Paper presented at: http://ist2016.ieee-ims.org/

Weed Recognition Framework for Robotic Precision Farming

Tsampikos Kounalakis*, Georgios A. Triantafyllidis† and Lazaros Nalpantidis*
*Department of Mechanical and Manufacturing Engineering, Aalborg University, Copenhagen, Denmark
Email: tkoun@m-tech.aau.dk, lanalpa@m-tech.aau.dk
†Department of Architecture, Design & Media Technology, Aalborg University, Copenhagen, Denmark
Email: gt@create.aau.dk

***Abstract*—In this paper, we introduce a novel framework which applies known image features combined with advanced linear image representations for weed recognition. Our proposed weed recognition framework, is based on state-of-the-the art object/image categorization methods exploiting enhanced performance using advanced encoding and machine learning algorithms. The resulting system can be applied in a variety of environments, plantation or weed types. This results in a novel and generic weed control approach, that in our knowledge is unique among weed recognition methods and systems. For the experimental evaluation of our system, we introduce a challenging image dataset for weed recognition. We experimentally show that the proposed system achieves significant performance improvements in weed recognition in comparison with other known methods.**

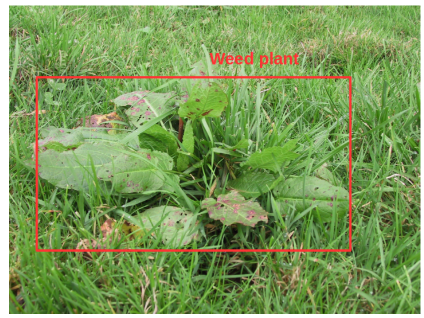

Fig. 1. An example output of a weed detection and recognition system.

I. Introduction

Weed control is a big problem for agricultural production. Weed plants can either compete with crops or be harmful when consumed by livestock. Currently, weed control in large scale agriculture is performed using pesticides, which diminish the quantity, quality and increases production cost of agricultural products. Others solutions, such as organic or manual treatments are not economically viable when applied in large scales. Therefore, automatic weed recognition and detection, an example is shown in Fig. 1, can play an important role towards the improvement of modern agriculture.

This research topic comes with great challenges due to the broad variety of weed plants and working environments. The majority of methods concerning weed recognition and detection have been using specialized robotic systems [1], [2], [3], [4], [5], [6], [7]. All systems propose a precise weed control, which meets specific key features. Firstly, weeds must be eliminated while preserving rest of flora. Also, speed of execution contributes to the production time and cost of agricultural products. Finally, the effectiveness of such systems, i.e, not wasting energy or supplies during operation, is also a major factor.

However, current weed control robotic systems use very simple detection and recognition based on simple visual recognition algorithms, which are not robust to environmental parameters or different weed control problems. Hence, each weed recognition system is an individual application with no connection to previous methods, resulting to an undefined research framework. On the other hand, recently, Kazmi in [8] proposed the use of feature detection and extraction methods, thus implementing object/image categorization techniques for weed recognition. Results show that local features are able to efficiently represent and categorize weed plants from sugar beets.

Inspired by state-of-the-art object/image recognition systems, we propose a framework for robust weed recognition that follows the structure of Fig. 2. Our proposed framework is compatible with all known image features, and does not require any segmentation for image pre-processing. The robust image features used from the proposed framework can be applied to challenging environments, any kind of weed plant and plantation type. Image features are either extracted using a feature detection algorithm or over dense grid of non-overlapping feature patches capturing feature information from the whole image. We also implement known feature encoding techniques [9], [10] in order to enhance the discriminatory capabilities of extracted features. We present benefits of image representations when combined with linear feature encoding techniques. These encoding techniques create linearly separable image representations, which can be used in combination with supervised learning linear classifiers yielding improved recognition results. The above conclude to a generic weed recognition system, which to our knowledge has not been attempted before.

The *contribution* of this work is twofold. Firstly, we propose a novel weed recognition framework that can be easily adapted to any kind of weed control problem. The proposed framework allows different combinations between feature de-

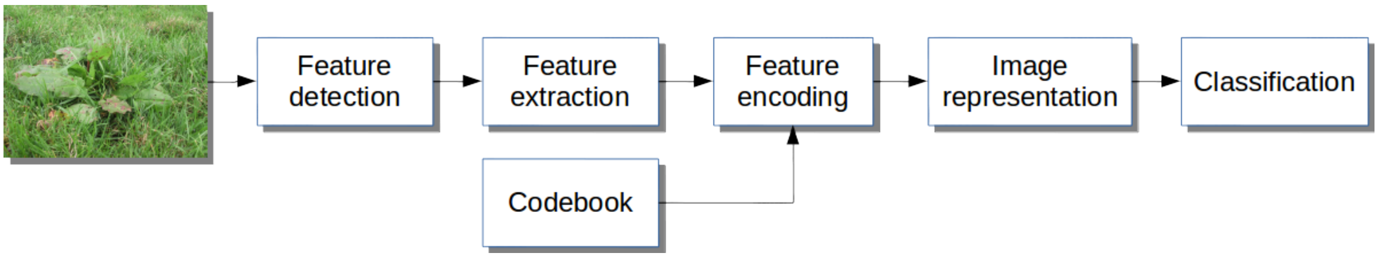

Fig. 2. Block diagram describing a generic weed recognition framework.

tection, extraction and encoding methods resulting to a weed control system of high recognition performance. Secondly, we introduce a new, challenging weed classification dataset. This dataset, dubbed *Rumex 100*, is compiled from images depicting (or not) Broad-leaved dock weed plants (Rumex obtusifolius L.)[1].

The rest of the paper is organized as follows, in Section II, we review related work from recent weed control robotic systems and weed recognition methods. In Section III, we present our novel weed recognition system. Our new publicly available weed recognition dataset is presented in Section IV. In Section V, we describe the advantages of our proposed weed classification method, followed by a detailed experimental assessment. Conclusions are drawn in Section VI.

II. Related work

As described in the introductory part, implementations of weed control robotic systems in precision agriculture is a rapidly developing research topic.

Authors in [1], propose a weed detection system where segmented weed images are represented using wavelet analysis. The system in [2], propose a multi-feature approach extracting shapes, skeletons and color. More recently, the work in [7] uses image correlation between the input image and sample images due to high operational speed. Moreover, weed detection is also applied to Unmanned Aerial Vehicles (UAVs), i.e, as weed mapping. [3], the system uses Hough transform as image features for small image patches. More recently, in [4] proposed the use of object-based image analysis features (OBIA). For the recognition stage, some of the aforementioned systems either use a simple comparison between features [7] or image features combined with classification algorithms such as k-means [3] and Support Vector Machines (SVMs) [1], [2], [4].

Nevertheless, robotic systems used for weed control are currently using simple recognition methods, which are outperformed by state-of-the art image categorization techniques. As a result, their proposed approach makes all aforementioned systems application-specific, thus restricting changes in implementation or plantation types.

More recently, the work in [8] proposed a weed detection and recognition method using image features and an object/image categorization framework [11]. Authors propose a feature detection and extraction method which is compared with other known feature detection [12], [13] and feature extraction methods [12], [14], [15]. The color-based feature extraction method uses image segmentation preprocessing, thus capturing information from sugar beet and thistle plants. For image representation, this method use a two-level spatial histogram model like the one proposed in [16] combined with a k-d tree encoding method. Classification is performed using a nonlinear SVM combined with the χ^2 kernel [17], due to the nonlinearity of produced image representation vectors. Color-based image features using the proposed framework [11], were experimentally shown to provide competitive results for the examined dataset.

The work in [8] presents an insight on the use of advanced image features and object/image categorization methods for weed recognition. However, the proposed color-based image features are not applicable in all type of weed images due to segmentation preprocessing. While focusing on image features, other important categorization parameters such as feature encoding, image representation and classification, where not completely studied thus adversely affecting recognition performance.

III. Proposed weed recognition system

Our proposed system, implements a framework inspired from state-of-the-art object/image classification algorithms. These algorithms use a series of individual processes, leading to a decision for each individual image. This series of processes include feature extraction, feature encoding and representation, concluding with the classification stage, as illustrated in Fig. 2. In the ensuing sections we present each process in our novel weed recognition framework.

A. Feature detection and extraction

Feature extraction can be considered as an early image representation. That is because each feature, describes an image region, termed image patch. These features are extracted from image patches that are either detected from a feature detection technique or by using a dense grid.

Feature detection is a process that retrieves regions robust to image scaling, rotation, distortion and illumination changes. As described earlier, authors in [8] conducted experiments on multiple feature extraction methods and their combination with feature descriptors. The top performing feature detection

[1]As part of this work, we have made publicly available our new dataset at: https://github.com/tkounalakis/weed-recognition

methods of [8] are also implemented and compared using our framework.

Object/image recognition algorithms compute image features over a dense grid of predefined image patches covering the whole image. The resulting image representation collects all available visual information, thus describing the whole image and not some part of it. Due to the lack of the excess computational complexity for feature detection, image features are computed much faster. Our novel framework implements this technique that to our knowledge, was never introduced in weed recognition.

B. Image representation and feature encoding

As described in the previous section, an image feature can be considered as a first-stage image representation. However, the discriminatory capabilities of feature individual vectors adversely affect their representation. A solution was provided by popular image representation methods, such as Bag-of-Features (BoF) model [18] and Spatial Pyramid Matching (SPM) [16], which are able to combine a set of image features. BoF and SPM use a vector quantization algorithm (VQ) to encode each image feature into a discrete visual word.

Feature encoding has the property of representing each individual feature vector in respect to a collection of codewords, i.e., a collection of representative feature vectors termed codebook. The resulting encoding vector is more discriminative than the feature vector from which it originates. The collection of visual words, results in a BoF histogram that represents the whole image. Due to the use of VQ, k-d tree encoding representation histograms can only be classified using a nonlinear classifier such as nonlinear SVM. Nonlinear classifiers require a $O(n^2 \sim n^3)$ computation complexity, n being the number of training images. However, even with advanced nonlinear classifiers and kernels, the recognition performance is adversely affected from nonlinearity.

Exploiting the advances in image representation and encoding, we implement two state-of-the-art linear feature encoding methods and classification algorithms. To our knowledge, this is the first implementation of linear encoding methods in the research topic of weed recognition. We implement the known Sparse coding Spatial Pyramid Matching (ScSPM) [9] which relaxes restrictive cardinality of VQ and regulates the sparsity of visual quantization by introducing an ℓ_1 regularization term. The resulting feature encodings favor sparsity, thus capturing salient image properties. Our novel framework can also implement the Locality-constrained Linear Coding (LLC). This method introduce a locality constraint which projects each feature vector into its local coordinate system. By doing so, LLC favors locality rather than sparsity, thus achieving better feature reconstruction, i.e., more efficient feature encoding leading to improved discrimination.

The aforementioned linear encoding methods can be combined with supervised learning linear classifiers [19]. The benefit from using these classifier lies in the simplification of computation, where linear classifiers require only $O(n)$ of computational complexity in their training phase. We show the implementation of linear feature encoding and linear machine learning techniques that result in weed recognition systems, as well as their contribution to recognition performance in the ensuing Section V.

Fig. 3. Example images from the *Rumex 100* weed recognition dataset.

IV. Rumex 100 weed recognition dataset

Most weed detection/recognition systems found in literature are specialized for and applied on very specific plantation or weed types. Also, a usual practice in such systems is a preliminary segmentation process. This kind of initialization using weed/background segmentation is possible only if the considered images allow for it. That is, if the Field of View (FoV) is very narrow, or if they depict isolated green plants on brown soil as background. As mentioned earlier, the lack of challenging and realistic, common benchmarks in weed detection does not favor comparisons, resulting in non reproducible results of weed detection and recognition.

In this paper, we introduce a weed detection and recognition dataset in order to evaluate our proposed experimental procedure. The dataset is compiled from 100 images depicting the Broad-leaved dock (Rumex obtusifolius L.). Broad-leaved dock is a very common weed found throughout Europe, North America and Oceania. This type of weed is harmful both in crops and dairy farming.

Images consisting our dataset are taken from multiple angles, with most of them being captured from a top-view. All pictures were captured outdoors using a high-resolution camera, under various natural illumination conditions. As seen in Fig 3, our dataset consists only of images depicting the Broad-leaved dock plant in its natural surrounding environment, making segmentation very difficult.

The classification process has to recognize images depicting the plant from the ones that do not, which is not possible in the default format of our database. Therefore, we have divided each initial image into a grid of rectangular regions. Each region can be considered as an individual image, which may contain the Broad-leaved dock plant or parts of it. In the case that the above conditions are met the examined region labeled as containing the Broad-leaved dock plant or parts of it. Otherwise, the image is labeled as surrounding environment. As a result, the annotation process provides

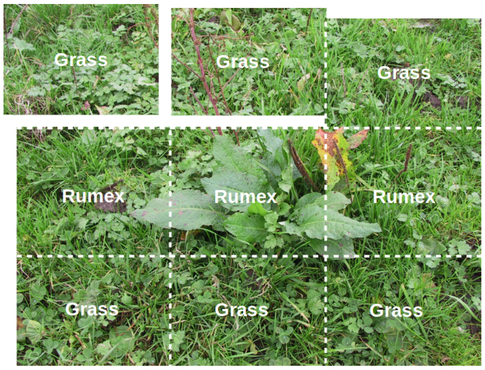

Fig. 4. An example depicting the annotation process. Each examined image is partitioned into a number of regions that are labeled into two classes.

two individual classes that make the classification process possible. The annotation process is done manually, with the user having the choice to decide about the context for each examined region. It has to be noted, that all images are partitioned in the same manner, i.e, the same number of regions, regardless initial image resolution. We consider that the annotation process images segmented into nine regions. An example of the annotation process can be seen in Fig 4. In this way, image annotation will provide regions depicting the surrounding environment but most importantly a central region which is crucial for the proposed dataset. That is because most plants in our dataset are located in the center of each initial image. However, finer annotation grids can be also considered. The annotation process results in a dataset with two classes and 900 images in total that are used in our experimental procedure.

From the above we can conclude that our novel dataset distinguish itself from conventional weed detection and recognition datasets. Henceforth, we denote our novel dataset *Rumex 100*.

V. Experimental results

A. Experimental settings

In order to evaluate our proposed framework and its produced systems, we had to establish a common experimental platform for all experiments. Specifically, we used the experimental framework of [11] with the experimental settings for weed detection proposed in [8].

We use Scale Invariant Feature Transform (SIFT) [12] which is compatible with all the aforementioned feature detection methods. Other features used in [8], like Shape Context (SC) [15] and Color Vegetation Index (CVI) [8] are not expected to perform well in our experiments. These features are computed using shape or color that require an image segmentation. Therefore, these features are not applicable in our dataset due to challenges regarding the depicted environment, i.e, weed plants surrounded by grass.

For feature detection, we use a variety of known methods including, the determinant of the Hessian detector (detHess) [13], difference of Gaussian in multi-scale regions (DoG) [12], Determinant of Hessian for space localization with trace of Laplacian for scale detection (HessLapl) [13] and Harris cornerness measure for space localization with trace of Laplacian for scale detection (HarrLapl) [13]. We also implement two feature extraction over dense image patches, as described in Section III-A. The first implementation describes an extraction of features over a dense grid, and a multi-resolution dense SIFT feature extraction method, the Pyramid Histogram Of visual Words (PHOW) proposed in [11].

For each feature extraction method, we compute an individual codebook consisting of 100 codewords using the k-means algorithm, as proposed in [8]. These codebooks are computed once and remain common for all examined feature encoding methods.

The image representation architecture in Section V-B was set as described in [8], [11]. The binning architecture consists of 20 spatial regions for each image, the result of 2×2 and 4×4 grids. The final image representation histograms are a concatenation of the aforementioned spatial histograms. In Section V-C, we examine the importance of image representations in recognition performance. Hence, each experiment uses different spatial representation as described in that ensuing section.

A compatible machine learning algorithm is combined with the examined feature encoding method in each experiment. For nonlinear image representations a SVM is used combined with a χ^2 homogeneous kernel map [17]. However, when linear separable image representations are used, then classification is performed using a linear classifier from LibLinear library [19]. All classifiers have a fixed cost equal to 10 and the same training/testing image set. The training set consists of 50% of the available image data, i.e, 450 randomly selected images, with the remaining used for testing. All experiments are performed using 10 different randomly selected training/testing set, that remain constant in every experiment.

B. Comparison on feature encoding using multiple feature detection and extraction methods

In order to assess the recognition efficiency of our proposed system, we compare it against the methodology in [8], explained in Section II. In this experiment we assess the representational efficiency of k-d tree nonlinear feature encoding [8], in comparison with ScSPM [9] and LLC [10] linear encoding methods implemented from our novel weed recognition system. Result are presented in Table I.

Regardless feature detection method, k-d tree feature encoding slightly outperforms the ScSPM in every experiment, except the PHOW dence feature extraction. As described in Section III-B, k-d tree computes nonlinear feature encodings that adversely affect the computational and recognition performance. However, k-d tree is able to capture the locality of features. Similar to LLC, this is achieved by representing features from codewords found close in the feature space. As presented in [10], the locality is more important than sparsity when computing feature encodings, a hypothesis also supported from the presented results.

TABLE I
RECOGNITION RATE (%) USING SIFT DESCRIPTOR WITH VARIOUS FEATURE ENCODING AND FEATURE DETECTION METHODS

Feat. Detection Method	Feat. Encoding Method		
	k-d tree	ScSPM	LLC
dense	79.60±1.30	78.83±1.34	82.09±1.18
PHOW	79.69±1.31	**80.93±1.25**	**83.09±1.23**
DoG	**80.40±0.96**	79.18±1.61	80.02±1.34
detHess	79.69±1.22	78.31±1.28	78.33±0.64
HessLapl	78.76±1.37	77.67±1.00	79.56±0.85
HarrLapl	80.02±1.00	79.24±1.02	79.96±1.01

As seen in Table I, the k-d tree feature encoding slightly outperforms the implemented LLC when describing image features using DoG, detHess and HessLapl feature detection methods. The experiments of Table I generally highlight that feature detection is not a strong suit for linear feature encoding methods. The implemented LLC feature encoding achieves to outperform k-d tree when combined with any dense feature extraction method and HarrLapl feature detection. However, the best overall recognition performance is achieved by LLC when using dense PHOW feature extraction. It should be noted, that the proposed in weed recognition, dense feature extraction outperforms every feature detection methods when using linear encoding. The performance of our proposed weed recognition system is favored from linear feature encoding implementation, yielding state-of-the-art results for weed recognition in the examined dataset.

TABLE II
MEAN RECOGNITION RATE (%) OF IMAGE REPRESENTATION ARCHITECTURES USING SIFT DESCRIPTOR AND DOG FEATURE DETECTION.

Im. represenation	k-d tree	ScSPM	LLC
[4 16]	80.40±0.96	79.18±1.61	80.20±1.34
[1 4 9]	79.40±1.38	79.04±1.15	**80.44±1.01**
[1 4 16]	**80.87±0.89**	79.16±1.65	80.29±1.35

TABLE III
MEAN FALSE-POSITIVE RATE (%) OF IMAGE REPRESENTATION ARCHITECTURES USING SIFT DESCRIPTOR AND DOG FEATURE DETECTION

Im. represenation	k-d tree	ScSPM	LLC
[4 16]	6.13±1.00	4.50±1.29	4.23±1.00
[1 4 9]	9.31±1.74	3.98±1.37	**4.59±0.99**
[1 4 16]	**5.89±1.03**	4.32±1.16	4.05±1.11

C. Choosing an image representation architecture

As described in Section II, image representation is an important parameter which is often not properly examined. Table I experimentally shows that the proposed system using linear feature encodings generally outperforms the method in [8]. In Section III-B we described known image representation methods and their combination with feature encoding methods, emphasizing their contribution to performance of image recognition systems. As seen in Tables II and IV, we experimentally assess the importance of image representation architectures in weed recognition.

The first experiment was performed using the DoG feature detector, where the method in [8] achieved its best recognition performance. We use three spatial representations, including the one proposed in [8]. The second, consisting of 3 levels of 1×1, 2×2 and 3×3 spatial cells respectively, resulting in 14 spatial bins. The final image representation architecture consists of 3 levels of 1×1, 2×2 and 4×4 spatial cells, resulting in 21 spatial bins. Table II shows that by changing the representation architecture of [8] the recognition performance of all examined feature encoding methodologies are favored. More importantly, we note that the linear feature encoding methods are only slightly outperformed in recognition. However, due to a more stable false-positive recognition percentage, presented in Table III, we claim that our system outperforms the competing methodology, when system implementation is taken under consideration.

The false-positive measure is inspired from robotic applications, presented in the introductory part and Section II. This measure is used to provide an insight on the efficiency of such systems. A false-positive decision describes an “imaginary success”, meaning that the system not only takes an incorrect decision, but also labels the examined image as the desirable output. In the case of weed recognition, the system recognize weed plant in pictures depicting surrounding environment. In detail, this measure is calculated by finding all images that represent the non desired class, i.e, surrounding environment, in the classification’s testing set. Then we manually count how many of those images were not correctly classified, i.e, described as the desired output class. By dividing the two, we compute the false-positive percentage for each examined experiment. Therefore, false-positive decisions waste system resources, i.e, energy and weed control supplies, a very crucial parameter when taking under consideration system integration. We consider this measure as interrelated with recognition results, thus creating an overall description of system performance.

TABLE IV
MEAN RECOGNITION RATE (%) OF IMAGE REPRESENTATION ARCHITECTURES USING SIFT DESCRIPTOR AND PHOW DENSE FEATURE EXTRACTION.

Im. represenation	k-d tree	ScSPM	LLC
[4 16]	79.69±1.31	80.93±1.25	83.09±1.23
[1 4 9]	**81.31±1.29**	81.56±1.03	**84.96±1.14**
[1 4 16]	81.02±1.21	**81.49±1.29**	84.11±0.88

In order to enhance our weed recognition system’s performance, we also examined the best performing feature extraction method, i.e, the PHOW dense feature extraction. We apply the same changes to image representation, repeating the aforementioned experimental process. The results in Table IV show that more intricate image representations favor recog-

TABLE V
MEAN FALSE-POSITIVE RATE (%) OF IMAGE REPRESENTATION ARCHITECTURES USING SIFT DESCRIPTOR AND PHOW DENSE FEATURE EXTRACTION.

Im. represenation	k-d tree	ScSPM	LLC
[4 16]	9.52±1.52	8.59±1.44	4.07±1.10
[1 4 9]	**9.72±1.12**	6.98±1.89	**4.02±0.90**
[1 4 16]	9.06±1.55	**8.07±1.40**	4.04±1.00

nition performance for both linear feature encoding methods, outperforming the method in [8].

Overall, the implementation of linear feature encoding and improved image representation by our proposed system, yields state-of-the-art results on weed recognition on the examined dataset. The proposed system also achieves less false-positive recognition rates, presented in Table V, thus making it more suitable for implementation in realistic weed control systems.

VI. Conclusion

In this work, we introduce a novel framework which applies known image features combined with advanced linear image representations for weed recognition. We have shown that the architecture of our proposed framework allows a universal application on weed recognition implementations, without any restrictions concerning environmental parameters, plantation or weed type. For this purpose, we compiled a new and challenging image dataset for weed recognition which is made publicly available. We experimentally shown that our system yields high accuracy recognition, but also low false-positive rates making it suitable for implementation on real robotic systems. Our resulting system achieves state-of the-art recognition performance in weed recognition and outperforms other known methods.

Acknowledgment

We want to thank Agroscope (Switzerland) for providing the images consisting the proposed dataset. This work has been supported by the DockWeeder project (project ID: 30079), administered through the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 618123 [ICT-AGRI 2]. The project has received funding from the Ministry of Economic Affairs (The Netherlands), from the Federal Office for Agriculture (Switzerland), and from Innovation Fund Denmark, the Ministry of Science, Innovation and Higher Education (Denmark).

References

[1] A. H. Kargar B. and A. M. Shirzadifar, "Automatic weed detection system and smart herbicide sprayer robot for corn fields," in *Robotics and Mechatronics (ICRoM), 2013 First RSI/ISM International Conference on*, Feb 2013, pp. 468–473.

[2] W. K. Wong, A. Chekima, M. Mariappan, B. Khoo, and M. Nadarajan, "Probabilistic multi svm weed species classification for weed scouting and selective spot weeding," in *Robotics and Manufacturing Automation (ROMA), 2014 IEEE International Symposium on*, Dec 2014, pp. 63–68.

[3] M. Perez-Ortiz, J.M. Pena, P.A. Gutierrez, J. Torres-Sanchez, C. Hervas-Martínez, and F. Lopez-Granados, "A semi-supervised system for weed mapping in sunflower crops using unmanned aerial vehicles and a crop row detection method," *Applied Soft Computing*, vol. 37, pp. 533 – 544, 2015.

[4] María Perez-Ortiz, Jose Manuel Pena, Pedro Antonio Gutierrez, Jorge Torres-Sanchez, Cesar Hervas-Martínez, and Francisca Lopez-Granados, "Selecting patterns and features for between- and within- crop-row weed mapping using uav-imagery," *Expert Systems with Applications*, vol. 47, pp. 85 – 94, 2016.

[5] Garford., "Robocrop inrow," 2011.

[6] Frank Poulsen Engineering., "Mechanical hoeing robot," 2014.

[7] A. Michaels, S. Haug, and A. Albert, "Vision-based high-speed manipulation for robotic ultra-precise weed control," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, Sept 2015, pp. 5498–5505.

[8] Wajahat Kazmi, Francisco Garcia-Ruiz, Jon Nielsen, Jesper Rasmussen, and Hans Jørgen Andersen, "Exploiting affine invariant regions and leaf edge shapes for weed detection," *Computers and Electronics in Agriculture*, vol. 118, pp. 290–299, 2015.

[9] Jianchao Yang, Kai Yu, Yihong Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, June 2009, pp. 1794–1801.

[10] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 3360–3367.

[11] Anna Bosch, Andrew Zisserman, and Xavier Munoz, "Image classification using random forests and ferns," in *IEEE International Conference on Computer Vision (ICCV)*, 2007.

[12] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proceedings of the 7th European Conference on Computer Vision-Part I*, London, UK, UK, 2002, ECCV '02, pp. 128–142, Springer-Verlag.

[14] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346 – 359, 2008, Similarity Matching in Computer Vision and Multimedia.

[15] Serge Belongie, Jitendra Malik, and Jan Puzicha, "Shape context: A new descriptor for shape matching and object recognition," in *In NIPS*, 2000, pp. 831–837.

[16] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, 2006, vol. 2, pp. 2169–2178.

[17] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 3, pp. 480–492, March 2012.

[18] L. Fei-Fei and P. Perona, "A bayesian hierarchical model for learning natural scene categories," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, June 2005, vol. 2, pp. 524–531 vol. 2.

[19] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin, "Liblinear: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, June 2008.