Image based estimation of oat panicle development using local patterns

Boyle, Roger David; Corke, Fiona; Howarth, Catherine Jane

Published in:
Functional Plant Biology

DOI:
10.1071/FP14056

Publication date:
2014

Citation for published version (APA):
Boyle, R. D., Corke, F., & Howarth, C. J. (2014). Image based estimation of oat panicle development using local patterns. *Functional Plant Biology*, *42*(5), 433-443. https://doi.org/10.1071/FP14056

Image based estimation of oat panicle development using local patterns

Roger Boyle, Fiona Corke and Catherine Howarth
National Plant Phenomics Centre, IBERS
University of Aberystwyth
{*rob21,fic5,cnh*}*@aber.ac.uk*

May 24, 2014

Abstract

Flowering time varies between and within species, profoundly influencing reproductive fitness in wild plants and productivity in crop plants. The time of flowering, therefore, is an important statistic that is regularly collected as part of breeding programs and phenotyping experiments to facilitate comparison of genotypes and treatments. Its automatic detection would be highly desirable.

We present significant progress on an approach to this problem in oats, an underdeveloped cereal crop of increasing importance. Making use of the many thousands of images of oat plants we have available, spanning different genotypes and treatments, we observe that during flowering, panicles (the flowering structures) betray particular intensity patterns that give an identifiable texture that is distinctive and discriminatory with respect to the main plant body and can be used to determine the time of flowering. This texture can be located by a filter, trained as a form of a 'Local Pattern'. This training phase identifies the best parameters of such a filter, which usefully discovers the scale of the panicle spikelets.

Results are presented that demonstrate the success of the filter. We proceed to suggest and evaluate an approach to using it as a Growth Stage detector. Preliminary results show very good correspondence with hand-measured ground truth, and are amenable to improvement in a number of ways. Future work will build on this initial success and will go on to locate fully mature panicles, which have a different appearance, and assess whether this approach can be extended to a broader range of plants.

1 Introduction

Cereal (and other plant) development goes through a number of well defined stages that are used globally to perform monitoring and comparison (BBC 2001, Zadoks et al. 1974). Thus, for example, the widely used Zadoks scale defines Growth Stage (GS) 0 as dry seed, GS 20 appearance of the first tiller, GS 50 appearance of first spikelets of the panicle, and so on. Catalogues – or atlases – of these stages are accompanied by representative drawings or images. This atlas-based approach to documenting development is familiar in many domains, in particular medical imaging where 'expert' judgements are recorded to assist classification of individual cases. These atlases often prove useful tools when automatic imaging techniques are later introduced to the domain (e-A 2014). Just one example, wrist radiography, is illustrated in (Gertych et al. 2007, Tanner & Whitehouse 1975).

Worldwide, there is increasing interest in applying imaging technologies to plant phenotyping (Furbank & Tester 2011), and a growing number of installations able to perform large scale phenotyping experiments – (APP 2014, JPP 2014, NPP 2014) are just some examples. Usually, these are based on automated greenhouses that can administer pre-programmed treatments to a number of plants, of which they likewise make regular automated measurements. These installations permit large scale experiments to be conducted over time within complex regimes, with minimal staff input. Much can be gained from the simplest of monitoring such as a photograph, but a variety of other image modalities (UV, IR, NIR, structured light), and root analysis, are also available. Measurements of benefit to biologists can then fall into a number of categories:

1. Replication/mimicry of 'simple' measurements performed manually. These include plant height and projected area (which can be used to approximate mass).

2. Replication/mimicry of less easily accessible measurement, for example, atlas growth stages.

3. Measurements that may be of benefit that have not been made systematically in the past.

The science of computer vision also continues to exhibit significant progress. In particular, many algorithms are now in every day use that operate on a 'train then classify' approach, where some form of automatic detector is built from knowledge of a (perhaps very large) number of training cases (Šonka et al. 2014). Such detectors have in recent years become increasingly sophisticated. High throughput phenotyping installations represent very fertile territory for many such algorithms, and coupled with good quality domain atlases, we might hope to build automatic systems that replicate significant parts of the work currently done very labour-intensively. More interestingly, we might seek to develop measurements accessible to computer extraction that would be difficult or costly if collected manually. Such activity has been growing in popularity in recent years, for example (Campillo et al. 2010, Hartmann et al. 2011, Reis et al. 2012, Sirault et al. 2013, Song et al. 2014).

In this paper we present work in progress on one such example: flowering in oats. This is an important property for commercial reasons, since it impacts on adapting varieties for particular agronomic purposes. However, the spikelets of the panicle are small and easily

obscured by the body of the plant making their reliable detection in images a challenge – we are unaware of attempts to solve this problem using computer vision, although studies of images of pre-isolated panicles have been conducted (Al-Tam et al. 2013, Huang et al. 2013). We find that areas of the image in which spikelets are emergent betray textural properties that are amenable to image-based extraction, and we show how this property can be used to estimate critical growth stages – we consider this work to be in the second category enumerated above. This work can be used as the basis of subsequent filters which will identify later growth stages of interest, such as full flowering. It may also be possible to generalise the approach to related cereals such as rice and millet.

(a) An oat plant – the plant is approximately at GS 60 (Zadoks et al. 1974).

(b) Close up of a panicle.

Figure 1: Images captured at the UK National Plant Phenomics Centre (see Figure 3); image quality is distinctly sub-optimal as the system was still in commissioning.

2 Background

2.1 Flowering time in oats

The oat plant in development goes through a number of well understood and documented phases (BBC 2001, Zadoks et al. 1974); one of particular interest is progress in flowering: phase GS 50 represents the appearance of the first spikelet (of the primary tiller), GS 60 would be full heading but not flowering, and GS 70 full flowering. This is straightforwardly observable in visual inspection of growing plants, within tolerable error limits. It is possible and normal for a plant to occupy more than one stage at any given time as successive tillers develop. Figures 1 and 8 gives some illustration.

Flowering is a major developmental transition in the life history of plants and has a major impact on grain yield in cereals such as oats. Control of flowering time is essential to maximise reproductive success, enabling completion of seed development in favourable environmental conditions. This adaptive effect has been exploited in agriculture to ensure that plants flower synchronously and at the optimal time to maximise seed yields (F & G 2012). Optimum floral initiation and development ensures the maximum use of resources available throughout the growing season, and minimises the exposure of sensitive floral tissue to biotic and abiotic stress (Worland 1996). In temperate environments with a long growing season, late flowering ensures a long vegetative phase for maximal resource capture leading to high grain yield

potential. Early flowering is important for environments where the effective growing season is short either due to extremes of temperature or water availability or where multiple cropping seasons within a year are possible (Locatelli et al. 2006). Considerable genetic variation exists for the control of flowering time and plant breeders continue to select for optimal flowering time to maximise yield for specific environments. The ability to quickly and accurately measure flowering time is important to characterise the genetic variation that exists for this trait and to determine the influence of the environment on the its regulation. This information can be combined with genetic analysis (Tinker et al. 2009) to identify regions of the genome controlling flowering time (Holland et al. 2002, Locatelli et al. 2006, Locatelli et al. 2008, Nava et al. 2012) Knowledge of both flowering time and the genes regulating it can then be used to precisely manipulate this trait within a plant breeding programme.

2.2 Classification from Binary and Ternary patterns

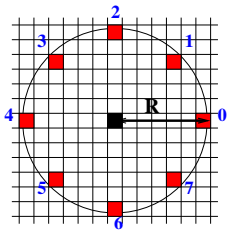
Texture characterisation is a well established branch of image processing and vision. It is very common for texture rather than shape or intensity to be the most accessible feature of certain image regions; it may be clear that this is frequently the case with images of plants. The range of well-established techniques is very wide (see, for example, (Šonka et al. 2014)). Recently, *Local Binary Patterns* (LBP) have found especial favour in this area (Ojala et al. 2002), demonstrating tolerance to a number of commonly encountered imaging problems while being appealingly simple. Since we determine (see Section 3) that panicles in development exhibit characteristic binary (in fact, red/green) patterns, we choose to experiment with these.

A LBP is derived at every pixel of an image; centred at the pixel, a circle quantised into q pixels is drawn at some radius R, and the pixels so defined thresholded by the intensity of the central pixel – thus a q-bit pattern is defined at each pixel. Figure 2a illustrates this for $q = 8, R = 6$ – a bit is determined at each red pixel according to whether it exceeds the central one or not, and a q-bit number is assembled in the order indicated, and associated with the black pixel. This representation very efficiently captures local contrast patterns.

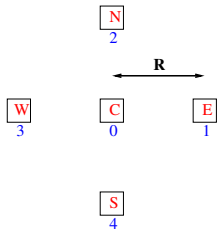
Now square windows of dimension W are considered, and a histogram of the responses within constructed. This histogram has W^2 responses in the range $[0, \ldots, 2^q - 1]$, and the vector it represents can be a very powerful texture descriptor; other areas of the image generating similar histograms will have similar texture. Such a histogram can be calculated at each pixel position by sliding the window through the image; this could become very costly for large images and it is often sufficient to tile the image with $W \times W$ windows, or perhaps compromise by sliding the window by $\frac{W}{2}$ pixels rather than 1.

LBPs lend themselves to very efficient implementation, and tuning to exploit the occurrence of areas of uniform intensity. It is further straightforward to adapt the idea to cope with reflective symmetries or to impose rotational invariance.

A companion approach would be to define a binarisation of the original image (at simplest, by thresholding), and derive per-pixel responses in a similar manner. Hereunder we describe a reduction of the image to three intensities, and a consequent ternary pattern to characterise local texture.



(a) Local Binary Patterns – a quantised circle around a central pixel is considered. Quantisation and radius are filter parameters.



(b) The local pixel pattern used in this application – a central pixel has four neighbours defined at distance R.

Figure 2: Texture description by local patterns.

2.3 Imaging environment

(a) One of the robotic greenhouses.

(b) Plants (here, maize) entering the imaging chambers.

Figure 3: The UK NPPC.

The UK National Phenomics Centre (NPPC) has recently been established at the university of Aberystwyth and exists to conduct large scale phenomics experiments. The full facility is described elsewhere (NPP 2014); it affords a variety of imaging modalities and opportunities for controlled environments and treatments. Here, it is sufficient to appreciate that up to 850 plants can be imaged daily under specified conditions. Imaging can include rotated and birds-eye view pictures of each plant.

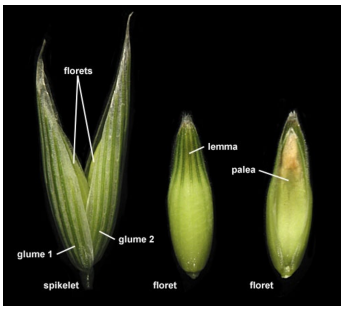
3 Detecting panicle emergence in images of oats

An experiment was conducted over a ten week period on a mapping population derived from the Buffalo and Tardis varieties. 3 individuals of each line and some parental controls provided 282 different plants, which generated a total of over 9700 RGB images[1]: The plants

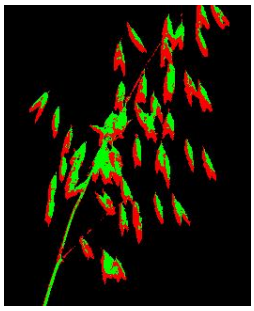
[1]Aspects of the system were being commissioned during this experiment. In normal circumstances many more images of higher quality and greater consistency, and other image modalities, would have been generated during this period.

develop from a single tiller through flowering to fully senesced, and we are interested in panicle development, particularly on the primary tiller.

Individual spikelets are small features, many of which are often occluded or obscured by other plant matter, meaning that attempts to locate or count them explicitly will be very challenging. On the other hand, to recognise the onset of flowering it would be sufficient to recognise an image area in which spikelets probably lie. Close inspection reveals that the rim of a visible spikelet has a narrow yellow band even during very early stages of development, implying that the red channel will be dominant in that band in an RGB representation. This is almost certainly due to the material of the glume being very different to the lemma it encloses. Figure 4 illustrates this effect.



(a) Structure and terminology of oat spikelets.



(b) Fig. 1b partitioned as background, red-dominant, green-dominant.

(c) Later in development, spikelets become predominantly yellow.

Figure 4: Panicles and spikelets (Image 4a: Anna Gardner, with permission from Iowa State University).

Segmentation of the plant tissue and suppression of the (blue) support frame is straightforward in the controlled imaging environment and is performed routinely on all images we collect. Thereafter we could simply seek dominance in the Red or Green channels; in fact, our approach is slightly more sophisticated. A (large) representative subset of plant matter pixels over the entire time series of images is considered: The RGB triples are 24 bit, but this very large data space is subjected to K-means clustering (Šonka et al. 2014) to facilitate various procedures independent of this application (such as detection of senescence). Empirically we have discovered that for all image sequences we have inspected, 25 clusters provide an adequate representation (that is, significantly increasing the number of clusters reduces information loss in the quantisation only marginally, while using 20 or fewer begins to increase this loss appreciably). No qualitative difference in results has been observed by changing the K-means initialisation, and this was done randomly.

We quantise the plant pixels using this approach, and then partition the cluster exemplars as either red or green dominant (it is no surprise that none are blue-dominant), which allows us to partition the image as black (for background), red and green – this approach encourages the small spikelet regions to emerge robustly and to reduce noise effects. The fine scale pattern evident in Figure 4b is rarely if ever evident in other areas of the image, where green and red regions are usually larger, and rarely in the geometric arrangement seen in the spikelets.

This leads us to suggest a texture detector that would highlight such regions: we experiment

with a very simple form illustrated in Figure 2b, considering a central pixel and its four orthogonal (E/N/W/S) neighbours at a distance of R. Arbitrarily labelling black as 0, green as 1 and red as 2, the per pixel texture is defined as

$$C + 3E + 9N + 27W + 81S$$

giving a value in the range $[0, 243]$. Then a histogram computed over some window would deliver a 243-dimensional feature. We can ameliorate this size by

- Neglecting responses of constant (all background, red or green) response, since the panicle is characterised mainly by the proximity of variable response. This reduces dimensionality to 240.
- Perhaps imposing a vertical symmetry constraint. The spikelets hang to left or right but we do not mind which; in each case they have a green apex and a red lower rim. If we consider the EW pixels of Figure 2b, and let them be *un*ordered (so, for example, Black/Red and Red/Black are taken as the same), the detector will become slightly less specific, but the dimensionality reduces to 159. (Since there are many pattern instances in which E and W pixels are the same this does not halve the dimensionality).

Accordingly, we have experimented with 240 and 159 dimensional detectors.

The approach is to define a square window of size W and radius R, and determines the frequency histogram in panicle areas of ground-truthed images. Unclassified images are then presented, marked as $(background, red, green)$ and a histogram computed at each pixel, which is then compared with the learned model. Histograms are compared using the Hellinger distance (Hellinger 1909): Choice of this metric was somewhat arbitrary, but it seems improbable that any alternative would significantly affect results. If $\mathbf{h}^1$ and $\mathbf{h}^2$ are two normalised histograms, this distance is

$$H(\mathbf{h}^1, \mathbf{h}^2) = \sqrt{1 - BC(\mathbf{h}^1, \mathbf{h}^2)}$$

where BC is the Bhattacharyya coefficient

$$BC(\mathbf{h}^1, \mathbf{h}^2) = \sum_i \sqrt{h_i^1 \cdot h_i^2}$$

where i counts through the components of the vectors $h_i^{\{1,2\}}$. This distance is then in the interval $[0, 1]$, and lower numbers will be indicative of spikelet presence.

In summary, we develop a panicle indicator by performing:

Algorithm: Determine panicle-like response in an image

1. Choose R and W.
2. For a small number of images, outline areas that contain panicles, or parts thereof. (This does not need to be done with great precision, making it a quick operation).
3. Convert the images to be 3-level, black, red, green. Select pixels randomly from the panicle areas and compute a 159-wide histogram with the given R, W.

4. Total the histograms and normalise, providing a 159-wide probability distribution $\mathbf{p}$ that describes panicle areas.

5. For an unexamined image, compute a histogram at each pixel and normalise it. Record the Bhattacharyya distance between the observation and $\mathbf{p}$ as the measure of panicle evidence at that pixel.

The choice of R and W at step 1 should be guided by the performance of the resulting filter – a systematic optimisation of this choice is discussed in section 4.1.

Step 5 here could be time-consuming and could be performed in a subset of pixels defined by a tesselated tiling, or some overlapping tiling of the image. In our experiments we have used a $W \times W$ window slid by quanta of $\frac{W}{2}$. In Section 4.2 we describe how this response filter may be used in a series of images to estimate day of onset of flowering.

4 Results

4.1 Choice of filter parameters

Five images with panicles at GS 50-60 were marked by hand, giving masks indicating positive/negative regions. Training was performed on pixels selected randomly from these positive masks, and the resulting histogram frequencies were then tested on marked image areas not used in training. For the sake of efficiency in testing, histograms were not computed at every pixel, but rather in square windows of size half that over which the histogram was computed.

Performance was measured by computing a Precision-Recall (PR) curve over the test-set, and for a given R, W pair recording the area under the curve. PR is often used in preference to ROC curves when there is a significant (order of magnitude) difference between the number of positives and negatives in a dataset, as here. In the normal manner, we define for a given threshold of the filter output

$$TP = \text{True Positives} \ , \quad FP = \text{False Positives}$$

and TN, FN similarly for negatives, then

$$P = \frac{TP}{TP + FP} \ , \quad R = \frac{TP}{TP + FN}$$

P and R are then plotted against one another - for a perfect classifier, the area under this curve is 1, while a random classifier will give area 0.5.

In all such experiments, the 159-dimensional feature performed very slightly but consistently better than the 240-dimensional one, and we settle on using it accordingly. There was no foreknowledge of 'good' values for R and W, and ranges of $R = 1, \ldots 12, W = 5, 10, 15 \ldots 120$ were selected arbitrarily.

Figure 5 illustrates performance: increasing brightness corresponds to better performance. As R increases beyond 2 overall performance deteriorates, with a best response at $R =$

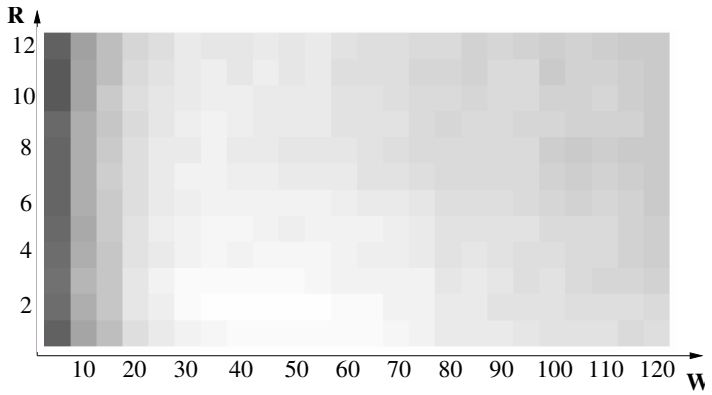


Figure 5: Areas under PR curves for various values of R and window sizes W, where light is high and dark is low (scaled for display). The best (brightest) response is seen at $R = 2, W = 45$.

$2, W = 45$ where the area is 0.74 (the worst response is at $R = 9, W = 5$). These figures are reasonable at the scale of image we have collected: Figure 6 shows two 45×45 windows of spikelets – it is clear that the best-performing detector has discovered the approximate scale of the feature of interest. At the right of the figure, a 10×10 window of a detail from one of the spikelets illustrates that $R = 2$, implying a 5×5 window around the central pixel, will capture the local red/green/black variation that characterises a panicle.

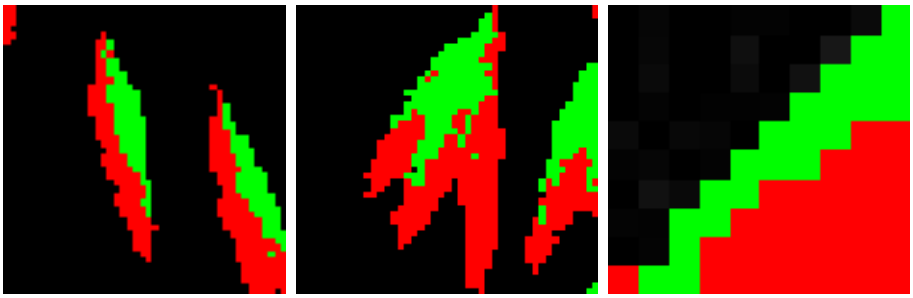


Figure 6: Close-ups of parts of a panicle: left and middle, 45×45 windows, right, a 10×10 window.

As further confirmation that the detector is functioning as it should, Table 1 shows the ternary patterns that dominate the trained histogram. Ten of the 159 dimensions provide over 50% of the response, and the patterns capture the co-occurrence of background, red and green as expected. In particular, the fifth, seventh, eight and ninth patterns illustrate that boundaries at the 'top' of regions are predominantly green, and at the 'bottom' red.

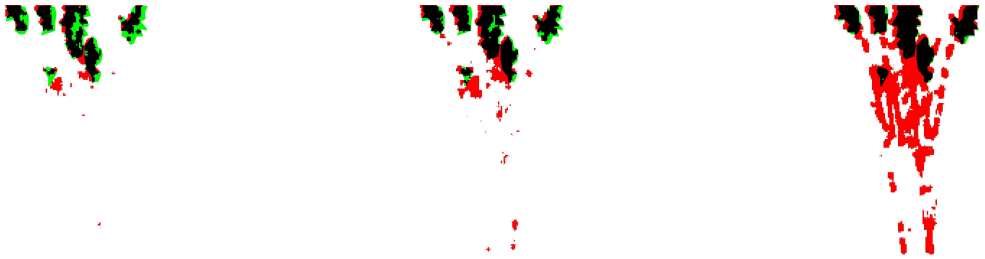
4.2 Derivation of GS estimator

We have used the output of the imaging system to exercise this ($R = 2, W = 45$) ternary filter on image sequences of 82 developing plants. The filter provides strength of belief in the existence of a panicle – Figure 7 illustrates thresholding selection of this measure, colour coding the True/False Positives and Negatives. It is clear that the central image provides

	0.105		0.039
	0.082		0.038
	0.069		0.034
	0.059		0.034
	0.041		0.029

Table 1: The ternary patterns that the filter predominantly seeks, with their histogram frequencies — these 10 (of 159) contribute over 50% of the observation in 'good' areas. (Black is background, green and red denote dominance of plant pixels in the green and red channels Note that the EW pixels are considered to be *un*ordered.)

the best indicator. This image also highlights the opportunity to restrict inspection to the top of the plant, having an immediate beneficial effect on the False Positive rate.



(a) Threshold 0.25 (b) Threshold 0.3 (c) Threshold 0.5

Figure 7: Thresholding the panicle filter: White pixels are True Negatives, black are True Positives, green are False Negatives and red are False Positives.

Figure 8 shows the development of a single plant over a 90 day period[2]. This confirms that we can probably neglect any response that is, say, below half plant height. More interestingly, it is possible to verify that all panicles produce a response of some kind, but the very strongest (red) responses are evidenced well into development.

It remains to indicate how this detector may automate the *estimated* measurement of GS. Figure 9a shows for the plant of Figure 8 the number of pixels within certain Hellinger thresholds (the colours corresponding to that Figure).

The precise patterns of these response curves are not easy to model – as the first tiller begins to senesce and the second and subsequent panicles begin to develop, we will expect a very noisy superposition of peaks and troughs. Nevertheless, the early phases of each response curve may be expected to be approximately zero, prior to the panicle emergence, followed by a sharp climb corresponding to the primary tiller's panicle which, while probably

[2]The day of sowing precedes day 1 of observation by some time – plants are not introduced to the imaging system before they are visible, at GS 20 or later

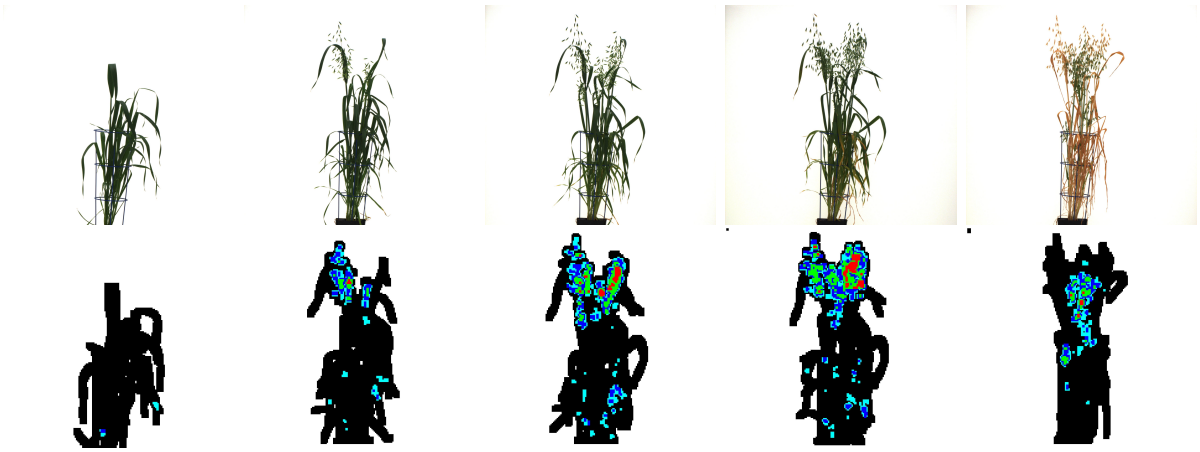


Figure 8: Development of a particular plant seen through the panicle detection filter. This plant was imaged over 90 days, day 1 being April 20^{th} 2013, which was 31 days after sowing: these images were recorded on days 36, 51, 56, 67, 85. Best estimate GS for the primary tiller are 39, 59, 65, 73, 85. Filter response (weight of evidence) is coded as black (low), light blue ($H < 0.4$), dark blue ($H < 0.35$), green ($H < 0.3$), red (high) ($H < 0.25$).

quasi-sigmoidal, may be approximated as linear. Piecewise linear approximation to noisy observations of sigmoidal responses has seen good success in other domains (Kubassova et al. 2007), and accordingly we seek a good fit to the early part of these curves by functions of the form

$$f(x) = \begin{cases} 0, & x \leq x_0 \\ m(x - x_0), & x_0 < x \leq x_1 \end{cases} \tag{1}$$

where the parameters x_0, x_1 give the start and end of the 'linear' upward segment, and $m > 0$ is the gradient. It is straightforward to minimise, in the least squares sense, over these parameters for a given signal. Supposing an observed signal is $y = (y_1, y_2, \ldots, y_T)$, and f is defined by equation 1, then we perform:

Algorithm: Determine best fit piecewise linear approximation

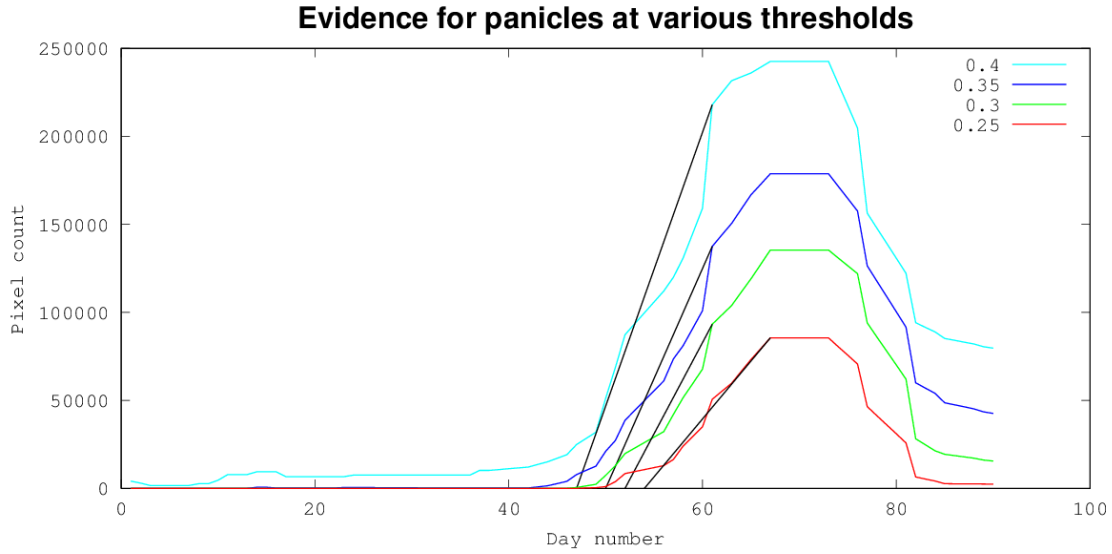
1. Set $x_0 = 1$.
2. For x_1 in the range $[x_0 + 1, T]$, determine

$$m_{x_0} = \max_{x_1} \frac{y_{x_1}}{x_1 - x_0}$$

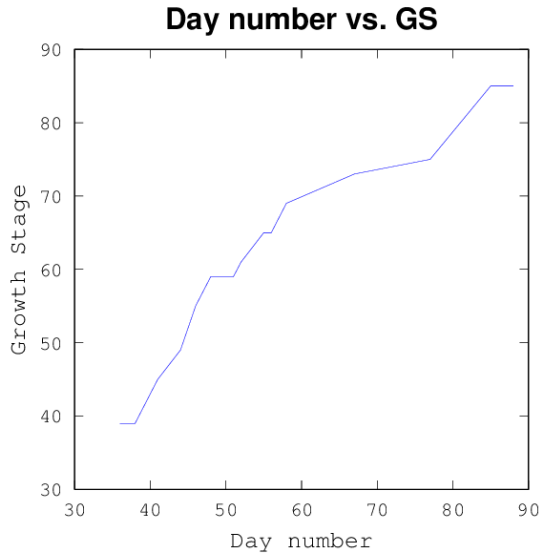
   the linear approximation following x_0 of maximal slope. For x_1 giving this maximal $m = m_{x_0}$, set $E(x_0)$ as the MSE between y and f so defined in the interval $[1, x_1]$.
3. While $x_0 < T - 1$, set $x_0 = x_0 + 1$ and go to 2.
4. Determine

$$\hat{x}_0 = \arg\max_{x_0} E(x_0)$$

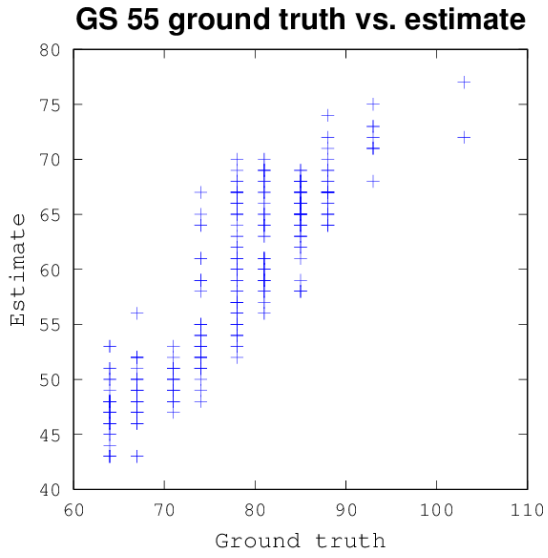
   as the best performing offset. Use this value, and the associated x_1 determined in step 2.



(a) Number of pixels giving certain strengths of response over time measured in experiment days – colours correspond to those of Figure 8. Black lines give LSQ best linear approximation to initial increase.



(b) For the primary tiller of a particular plant, GS plotted against day. For the 3 week period between days 40 and 60, more than 1 GS per day is being achieved.



(c) Observed GS 55 plotted against the proposed indicator – the correlation coefficient is 0.89 and the best line gradient 0.87.

Figure 9: Graphs of experiment and observation

We perform this algorithm on the signals extracted from the thresholded Hellinger distance images which are median filtered to reduce the (considerable) noise that they inevitably display.

There are clearly many ways this algorithm could be amended and improved, and the results which follow may accordingly be seen as a low water mark for success. Figure 9a depicts these line segments for four representative signals derived from one plant. In this example, the values of x_0 are 45, 48, 51 and 54 days.

We have experimented using the value x_0 for various distance thresholds as an indicator of GS. Our experiment was ground-truthed for GS 55 for some 270 plants and we have correlated these observations with this predictor for the arbitrarily chosen distances 0.4,0.35,0.3,0.25 (illustrated in Figure 8 and 9a). Of these, the distance 0.3 provided a correlation coefficient of 0.89 with ground truth – this is illustrated in Figure 9c. We are not predicting GS 55 in any absolute sense although the slope of the linear regression is 0.87 which is encouraging close to 1. The mean offset (underestimate) of 8.9 days would be added to the predictor to acquire GS55.

Collecting ground truth is labour-intensive, and so observations are only made every 2-4 days (this is clear in Figure 9c). Simultaneously, judgement of GS with precision by one individual, howsoever experienced, is very difficult to implement consistently over time and across seasons. Thus we can argue that the imperfect correlation we see (notably at ground truth 80) might well be considerably better as errors in GS 55 observations could very easily be up to 4 days. For one plant, we have ground-truthed estimates of GS, plotted in Figure 9b; we believe the patterns in these observations to be characteristic. GS increases by more than 1 per day over a 20 day period, then slows, then accelerates again as senescence sets in. During periods of high gradient we might expect estimates from image data to be less reliable, and stage 55 falls within this sensitive interval, further jeopardising accuracy. It is plausible to expect that mean observations of replicates of experimental conditions, for example over different genotypes, would further reduce inaccuracies in the predictor.

5 Conclusions and further work

We have demonstrated an approach to identification of oat panicle spikelets in the bulk collection of images of developing plants. The algorithms deployed adopt the usual classification approach of training on known data, and use a variant of the widely used and robust Local Pattern texture detector. All codes were written in standard, portable Matlab[3] or C[4]. The time required to process a single image is negligible – a small fraction of a second – and the time to process a series of images for a plant and deliver a prediction correspondingly tiny. Manual ground-truthing, of course, is very costly in resource.

Significant success in identification of image regions with young spikelets can be evidenced. This success can be reinforced by deploying obvious and justifiable domain knowledge such as confidence that panicles will appear in the upper part of the plant, and will almost certainly be enhanced by further experiments on a fully commissioned system delivering higher quality,

[3]MATLAB is a registered trademark of The MathWorks, Inc.

[4]The authors are happy to share codes on request.

consistent images. We also demonstrate the ability to detect second and subsequent flowering tillers, permitting automated determination of the range of commencement of flowering on a plant and its duration.

The detector has been exploited to predict a day on which growth onset commences; this measure shows good correlation with ground-truthed GS 55 for a large sample. It is highly probable that this correlation can be improved by refinements to the detector and better measurement.

We have reported here work in progress, which we can be developed in several directions:

- The detector we have built was – albeit successful – the first and simplest experiment. Extending this to more sophisticated variants is an obvious avenue of research: a finer quantisation of the 'circle' surrounding the target pixel, and/or a finer colour model, may improve performance and open the possibility of estimating the precise numbers of spikelets present. It is moot whether the implied significant growth in dimensionality would be worthwhile. It is also possible that entirely different texture analysis techniques, or exploration of colour spaces other than RGB, would assist.

- The indicator we derive from the detector is open to significant improvement. At least, the fitting to Equation 1 may be made much more robust in a number of ways (RANSAC (Fischler & Bolles 1981) is just one). More constructively, optimal extraction of day from thresholded signal, and optimal choice of threshold given this, may both be explored with every likelihood of improving results.

- The imaging system usually captures more than one view of the plant at each visit to the cabinets. Routinely, this is a side view at 0, another at $\frac{\pi}{2}$, and a third, top-down, view. We have performed no experiments on the top-down views but it is highly plausible that the two side views, coupled with the knowledge that spikelets represent a cluster in 3-space, would allow a 3-D reconstruction of the volume(s) occupied by the clusters.

- More generally, in all the image sequences we collect, the temporal evolution of the plant is of interest. Clearly, if a panicle is evident on a given day, it may be expected to be present in a similar location the next day, thereby easing and encouraging identification. More interestingly, in pinpointing GS, we might possibly deduce likely evidence the *preceding* day as well (this *forward-backward* reasoning has seen success in plant imaging elsewhere (Li et al. 2013)). As mentioned above, the trivial foreknowledge that panicles form above the (vertical) mid-point of the plant would immediately assist measurements.

  There remains interest in tracking the development of the panicle after GS 50-60. As is evident from Figure 8, the detector we present here, by design, highlights the earlier stages. Approaching Stage 70 the spikelets are almost exclusively yellow (see Figure 4), and the detector fails to 'see' the panicle. We have experimented with a detector trained on these which showed only modest success since the much simpler local patterns do not provide the discrimination of the red/green patterns. We are confident that this problem is amenable to efficient solution by using geometric foreknowledge (Stages 60+ must succeed Stages 50-59), and more complex colour or filter design models

- The emergence of anthers is another point of interest during panicle development – GS 65 represents 50% of the primary tiller's anthers being mature. In some plants this is visually observable, but in oats anthers are very slender and difficult to observe with any reliability in the images we acquire. Nevertheless, we are optimistic about pinpointing stages in excess of 70 as the spikelets yellow, which would then allow some interpolation of GS 65 given the work outlined above.

While this work is of direct benefit in existing experiments and high throughput installations, we see it further as an exemplar of the practicality of applying established computer vision techniques in plant breeding and biology. As is customary in cross-disciplinary work, it is critical for the computer scientist to engage properly with what the domain experimenter is trying to find out: thereafter it is possible that information automatically extractable may be of great benefit, but may not correspond directly with traditional approaches. Specifically, we suspect that Growth Stages habitually measured by hand may not be the simplest to extract automatically, but a reproducible and reliable identification of other criteria would prove to be of equal or more value.

6 Acknowledgements

We gratefully acknowledge the work of Andrew Rawlins and Andreu Alcalde in keeping the robotic greenhouse system functioning, and Hannah Dee for discussions on image analysis.

The National Plant Phenomics Centre is a BBSRC-funded national facility hosted by IBERS. RB was funded by BBSRC grant BB/J004464/1 and CH by BB/J004405/1.

Ground truth was provided by Markku Farel and Matt Lowe.

Anonymous referees provided very useful criticism and comment on the first draft of this paper.

References

Al-Tam, F., Adam, H., dos Anjos, A., Lorieux, M., Larmande, P., Ghesquière, A., Jouannic, S. & Shahbazkia, H. R. (2013), 'P-TRAP: a panicle trait phenotyping tool', *BMC Plant Biology* **13:122**. 14pp, `http://www.biomedcentral.com/1471-2229/13/122`.

APP (2014), 'Australian Plant Phenomics Facility'. `http://www.plantphenomics.org.au/`.

BBC (2001), Growth stages of mono-and dicotyledonous plants, Technical report, Federal Biological Research Centre for Agriculture and Forestry. BBCH Monograph. Available at `http://www.bba.de/veroeff/bbch/bbcheng.pdf`.

Campillo, C., Garcia, M., Daza, C. & Prieto, M. (2010), 'Study of a non-destructive method for estimating the leaf area index in vegetable crops using digital images', *Hortscience* **45**(10), 1459–1463.

e-A (2014), 'e-Anatomy'. `http://www.imaios.com/en/e-Anatomy`.

F, A. & G, C. (2012), 'The genetic basis of flowering responses to seasonal cues', *Nature Reviews Genetics* **13**, 627–639.

Fischler, M. A. & Bolles, R. C. (1981), 'Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography', *Communications of the ACM* **24**(6), 381–395.

Furbank, R. T. & Tester, M. (2011), 'Phenomics technologies to relieve the phenotyping bottleneck', *Trends in Plant Science* **16**(12), 635 – 644.

Gertych, A., Zhang, A., Sayre, J., Pospiech-Kurkowska, S. & Huang, H. K. (2007), 'Bone age assessment of children using a digital hand atlas', *Computerized Medical Imaging and Graphics* **31**(4-5), 322–331.

Hartmann, A., Czauderna, T., Hoffmann, R., Stein, N. & Schreiber, F. (2011), 'HTPheno: An image analysis pipeline for high-throughput plant phenotyping', *BMC Bioinformatics* **12**(1), 148.

Hellinger, E. (1909), 'Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen', *Journal für die reine und angewandte Mathematik* **136**, 210271.

Holland, J. B., Portyanko, V. A., Hoffman, D. & M, L. (2002), 'Genomic regions controlling vernalization and photoperiod responses in oat', *Theoretical and Applied Genetics* **105**, 113–116.

Huang, C., Yang, W., Duan, L., Jiang, N., Chen, G., Xiong, L. & Liu, Q. (2013), 'Rice panicle length measuring system based on dual-camera imaging', *Computers and Electronics in Agriculture* **98**, 158–165.

JPP (2014), 'Jülich Plant Phenotyping Centre'. `http://www.fz-juelich.de/ibg/ibg-2/DE/Organisation/JPPC/JPPC_node.html`.

Kubassova, O. A., Boyle, R. D. & Radjenovic, A. (2007), 'Quantitative analysis of dynamic contrast-enhanced MRI datasets of the metacarpophalangeal joints', *Academic Radiology* **14**(10), 1189–1200.

Li, Y., Fan, X., Mitra, N. J., Chamovitz, D., Cohen-Or, D. & Chen, B. (2013), 'Analyzing growing plants from 4D point cloud data', *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2013)* **32**(6), 157.
URL:`http://doi.acm.org/10.1145/2508363.2508368`

Locatelli, A. B., Federizzi, L. C., Milach, S. C. K. & McElroy, A. R. (2008), 'Flowering time in oat: Genotype characterization for photoperiod and vernalization response', *Fields Crop Research* **106**(3), 242–247.

Locatelli, A. B., Federizzi, L. C., Milach, S. C. K., Wight, C. P., Molnar, S. J., Chapados, J. T. & Tinker, N. A. (2006), 'Loci affecting flowering time in oat under short-day conditions', *Genome* **49**, 1528–1538.

Nava, I. C., Wight, C. P., Pacheco, M. T., Federizzi, L. C. & A, T. N. (2012), 'Tagging and mapping candidate loci for vernalization and flower initiation in hexaploid oat', *Molecular Breeding* **30**, 1295–1312.

NPP (2014), 'The UK National Plant Phenomics Centre'. http://www.plant-phenomics.ac.uk/en/.

Ojala, T., Pietikainen, M. & Maenpaa, M. (2002), 'Multiresolution gray-scale and rotation invariant texture classification with locally binary patterns', *IEEE Transactions of Patttern Analysis and Machine Intelligence* **24**, 971–987.

Reis, M., Morais, R., Peres, E., Pereira, C., Contente, O., Soares, S., Valente, A., Baptista, J., Ferreira, P. & Bulas Cruz, J. (2012), 'Automatic detection of bunches of grapes in natural environment from color images', *Journal of Applied Logic* **10**(4), 285–290.

Sirault, X., Fripp, J., Paproki, A., Guo, J., Kuffner, P., Daily, H., Li, Rongxina nd Sirault, X. & Furbank, R. (2013), PlantScan: a three-dimensional phenotyping platform for capturing the structural dynamic of plant development and growth, *in* '7th International Conference on Functional-Structural Plant Models', p. 75.

Song, Y., Glasbey, C., Horgan, G., Polder, G., Dieleman, J. & van der Heijden, G. (2014), 'Automatic fruit recognition and counting from multiple images', *Biosystems Engineering* **118**, 203–215.

Šonka, M., Hlaváč, V. & Boyle, R. (2014), *Image processing, analysis, and machine vision*, 4^{th} edn, CEngage.

Tan, X. & Triggs, B. (2010), 'Enhanced local texture feature sets for face recognition under difficult lighting conditions', *IEEE Transactions on Image Processing* **19**, 1635–1650.

Tanner, J. M. & Whitehouse, R. H. (1975), *Assessment of Skeletal Maturity and Prediction of Adult Height (TW2 Method)*, Academic.

Tinker, N. A., Kilian, A., Wight, C. P., Heller-Uszynska, K., Wenzl, P., Rines, H. W., Bjornstad, A., Howarth, C. J., Jannink, J.-L., Anderson, J. M., Rossnagel, B. G., Stuthman, D. D., Sorrells, M. E., Jackson, E. W., Tuvesson, S., Kolb, F. L., Olsson, O., Federizzi, L. C., Carson, M. L., Ohm, H. H., Molnar, S. J., Scoles, G. J., Eckstein, P. E., Bonman, J. M., Ceplitis, A. & Langdon, T. (2009), 'New DArT markers for oat provide enhanced map coverage and global germplasm characterization', *BMC Genomics* **10:39**. 22pp, http://www.biomedcentral.com/1471-2164/10/39.

Worland, A. J. (1996), 'The influence of flowering time genes on environmental adaptability in European wheats', *Euphytica* **89**, 4957.

Zadoks, J. C., Chang, T. T. & Konzak, C. F. (1974), 'A decimal code for the growth stages of cereals', *Weed Research* **14**, 415–421.