab, green and unblemished. The white potato data set contains 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 contains 48 images including between 10 and 15 examples of each blemish class.

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 types and to non-blemish. It is not necessary to label all pixels in an image: some areas of high uncertainty or ambiguity are left unmarked, and these pixels are ignored during training of the classifier. Background pixels are also omitted from the subsequent calculations. An example of training and ground truth images can be seen in Figure 2.

### 2.3 Feature extraction

The first step of the procedure is to extract image features that should indicate the type of blemish present in a potato image. The features include statistical summaries of the whole potato and square regions of interest centred on each pixel as well as the data of the pixel itself. The statistics used for each feature type and region were the mean, variance, skew, maximum and minimum values in the area. The proposed system uses the RGB colour space - the original colour format of the camera output. The system software was implemented in MATLAB.

The square regions used in our experiments were of size  $33 \times 33$ ,  $65 \times 65$ ,  $97 \times 97$ ,  $129 \times 129$  and  $161 \times 161$ , giving 5 regions in total, as well as the pixel.

Our system uses seven colour channels; raw RGB, normalised RGB and the intensity channel. From these channels we consider the following image properties:

*Colour channels and intensity:* 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.

*Edge Gradient:* 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  $3 \times 3$  kernel size.

*Edge Length:* An extension of the Edge Gradient feature, the edge length is determined by using a thresholded sobel edge detector. Pixels are first labelled as edge or non-edge, then edge pixels are used to form connected components. Each pixel within a component is then given a value equal to the number of pixels which form that component. Non-edge pixels are given a value of zero. Larger edge components tend to be found around edges in particular.

*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 image. The range filter was run on the same seven channels with a  $5 \times 5$  neighbourhood.

In summary there are 7 colour channels  $\times$  4 feature types  $\times$  5 statistics making 140 features for each region and  $7 \times 4 = 28$  features for the pixel itself. All these features are used as the candidate feature set. Since there are 5 different regions, this gives  $140 \times 5 = 700$  features which, with additional 28 features for the pixel itself, gives us  $F_c = 728$  candidate features in total. These features were used as the training input to our classifier.

### 2.4 AdaBoost

The AdaBoost algorithm [10] is used to build a classifier, which combines results from so-called “weak” classifiers (each a decision stump using one of the candidate features) into one “strong” classifier that performs better than any of the

weak classifiers alone. 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 has a weighted vote in the classification of a given example.

Real AdaBoost [12] 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. Using MATLAB we have used the Real AdaBoost implementation within the GML AdaBoost Toolbox [13] for these experiments.

Minimalist AdaBoost [9] is a method of limiting the number of unique features used by Real AdaBoost while still allowing free selection of a larger number of weak classifiers from the whole set of candidate features. This is achieved by training on the full set of features until AdaBoost has selected weak classifiers using a chosen number of unique features, then continuing using only those selected features up to the chosen maximum number of weak classifiers. This can improve the speed of a system which has to extract features in real time, in tasks such as produce quality control.

To extend Minimalist AdaBoost to multi-class recognition problems, it was trained for each class with 10,000 positive examples and 10,000 negative examples, the negative examples being split equally between all other classes. First the classifier was used to select features for each class, then each binary classifier was trained on all selected features, allowing 40 weak classifiers. When classifying test data, the five different binary classifiers were used and their output certainties compared. The pixel was then classified according to the classifier which returned the highest certainty of a positive result for its class, or the lowest certainty of a negative result (a winner-takes-all approach).

## 3 Results and discussion

### 3.1 Training and testing

To begin, the images were split into training and testing sets. These were selected to ensure that a representative number of examples of every blemish was present in both sets and also that those potatoes which had been photographed twice, front and back, did not appear in both sets. From this data, 10,000 examples of each of the 5 classes were extracted, spread out within all training images.

Using minimalist AdaBoost with a winner-takes-all approach, each example (an individual pixel in an image) was classified with the binary classifiers and assigned to the class with the highest certainty.

### 3.2 Selected features

In many cases the top features selected for each class had an intuitive explanation for why they were chosen.

In white potatoes, the top feature selected for scabs was the variance of the intensity range filter over the smallest area, representing how much the texture changes in the immediate vicinity of this blemish. The other four were all related to edge detectors. The top feature selected for green blemishes was the minimum value of the green colour channel in the 65x65 region. Two interesting parallels occur between silver scurf and black dot; the presence of the maximum red for the smallest area and the presence of the variance of a normalised colour followed immediately by the maximum value of the same, non-normalised colour, which may be representative of their similarity as blemishes. For the good potato class, the top five selected features all relate to either the red or normalised red colour channels.

In red potatoes, the top feature selected for scabs is the maximum level of the red channel over the smallest area, possibly because there is more of a colour difference between scab and red skin than white, although three of the next four features are edge related. Green begins with two features relating to the intensity channel, which may be because field greening of red potatoes is closer to black than green. Features selected for black dot have some correlation with those on white potatoes, notable the first feature relates to the blue channel in the largest region, compared to the normalised blue channel in the largest region for white potatoes. Also the maximum red value in the smallest region is used again. For silver scurf the variance and skew of the edge detector for the normalised blue colour channel over the largest region are selected as well as two features for the green colour channel. For good potato skin, the most obvious difference from white potatoes is that only one of the five selected features relates to the red channel. Edge length is selected twice, once for green and once for normalised green.

### 3.3 Classification performance

(a) White potatoes						(b) Red potatoes					
	Bd	Ss	Sc	Gr	Gd		Bd	Ss	Sc	Gr	Gd
Black dot	<b>80%</b>	9%	1%	0%	9%	Black dot	<b>63%</b>	5%	2%	5%	25%
Silver scurf	8%	<b>82%</b>	5%	0%	4%	Silver scurf	6%	<b>72%</b>	3%	3%	17%
Scab	4%	7%	<b>87%</b>	0%	2%	Scab	1%	1%	<b>97%</b>	1%	2%
Green	1%	0%	0%	<b>96%</b>	3%	Green	1%	2%	1%	<b>92%</b>	4%
Good	9%	3%	1%	2%	<b>85%</b>	Good	8%	5%	3%	1%	<b>83%</b>

**Table 1.** Pixel-wise confusion matrices, showing the ground truth classes in rows and the classifier output classes in columns.

Table 1 shows a comparison between the classifier output and ground truth data. The main disagreements are fairly consistent. In white potatoes the biggest confusion is between silver scurf and black dot, which are known to look very similar at times, even confusing human experts. Figure 2 shows a very good example of the other common disagreement between the classifier and ground truth, which is the exact extent of a speckled blemish, in this case black dot.

It is likely that potential users of this system will be more interested in the performance of this classifier on a potato-wise rather than pixel-wise basis, since they will either be grading whole potatoes or monitoring, for instance, the number of potatoes afflicted by specific blemishes in certain fields, with reference either to location or to experiments such as pesticide trials. Potato-wise performance is shown in Table 2, with each class being given a positive result if it is detected in more than 10% of the pixel-wise results for that image, so 10% Scab coverage was either wrongly detected or missed in a total of 5 potatoes. The exact coverage that is considered as a major blemish varies by customer, typically in the range of 1% - 10% of surface area. Using 10% as our threshold for individual images we achieved a potato-wise classification accuracy of 93% and 94% for white and red potatoes, respectively.

	White Potatoes	Red Potatoes
Black dot	90%	92%
Silver Scurf	88%	90%
Scab	95%	100%
Green	98%	96%
Good potato	95%	96%
Overall	93%	95%
Blemish	93%	94%

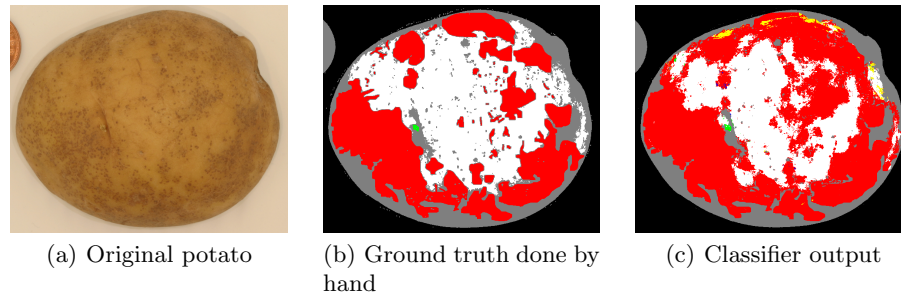
**Table 2.** Potato-wise accuracy for detecting significant (10%) coverage per class.

## 4 Conclusions and further work

We have presented a trainable classifier using a multi-class adaptation of minimalist AdaBoost which detects blemishes in potatoes with 93-94% accuracy.

The biggest concern we identified was the disagreement as to the exact locations of blemishes between the hand-drawn ground truth and the pixel-wise classification. Some of these disagreements are due to the inaccuracy of hand-drawing as a means of ground truthing. It is possible that this could be improved with an assisted ground truthing method which might include some automatic classification during the markup process.

So far the presented method does not consider the 3d shape of potatoes, whereby a blemish seems smaller toward the edge of a potato since it is at a steeper angle to the camera. This challenge could be approached in several ways, either by increasing the weight given to blemishes near the potato's edge, using a mathematical model or a 3d sensor to decide the weighting, or by taking a number of images and averaging the detected blemish areas.



**Fig. 2.** A potato classified by the system, showing the similarity between classifier output and ground truth, with differences in the exact boundaries of blemished areas. Red = black dot, green = green, white = unblemished, grey+black = untested

## References

- [1] Unay, D., Gosselin, B.: Stem and calyx recognition on 'jonagold' apples by pattern recognition. *Journal of Food Engineering* **78** (2006) 597 – 605
- [2] Bolle, R., Connell, J., Haas, N., Mohan, R., Taubin, G.: Veggievision: A produce recognition system. In: *Proc. 3rd IEEE Workshop on Applications of Computer Vision (WACV '96)*, Washington, DC, USA, IEEE Computer Society (1996) 244
- [3] Tao, Y., Morrow, C.T., Heinemann, P.H.: Fourier-based separation technique for shape grading of potatoes using machine vision. *Trans. American Society of Agricultural Engineers* **38**(3) (1995) 949–957
- [4] Muir, A.J., Ross, D.W., Dewar, C.J., Kennedy, D.: Defect and disease detection in potato tubers. In: *Proceedings of SPIE - The International Society for Optical Engineering*. Volume 3543. (1999) 199–207
- [5] Tao, Y., Heinemann, P.H., Varghese, Z.: Machine vision for color inspection of potatoes and apples. *Trans. American Society of Agricultural Engineers* **38**(5) (1995) 1555–1561
- [6] Heinemann, P.H., Pathare, N.P., Morrow, C.T.: An automated inspection station for machine-cision grading of potatoes. *Machine Vision and Applications* **9**(1) (1996) 14–19
- [7] 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) (1998) 423–433
- [8] Guannan, Z., Yuzhi, T., Junxiong, Z., Wei, L.: Automatic detecting and grading method of potatoes with computer vision. *Nongye Jixie Xuebao / Trans. Chinese Society of Agricultural Machinery* **40**(4) (2009) 166–168+1
- [9] 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) (2010) 339 – 346
- [10] Freund, Y., Schapire, R.: A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence* **14** (5) (September 1999) 771 – 780
- [11] De Haan, E.G., van den Bovenkamp, G.W.: Improved diagnosis of powdery scab (*spongopora subterranea* f.sp. *subterranea*) symptoms on potato tubers (*solanum tuberosum* l.). *Potato Research* **48**(1-2) (2005) 1–14
- [12] Schapire, R., Singer, Y.: Improved boosting algorithms using confidence-rated predictions. In: *Proc. Conf. on Computational Learning Theory*. (1998) 80–91
- [13] Vezhnevets, A.: *GML AdaBoost MATLAB Toolbox* (2006)