Rumex and Urtica Detection in Grassland by UAV

Adam Binch[2] Nigel Cooke[2] Charles Fox[1,2]

[1] School of Computing, The University of Lincoln, UK

[2] Ibex Automation Ltd, UK

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – June 27, 2018
Montreal, Quebec, Canada

Abstract. *Previous work (Binch & Fox, 2017) used autonomous ground robotic platforms to successfully detect Urtica (nettle) and Rumex (dock) weeds in grassland, to improve farm productivity and the environment through precision herbicide spraying. It assumed that ground robots swathe entire fields to both detect and spray weeds, but this is a slow process as the slow ground platform must drive over every square meter of the field even where there are no weeds. The present study examines a complimentary approach, using unmanned aerial vehicles (UAVs) to perform faster detections, in order to inform slower ground robots of weed location and direct them to spray them from the ground. In a controlled study, it finds that the existing state-of-the-art (Binch & Fox, 2017) ground detection algorithm based on local binary patterns and support vector machines is easily re-usable from a UAV with 4K camera despite large differences in camera type, distance, perspective and motion, without retraining. The algorithm achieves 83-95% accuracy on ground platform data with 1-3 independent views, and improves to 90% from single views on aerial data. However this is only attainable at low altitudes up to 8 feet, speeds below 0.3m/s, and a vertical view angle, suggesting that autonomous or manual UAV swathing is required to cover fields, rather than use of a single high-altitude photograph. This demonstrates for the first time that combined aerial detection with ground spraying system is feasible for Rumex and Urtica in grassland, using UAVs to replace the swathing and detection of weeds then dispatching ground platforms to spray them at the detection sites (as spraying by UAV is illegal in EU countries). This reduces total time requires to spray as the UAV performs the survey stage faster than a ground platform.*

Keywords. *UAV, weeding, Rumex, Urtica, robotics*

1 Introduction

Autonomous weeding of grassland may enhance dairy and sheep farm productivity in less favoured farm areas, and also reduce chemical herbicide usage in these environments and run-off into their surrounding ecosystems, including the human water supply. Many such farms, unlike large agribusiness operations, are small family-owned businesses operating at low margins. This includes areas of hill grassland which is currently uneconomical to spray either by precise manual backpack sprayer application to weeds of generic herbicides (at minimum labour wage costs) or by tractor-mounted bulk spray (due to the higher costs of selective herbicides). Hence, these areas remain unsprayed and the benefits of additional food production via grazing of the areas occupied by weeds is not realised. Autonomous robotic precision spraying of these weeds provides a possible alternative solution, which could spray low doses of generic herbicide onto individual weeds, like a manual backpack sprayer, but without the associated and prohibitive minimum wage labour costs.

In a recent study [1], we performed a comprehensive evaluation and optimisation of all known machine vision approaches for detection of Rumex and Urtica in grassland from a ground robot platform. After standardising and optimising the methods, for the first time on a "big" data set large enough to report the comparison with high statistical significance, we found that a combination of Local Binary Patterns (LBP) with Support Vector Machine (SVM) with certain parameters gave practically usable performance, giving 83% accuracy on classification of 12cm square regions of ground, and rising to 95%+ where two or more observations of the same region are fused. This was notably lower than previously claimed results by original authors of the algorithms tested, due to the higher standard of independent testing imposed in the comparison study, and the new and larger data sets used.

To move further towards a practical system for weed control, it is next necessary to consider how such recognition should be used as a part of practical agri-systems methods. The previous study initially assumed that weed detection and spraying would be performed by the same ground robot, swathing (i.e. driving up and down in a ‘lawnmower stripes pattern’) over each square meter of a field, detecting and spraying any weeds found there. We have performed pilot studies with ground platforms following this approach, and along with the results of the previous paper, have found two issues with it. First, swathing with a ground platform is inefficient because weeds are not evenly distributed in most fields, and there are many areas which have no weeds in them at all. Driving over every square meter is a waste of time and (more importantly for unmanned systems) energy. Human manual sprayers do not blindingly swathe fields, but use both their prior knowledge (weeds are more likely to grow around edges of fields, and in shaded, damp areas) and their distant vision (distant uncertain dark green blobs may or may not be weedy areas, so warrant travelling to investigate them in more detail). Second, the previous study discovered a strong drop-off in recognition performance as a function of distance from the ground robot, due to its perspective warping assumptions breaking down with distance. While it would be possible, as in other studies (most of which investigated the much simpler problem of weed detection in arable row crops rather than grassland), to mount the robot’s camera facing vertically downwards instead of its current 22.5 degrees, this would preclude it from detecting weeds not directly under it, which in turn would preclude the fusion of multiple observations of weeds ahead of the vehicle before spraying which were shown to be needed to obtain the higher 95% accuracies.

Both of these issues could be addressed by a two-platform system, using unmanned aerial vehicles (UAVs) to perform detection and the ground platform to relocate the detections and perform the spraying. (EU law prohibits spraying from UAVs directly). UAVs can view much larger areas - potentially whole fields - in single camera frames, and/or may also aerially swathe over areas at low altitudes more quickly than ground platforms if higher resolution is required. Views from UAVs are different than from ground platforms, and it is not obvious whether systems developed for ground use are re-usable from the air. The present study tests this hypothesis, including UAV images both directly vertical and at a 45 degree tilt (similar to training data from the ground platform). It also measures the effect of altitude to give the first indication of whether aerial swathing is required or whether single high-altitude images can be used, thus providing the first estimates of the time required to detect weeds over whole fields in this manner.

2 Methods

The objective of the experiments outlined here is to report the performance of a method utilised for the classification of grass vs weeds (i.e. a two-way classification). More specifically we wish to assess the performance of the method when there is a miss-match between the training data and the test data, where the training data was captured from cameras mounted to a ground robot and test data was captured from cameras mounted to a UAV. Further, we assess the performance of the method for two types of test data captured using the UAV. Firstly using images captured at 90° pitch down (i.e. a birds-eye view) and secondly using images captured at 45° pitch down (yielding a perspective view of the ground).

The performance of the method was assessed by training it on the training images and then using the method's (learned) parameters to predict the test images (i.e. an image containing grass or weeds).

2.1 Training image acquisition and pre-processing

A plot of Rumex and a plot of Urtica, each measuring approximately 3m squared, were built on a dairy grass-land farm (South Yorkshire, UK) in a region of the farm exposed to direct sunlight throughout the day. This was achieved by taking slabs of Rumex and Urtica, each measuring approximately 0.2 m squared, from a working field and transplanting them into their respective trenches.

An auto-focusing camera (C920, *www.logitech.com*) was mounted to a tracked robot on its left side facing out at right angle yaw to the direction of travel, and at 22.5 degrees pitch down. The robot travelled in circles around each plot (figure 1) capturing stereo pair images at 1080HD resolution. The distance between the robot and the edge of the plots was randomised between 0 m and 1m. This setup was devised so that the plot contents filled the images and gave a clear view over approximately 1 m square area of ground. Images of grass were also acquired by driving the robot on a random path through working fields. Approximately half of this data was acquired under bright or sunny weather conditions, whilst the other half was acquired under overcast weather conditions. Data was captured during May 2016, over the course of 4 days at random times of day from sunrise to sunset. Approximately a third of a terabyte of usable image data was acquired in total.

Images were pre-processed in three steps: colour calibration, perspective dewarping, and windowing. Images undergo a continuum of colour transformations due to lighting and weather conditions. In order to compensate for these effects images were modified using a 'colour calibration' algorithm. To a certain extent this operation acts to colour standardise images. Images were colour calibrated by measuring the blue, green and red light intensities from a colour bar (figure **Error! Reference source not found.**) present in images and adjusting the colour intensities in comparison images (i.e. images that were to be colour calibrated) to be as close as possible to the colour intensities of a reference image (for a full account of this process see [1]).

Perspective normalisation was performed via a projective transform which maps a 1.16 m width by 1.0 m depth ground area (figure **Error! Reference source not found.**) into a 700 pixel width by 600 pixel height image, whose geometry is identical to that of a vertical, overhead camera. A single spray target radius was judged to be a ground area of 106 mm squared which corresponds to a 64 × 64 square pixel region of the image. Thus the dewarped image was split into regular 64 × 64 contiguous and non-overlapping windows.

(a) *(b)*

Figure 1*: Training image acquisition (from [1]).*
(a) Plot of Urtica and ground robot (IBEX2) capturing training images from the plot using its on-board cameras. (b) Raw training image of Urtica obtained form the ground robot's cameras. Note the colour bar used for the colour calibration process and the grid annotation (outlined area) used for the perspective normalisation described in the main text (section 2.1). Image windows were obtained from the annotated region.

2.2 Test image acquisition and pre-processing

Images of grass, Rumex and Urtica were captured from 4K resolution, autos-stabilised cameras mounted to a DJI Phantom 4 UAV (*www.dji.com*) in May 2017, on the same field from which the ground robot images were acquired. All images were acquired over the course of a single afternoon under overcast weather conditions. Images of Rumex were captured by flying the UAV over a large region of a working field containing large patches of Rumex. Images of grass were captured by flying the UAV over a large region of the field containing grass only (care was taken to avoid the occurrence of weeds in these images). Images of Urtica were acquired by hovering the UAV (and moving it back and forth slightly) over the Urtica garden (note that the Urtica garden had undergone a year's worth of growth since the acquisition of the ground robot images, whilst the Rumex garden had been destroyed prior to the acquisition of the UAV's images). The UAV captured images at a vertical distance of 1 m above the ground contents. Two sets of UAV images (where a single set consists of images of grass, Rumex and Urtica) were acquired. In the first set the UAV's cameras were adjusted to face 90° pitch down yielding images with a 'birds-eye-view' of the ground (figure 2), while in the second set the UAV's cameras were adjusted to face 45° pitch down yielding images with a perspective view of the ground (figure **Error! Reference source not found.**).

For the 45° images, perspective normalisation was performed via a projective transform which maps the ground area indicated by the annotated region in figure **Error! Reference source not found.** into a 1249 pixel width by 580 pixel height image whose geometry is identical to that of a vertical, overhead camera. The 1249×580 image was split into regular 64 × 64 contiguous and non-overlapping windows. For the 90° images the *whole* image was split into regular 64 × 64 contiguous and non-overlapping windows. Some of the Rumex images (both 90° and 45°) had to be removed because they contained large patches of grass.

(a) *(b)*

***Figure 2**: Test Image Acquisition.*

(a) Raw 90° test image of Urtica captured from the UAV's cameras. (b) Raw 45° test image of Urtica captured from the UAV's cameras. The grid annotation (outlined area) used for the perspective normalisation is also shown. Image windows were obtained from the annotated region.

2.3 Classification method

The classification method used in the present analysis involves extracting features known as local binary patterns (LBPs) [2] (figure **Error! Reference source not found.**) from image windows which are then passed into a support vector machine (SVM) [3] for classification.

Local binary patterns (LBP) are family of texture description features. Local means they are computed on local sub-windows of an image only, as a function of a center pixel and either its immediate or *r* pixel radius neighbours; binary means that the feature vector is binary, with each feature classed as either present or absent. Windows are converted to greyscale, then for each pixel $x_{i,j}$ in the window, LBP computes a vector,

$[(x_{i;j} > x_{i;j+r}); (x_{i;j} > x_{i-r;j+r}); (x_{i;j} > x_{i-r;j}); (x_{i;j} > x_{i-r;j-r});$
$(x_{i;j} > x_{i;j-r}); (x_{i;j} > x_{i+r;j-r}); (x_{i+r;j} > x_{i;j}); (x_{i;j} > x_{i+r;j+r})].$

There are 2^8=256 possible values of this feature vector. LBP computes the vector for each pixel in the window, then computes a 256-point histogram of the obtained values over the window. Further details are given interval [1] and the references therein.

Support vector classifiers (SVCs) separate images into their classes, firstly by creating a finite dimensional feature vector space and secondly by finding the linear affine hyper-plane that maximises the distance (known as the margin) between image classes in this feature space. Support vector machines (SVMs) extend the SVC by using non-linear functions known as kernels to enlarge the feature space. In the higher dimensional space the hyper-plane is linear, but non-linear when projected back onto the original feature space. As a consequence SVMs are more effective than SVCs as there are more dimensions over which different classes can be separated. Both SVCs and SVMs are modified by a 'tuning' hyper-parameter *C* which controls how many observations (training images) are permitted to violate the margin. The kernel used in the present analysis is the radial (RBF) kernel which has its own hyper-parameter γ (for a fuller yet concise explanation of the SVM with RBF kernel and its hyper-parameters see [1]).

(a) *(b)*

***Figure 3**: Feature extraction.*
(a) Image window of Urtica. (b) The local binary pattern (LBP) image of 3, with n=24, r=4. Note that the jagged pattern of the Urtica leaves is apparent in the LBP image.

For both training and testing, the hyper parameters of the LBP were set to n=24, r=4 and the hyper-parameters of the SVM were set to $C=2^9$, $\gamma=2^3$. These hyper-parameter values were obtained using the optimisation method outlined in [1].

2.4 Experiments

To train the classifier, a training set of 40,000 image windows was selected from the data set obtained from the ground robot's cameras. Half the image windows contained grass under an equal mixture of sunny and overcast weather conditions, whilst the other half contained weeds (an equal mixture of Rumex and Urtica) under an equal mixture of sunny and overcast weather conditions. This selection was achieved by uniform random sampling across the subsets of the data containing grass and weeds, respectively. Grass image test sets and weed image test sets (for any given experiment) each contained 1710 image windows selected using uniform random sampling. Weed image test sets contained an equal mixture of Rumex and Urtica.

We wished to assess classification performance as a function of the speed of the UAV. However, there are two problems with regard to obtaining speed results from the test data. Firstly, the test data was not acquired with the intention of obtaining speed results. Thus there were no subsets of the data acquired at pre determined speeds. In order to obtain these subsets we computed the dense optical flow occurring in a given image using a method based on Gunner Farneback's algorithm [4]. Note that this method computes optical flow for all points in the image and is therefore a global measure of optical flow (as opposed to a local measure of flow computed at interest points in the image). Due to the forward motion of the UAV and the (relatively) static nature of the ground contents a global measure of flow was deemed preferable as all pixels were assumed to be moving across the image at equal speeds and in the same direction. For two consecutive images, the Gunner Farneback method yields a change in position (i.e. a change in x and a change in y) for every pixel in the image and thus a Euclidean distance for all pixels in the image. Taking the median Euclidean distance provides a global measure of speed (distance over time) for each image. Then images can be binned into various speed intervals. Each bin thus contains a set of images that can be windowed and stored for analysis.

The second problem with obtaining speed results was due to a lack of speed variation in the weed images (Rumex and Urtica). For Rumex, there were only a few images containing mostly

Rumex (without large patches of grass). Those images were all captured when the UAV was moving at an approximately uniform speed. For Urtica, the UAV captured images over the Urtica garden. Thus the UAV remained localised (i.e. not moving or only moving very slightly) when capturing images of Urtica. In order to acquire classification performance for overall grass vs weed detection we assumed that the ratio between detection rates for grass and detection rates for weeds remains constant with respect to the speed of the UAV. To obtain this ratio we first measured classification performance using images captured at speeds ranging between 0.09 and 0.16 m/s (i.e. the smallest UAV speed bin considered in these experiments). For this experiment, we obtained the break down accuracies for grass and for weeds, and thus the ratio between them. Then, grass classification accuracy as a function of UAV speed was obtained from the grass image data (downsampled to 1080HD) and those accuracies and the ratio was used to obtain a proxy for weed classification accuracy as a function of UAV speed. Taking the average of those two values yields the overall accuracy as a function of UAV speed. Note that this experiment was carried out using the 90° only.

We also wished to assess classification performance as a function of the simulated altitude of the UAV. As the UAV increases in altitude its cameras cover a larger area of ground. At the same time a given (for example a 1m by 1m) area of ground is represented at a correspondingly smaller resolution. Specifically, if the UAV doubles its altitude, then the 1m square area of ground is represented in the image by half the number of pixels (a 3840×2160 pixel image covering a 1m square area of ground taken at an altitude of 3.28 feet (1m) corresponds with a 1920×1080 pixel image covering a 1m square area of ground taken at an altitude of 6.56 feet (2m)). Thus, simulating images taken by the UAV at different altitudes can be achieved by down-sampling its 4K images to different values. However, these images need to be subsequently up-sampled to 1080HD to correspond with the ground robot training data (such that a 64×64 windowed image covers a ground area of 106 mm squared - a single spray target radius). Note that these operations were performed using the 90° (birds-eye-view) images only. Also note that classification performance as a function of the simulated altitude of the UAV was assessed using images captured in the lowest speed bin.

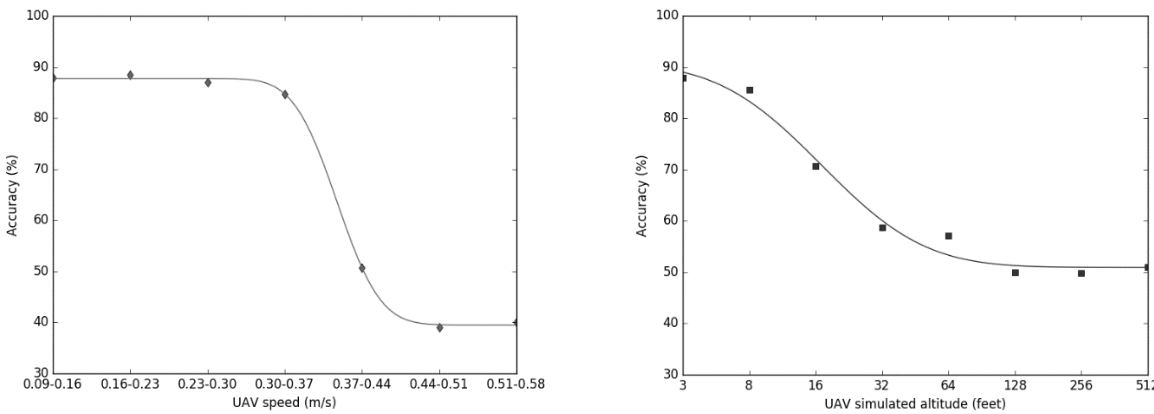
3 Results

Table 1 gives the results of the experiment in which the ratio of grass to weed classification accuracy was determined (break down accuracies for grass and weeds) (first row: 45°, second row: 90°) which are compared with the results (third row) of the experiment with the ground robot test data, which was reported in [1]. ACC gives the proportion of images correctly classified (i.e. the overall accuracy of the method) and CI are confidence intervals justifying the significance of the accuracy percentages to two decimal places. The breakdown accuracies for grass, weed, Rumex and Urtica presentations are also given.

Experiments using the UAV test data (for both 90° and 45° images) yielded higher accuracies than the experiment using the ground robot test data, and the 45° UAV images yielded a higher accuracy than the 90° UAV images. Specifically, the 90° and 45° UAV images yielded accuracies of 87.89% and 90.39%, respectively, whilst the ground robot images yielded an accuracy of 82.88%. The better performance of the experiment with the 45° images as compared with the performance of the experiment with the 90° images is due (mainly) to their difference in grass image classification, with approximately 5% greater accuracy for the 45° images. For the UAV test data, Urtica classification was more accurate than Rumex classification, whilst the opposite pattern was found for the ground robot test data.

Figure 4(a) gives the accuracy of the classification method as a function of the speed of the UAV, in intervals spanning 0.07 m/s. The diamond shaped markers are the data points and the curved line is the best fitting mirrored cumulative Gaussian modified with an additive (controlling intercept) and multiplicative (controlling magnitude) constant. The accuracy of the classification method decreases sharply beyond speeds of 0.3 m/s.

Figure 4(b) gives the accuracy of the classification method as a function of the simulated altitude of the UAV increasing in powers of 2. The square markers are the data points and the curved line is the best fitting cumulative Gaussian, again modified with an additive and multiplicative constant. The accuracy of the classification method decreases sharply beyond altitudes of 8 feet.



(a) (b)

***Figure 4**: Accuracy vs UAV speed and accuracy vs simulated UAV altitude.*
(a) Accuracy of the classification method as a function of simulated UAV altitude. UAV altitude is given on a logarithmic scale and increases in powers of 2. The square markers are the data points and the curved line is the best fitting cumulative Gaussian (see main text). (b) Accuracy of the classification method as a function of UAV speed with speed bins spanning an interval of 0.07 m/s. The square markers are the data points and the curved line is the best fitting cumulative Gaussian (see main text).

Experiment	ACC	CI	GRASS	WEEDS	RUMEX	URTICA
UAV 90°	87.89	3.94×10^{-3}	90.4	85.4	74	96.8
UAV 45°	**90.39**	3.56×10^{-3}	95.4	85.4	73.1	97.6
Ground Robot	82.88	5.96×10^{-3}	87.1	78.7	86	78.3

***Table 1.** Results of applying the classification method to UAV 90° and 45° test images. The results (reported in [1]) of applying the same method to test images captured from the ground robot are also shown for comparison. ACC is the over all accuracy, i.e. the probability that a random image is classified correctly. CI is the confidence interval in the estimate of ACC. GRASS and WEED are the probabilities that images of grass, or weeds, respectively, are correctly classified. RUMEX and URTICA are the probabilities that images of Rumex, or Urtica, respectively, are correctly classified.*

4 Discussion

The results show that simple re-use of the existing state-of-the art ground-platform algorithm [1] does perform sufficiently accurately on UAV imagery to make hybrid UAV-detection and ground-spraying weed control possible and economical. The system performs well at low altitudes, up to around 8 feet, then quickly falls off with higher altitudes. This is likely due to 8 feet being the greatest altitude at which individual blades of grass can be seen in the 4K pixel images; beyond this, regions of grass becomes blurred and indistinguishable from broader-leafed weeds. This in turn suggests that to fully check whole fields for weeds, UAVs must perform some swathing rather than work from single high-altitude images. Such swathing could be automated, requiring more effort than taking a single manual high altitude image; but would operate much more quickly and using less energy than swathing with heavy ground platforms, so is worth further investigation.

It is interesting that the accuracy on the UAV test set is actually *higher* than the state-of-the-art results of the same algorithm on ground platform images, even though the algorithm is trained entirely from mis-matched ground platform image data. The training data used from the ground platform were colour calibrated, using physical color charts mounted in camera view on the ground robots. It was not possible to mount similar physical charts to the UAV in this study due to weight requirements. The training data were taken from ground platforms at 22.5 degree down-facing cameras, while the UAV images were taken vertically or at 45 degrees. The gain in performance was thus unexpected. Possible explanations may revolve around the use of perspective dewarping assumptions in the ground platform case. There, we assumed that the ground was perfectly flat in order to enable an affine transform to de-warp perspective images to give estimates of what a true vertical view would look like. As weeds get taller, this assumption becomes less valid, and results in more distant regions becoming more distorted from the true aerial view. Although the system was fully trained on such images, including distant distorted images, it may have learned to recognise ideal true vertical views despite the noise in such training images. Then when tested on ground images, it suffers a performance penalty from distance distorted images in the test set; while when tested on real vertical UAV images it does not. For example, in the limiting case, the most extremely distorted in both ground training and ground test sets might become completely blurred and contain zero information. For ground training they will be effectively ignored, and for ground testing they will lower the accuracy rate, without lowing performance on the UAV data. Similarly, it is possible that a higher quality camera on the UAV (specifically, it is equipped with motion stabilisation while the ground platform was not) has the same effect. Or the improvement may be simply due to chance because the aerial test set is a different dataset from the ground test set. It includes some of the same physical weed plants as the ground data, but is not directly comparable, being recorded several months later under different growing and image lighting conditions. Further work is needed to test this in a more controlled setting, as in the original state of the art ground study which made very careful controls of all such factors.

For the UAV images, there are only a few images of Rumex that contain mostly Rumex (without big patches of grass). But these images still contain small patches of grass and are blurry. The images of grass and Urtica that were selected for windowing are very clear by comparison. This explains the poorer performance for Rumex compared with grass and Urtica. Future work should use a more controlled Rumex environment such as the controlled, deliberately cultivated "weed gardens" of the state of the art ground study to replace the "found data" used here. Similarly, use of controlled data would improve UAV accuracy if enough could be collected to provide new matched aerial training sets, including exact matching of perspective and camera angles between training and test sets. The height of the UAV was controlled by a human pilot in this study for technical reasons which varied somewhat and introduced size distortions to the dewarped images. Classification may be improved if automated height control, or measurement, is used to compensate. These considerations suggest that accuracy can be improved further given additional research work.

The speed of the UAV during imaging may be important if it is found to blur the images and lower the accuracy. New business-based studies should balance the business costs of longer flight times against accuracies as both altitude and speed are varied.

Acknowledgements

This work was funded in part by InnovateUK grant number 131790 IBEX2.

References

[1] Binch A, Fox CW (2017). Controlled comparison of machine vision algorithms for Rumex and Urtica detection in grassland. *Computers and Electronics in Agriculture* 140: 123-138.

[2] He, Dong-Chen, Li-Wang (1990). Texture unit, texture spectrum, and texture analysis. *IEEE transactions on Geoscience and Remote Sensing* 28.4: 509-512.

[3] Vapnik Vladimir (2013). The nature of statistical learning theory. *Springer, Berlin, Heidelberg.*

[4] Farneback G (2003). Two-Frame Motion Estimation Based on Polynomial Expansion. In: Bigun J., Gustavsson T. (eds) Image Analysis. SCIA 2003. Lecture Notes in Computer Science, vol 2749. *Springer, Berlin, Heidelberg.*