ISPRS Journal of Photogrammetry and Remote Sensing 131 (2017) 65–76

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Unsupervised domain adaptation for early detection of drought stress in hyperspectral images

P. Schmitter [a,*], J. Steinrücken [a], C. Römer [a], A. Ballvora [b], J. Léon [b], U. Rascher [c], L. Plümer [a]

[a] Institute of Geodesy and Geoinformation, Department of Geoinformation, University of Bonn, Meckenheimer Allee 172, 53115 Bonn, Germany
[b] Institute of Crop Science and Resource Conservation, Plant Breeding and Biotechnology, University of Bonn, Katzenburgweg 5, 53115 Bonn, Germany
[c] Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich, Leo-Brandt-Str., 52425 Jülich, Germany

ARTICLE INFO

Article history:
Received 27 May 2017
Received in revised form 25 June 2017
Accepted 6 July 2017

Keywords:
Unsupervised domain adaptation
Machine learning
Support vector machine
Hyper spectral
Agriculture

ABSTRACT

Hyperspectral images can be used to uncover physiological processes in plants if interpreted properly. Machine Learning methods such as Support Vector Machines (SVM) and Random Forests have been applied to estimate development of biomass and detect and predict plant diseases and drought stress. One basic requirement of machine learning implies, that training and testing is done in the same domain and the same distribution. Different genotypes, environmental conditions, illumination and sensors violate this requirement in most practical circumstances. Here, we present an approach, which enables the detection of physiological processes by transferring the prior knowledge within an existing model into a related target domain, where no label information is available. We propose a two-step transformation of the target features, which enables a direct application of an existing model. The transformation is evaluated by an objective function including additional prior knowledge about classification and physiological processes in plants. We have applied the approach to three sets of hyperspectral images, which were acquired with different plant species in different environments observed with different sensors. It is shown, that a classification model, derived on one of the sets, delivers satisfying classification results on the transformed features of the other data sets. Furthermore, in all cases early non-invasive detection of drought stress was possible.

1. Introduction

In recent years, several research groups have successfully demonstrated the detection of plants' physiological processes (e.g. plant stress) from hyperspectral images by using (supervised) methods of Machine Learning (e.g. Karimi et al. (2006), Mucherino et al. (2009), Römer et al. (2011)). Typically, each approach has been developed from scratch, i.e. a new model has been derived from data, which were measured in a specific experiment. However, considering the effort of labelling training data, the question arises how far a model is transferable to other data sets obtained by different sensors observing different plants in different environments.

Generally, most methods of Machine Learning are based on the fundamental assumption that training and test data have the same underlying feature space and distribution (Pan and Yang, 2010). In most real world applications this constraint is violated by varying measuring setup, different sensors or changing environment (e.g. illumination or background).

Strategies of reusing knowledge from a *source domain* (consisting of a feature space and feature distribution) and a *source task* (consisting of a label space and a predictive function) for learning a predictive function in a *target domain* with a *target task* are addressed in *Transfer Learning* (Pan and Yang, 2010). Subject to different settings, specific subcategories have been identified. In *Transductive Transfer Learning* the domains of source and target are different, whereas the tasks are the same. Labelled data are only available in the source domain. Different domains are caused either by different feature spaces or by different feature distributions. The latter case is also known as *Domain Adaptation* (Arnold et al., 2007).

In this paper we propose a Domain Adaptation approach for the detection of water limitation based on hyperspectral images. We focus on stress responses which

- are in early states and cannot be perceived by the naked eye,
- progress continuously and can be characterised by a set of ordinal, subsequent stages.

* Corresponding author.
E-mail address: schmitter@igg.uni-bonn.de (P. Schmitter).

http://dx.doi.org/10.1016/j.isprsjprs.2017.07.003
0924-2716/

The detection of such physiological processes is challenging as the hyperspectral data sets are influenced by the environmental factors of real world applications, and the invisibility of processes prevents a labelling.

Based on the hypothesis that this knowledge allows the reuse of an existing classification model in a different target domain, we propose a transformation of the features of the target domain and formulate an objective function, which enables an evaluation of the transformation parameters without using labelled data in the target domain. The objective function is based on general characteristics about classification and the biological knowledge about the ordinal scale of the process.

In this study we focus on the automated detection of drought stress induced changes in leaf pigments and leaf structure. Drought stress, caused by water scarcity, is one of the biggest challenges in global crop production and it has been estimated that drought can cause a depreciation of crop yield up to 70% in conjunction with other abiotic stresses (Pennisi, 2008; Tuberosa and Salvi, 2006). Prolonged water scarcity initiates invisible biochemical and physiological processes and, subsequently leads to an impairment of crop growth and yield.

The recent rapid developments of destructive and non-destructive technologies offer new opportunities for an effective and high-throughput analysis of plant characteristics. The connectivity and flow of these data, towards molecular breeding and farming, has been hampered by a bottleneck at the level of phenotyping, the so called phenomics bottleneck (Tardieu and Schurr, 2009). Effective use of sensors could contribute towards a pinpoint accuracy in phenotyping, reduced experimental requirements and will enable multiple, simultaneous and objective data to be collected.

Drought is a spatial-temporal process, which triggers various reactions within the plants. As a first intermittent reaction many plants close their leaf stomata, limiting the transpiration loss in periods of limited soil water availability. This physiological reaction normally cannot be detected by hyperspectral imaging as no plant pigments are affected. If water however remains a limiting resource for a longer time period, plants react by changing growth patterns and resource allocation in their organs. Leaf pigments may be broken down, leaf surface may be come pigmented and leaf anatomy changes. These responses however are very variable. Depending of the ecological niche of each species and the developmental stage plants may react very different to a limitation of water. Often drought responses proceed from older to younger leaves and, within a leaf, from the tip towards the leaf base.

The derivation of a source model requires labelled data for the invisible stages of drought stress. However, while it is quite easy to manually identify perfectly healthy and dead leaf pigments, it is not possible to visually grade the stage of senescence for presymptomatic stress detection. Therefore, while the change from healthy to senescent pigments is a continuous process, it is not possible to manually get continuous labels from the image alone. Hence, either an unsupervised regression model or a discretization into ordinal classes are feasible options. Römer et al. (2012) have presented an unsupervised regression method for early detection of drought induced stress with hyperspectral images based on cluster analysis. However, as described in detail later in this manuscript, the mightiness of classes from similar spectra is an important optimization criteria. Therefore, discrete ordinal classes as used in Behmann et al. (2014b) are preferable for the presented method. In Behmann et al. (2014a) the authors show that this ordinal classification approach outperforms non-linear regression with regard to plant stress detection in hyperspectral data. This multi-class model is based on the previous knowledge that process the of the drought induced stress forms an ordinal order, mainly related to chlorophyll degradation (Merzlyak et al., 1999). The model classifies each pixel into a drought class, following an ordinal scale from healthy to stressed. The drought states are a discrete representation of the progressive process of the drought induced stress. The model provides a description of the drought state of a plant; the relative frequencies of the drought classes enable an early detection of drought induced stress (Behmann et al., 2014b).

Furthermore, we assume, that the spectral information of the hyperspectral images is adequately represented by a set of commonly used Vegetation Indices (VI, Table 1). However, although the indices are usually invariant to changes in the environment (Jensen, 2009), a simple reuse of the model does not provide sufficient results (Fig. 1).

The contribution of this paper is an unsupervised method, which allows the reuse of a classification model for the early detection of physiological plant processes from hyperspectral images. The method is characterized by

- the handling of data without any labels in the target domain, as well as without the need for labels at training time,
- the transformation of the feature space of the target domain,
- the use of biological knowledge for evaluating the transformation parameters, and
- the application of a source model which is not changed in the target domain.

We demonstrate the applicability of the proposed approach on three sets of hyperspectral images. Two sets have been measured in drought stress experiments on single barley plants. The experiments were set up in two consecutive years in *foliar tunnels* in Germany. The third data set was collected on maize, grown up in *the field* under different treatments in Italy. It is obvious that the transfer from barley in foliar tunnels in Germany to maize on the field in Italy is rather challenging. However, we will show that an ordinal classification model, derived from one of the sets, delivers satisfying classification results on the transformed features of the other data sets and that the classification result enables an early detection of drought effects.

2. Related work

Classic machine learning methods assume the same domain $\mathscr{D}$ and the same learning task $\mathscr{T}$ for training and test data. A domain consists of a feature space $\mathscr{X}$ and a probability distribution $P(X)$: $\mathscr{D} = \{\mathscr{X}, P(X)\}$. For a given domain, a task consists of a label space $\mathscr{Y}$ and a predictive function $f(\cdot)$: $\mathscr{T} = \{\mathscr{Y}, f(\cdot)\}$ (Pan and Yang, 2010).

Transfer Learning differentiates between a source domain $\mathscr{D}_{\mathscr{S}} = \{\mathscr{X}_{\mathscr{S}}, P(X_S)\}$ with a source task $\mathscr{T}_{\mathscr{S}} = \{\mathscr{Y}_{\mathscr{S}}, f_S(\cdot)\}$, and a target domain $\mathscr{D}_{\mathscr{T}} = \{\mathscr{X}_{\mathscr{T}}, P(X_T)\}$ with a target task $\mathscr{T}_{\mathscr{T}} = \{\mathscr{Y}_{\mathscr{T}}, f_T(\cdot)\}$, where $\mathscr{D}_{\mathscr{S}} \neq \mathscr{D}_{\mathscr{T}}$, or $\mathscr{T}_{\mathscr{S}} \neq \mathscr{T}_{\mathscr{T}}$ (Pan and Yang, 2010). The target predictive function $f_T(\cdot)$ is to be learned by reusing knowledge from $\mathscr{D}_{\mathscr{S}}$ and $\mathscr{T}_{\mathscr{S}}$.

Subject to the availability of labels and different relations between the source and target domain and the source and the target task, specific subcategories of Transfer Learning have been identified (for a full taxonomy see, e.g. Pan and Yang, 2010). Problems in which labelled data are only available in the source domain are addressed in Transductive Transfer Learning. Furthermore, the source and the target domain are different ($\mathscr{D}_{\mathscr{S}} \neq \mathscr{D}_{\mathscr{T}}$), while the source task and the target task are the same ($\mathscr{T}_{\mathscr{S}} = \mathscr{T}_{\mathscr{T}}$). The different domains are caused either by different feature spaces ($\mathscr{X}_{\mathscr{S}} \neq \mathscr{X}_{\mathscr{T}}$) or by different feature distributions ($\mathscr{P}(\mathscr{X}_{\mathscr{S}}) \neq \mathscr{P}(\mathscr{X}_{\mathscr{T}})$). In this paper, we focus the latter case, which is also known as unsupervised Domain Adaptation.

Table 1
List of the used vegetation indices.

Name	Short	Formula	Reference
Normalized Vegetation Index	NDVI	$\frac{R_{800}-R_{680}}{R_{800}+R_{680}}$	Rouse et al. (1973)
Simple Ratio Index	SR	$\frac{R_{800}}{R_{670}}$	Rouse et al. (1973)
Enhanced Vegetation Index	EVI	$2.5\frac{R_{800}-R_{670}}{R_{800}+6R_{670}-7.5R_{490}+1}$	Huete et al. (1997)
Atmospherically Resistant Vegetation Index	ARVI	$\frac{R_{800}-2(R_{670}-R_{490})}{R_{800}+2R_{670}-R_{490}}$	Kaufman and Tanré (1996)
Sum Green Index	SG	$\frac{1}{n}\sum_{i=500}^{599}R_i$	Gamon and Surfus (1999)
Red Edge NDVI	RENDVI	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}}$	Gitelson and Merzlyak (1994)
Modified Red Edge Simple Ratio Index	mRESR	$\frac{R_{750}-R_{445}}{R_{705}-R_{445}}$	Sims and Gamon (2002)
Modified Red Edge NDVI	mRENDVI	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}-2R_{445}}$	Sims and Gamon (2002)
Vogelmann Red Edge Index 1	VOG1	$\frac{R_{740}}{R_{720}}$	Vogelmann et al. (1993)
Vogelmann Red Edge Index 2	VOG2	$\frac{R_{734}-R_{747}}{R_{715}+R_{726}}$	Vogelmann et al. (1993)
Vogelmann Red Edge Index 3	VOG3	$\frac{R_{734}-R_{747}}{R_{715}+R_{720}}$	Vogelmann et al. (1993)
Red Edge Position Index	REP	$\mathrm{argmax}(R_{690-740})'$	Curran et al. (2001)
Photochemical Reflectance Index	PRI	$\frac{R_{531}-R_{570}}{R_{531}+R_{570}}$	Gamon et al. (1992)
Structure Insensitive Pigment Index	SIPI	$\frac{R_{800}-R_{445}}{R_{800}+R_{680}}$	Penuelas et al. (1995)
Red Green Ratio Index	RGRI	$\frac{\mathrm{Mean}R_{500-600}}{\mathrm{Mean}R_{600-700}}$	Gamon and Surfus (1999)
Plant Senescence Index	PSRI	$\frac{R_{680}-R_{500}}{R_{750}}$	Merzlyak et al. (1999)
Carotenoid Reflectance Index 1	CAR1	$\frac{1}{R_{510}}-\frac{1}{R_{550}}$	Gitelson et al. (2002)
Carotenoid Reflectance Index 2	CAR2	$\frac{1}{R_{510}}-\frac{1}{R_{700}}$	Gitelson et al. (2002)
Anthocyanin Reflectance Index 1	ANTH1	$\frac{1}{R_{550}}-\frac{1}{R_{700}}$	Gitelson et al. (2001)
Anthocyanin Reflectance Index 2	ANTH2	$R_{800}(\frac{1}{R_{550}}-\frac{1}{R_{700}})$	Gitelson et al. (2001)
Datt Index 1	Datt1	$\frac{R_{860}}{R_{550}}$	Datt (1998)
Datt Index 2	Datt2	$\frac{R_{860}}{R_{708}}$	Datt (1998)
Datt Index 3	Datt3	$\frac{R_{860}}{R_{550}R_{708}}$	Datt (1998)

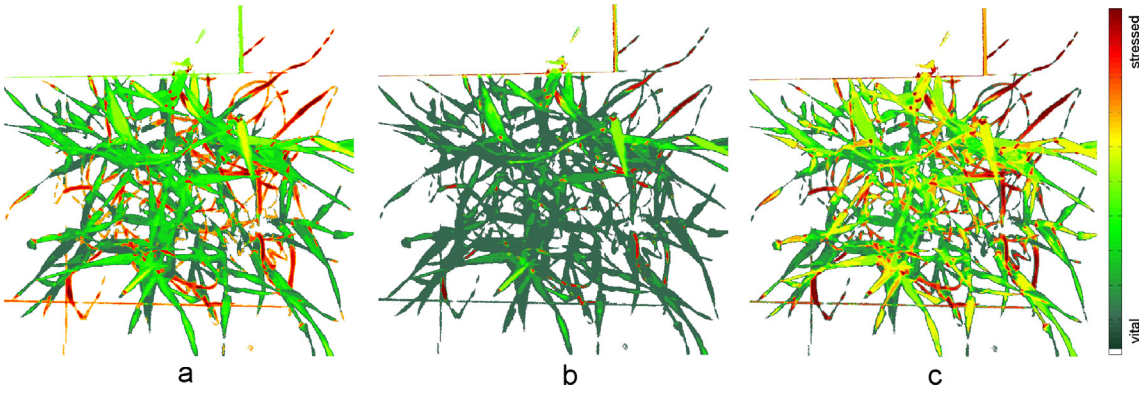

Fig. 1. Classification of barley leafs into senescence classes, ordered from vital to stressed using data from 2010 with a model from 2010 (a). Classification of data from 2010 with a model from 2011 leads to an insufficient result (b) and reclassification by applying the Domain Adaptation approach (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The challenge of unsupervised Domain Adaptation is the lack of labels in the target domain which prevents the evaluation by classic evaluation procedures like a covariance matrix. Previous work in the field of unsupervised Domain Adaptation has used either subspace-based approaches or sample reweighting approaches. Blitzer et al. (2011) proposed a method which projects the features of source and target domain into a subspace where a common linear model is learned. Their approach requires both, unlabelled source and target data at training time.

Huang et al. (2006) proposed an universal method which overcomes slight differences between probability distributions in source and target domain by sample reweighting. The weights for the training points are estimated by matching the training and test points in a reproducing kernel Hilbert space.

Remote sensing of hyperspectral images in general and domain adaptation in particular have been intensively studied for remote sensing, overviews are given in Gómez-Chova et al. (2015) and Tuia et al. (2014).

In remote sensing, Vogt et al. (2017) present a multiple-source selection approach, were labelled data from a subset of domains is combined. The weights for the combination are estimated by boosting. Liu and Li (2014) use boosting to estimate weights to combine features with small divergences in different domains.

An overview of SVM in remote sensing is given by Mountrakis et al. (2011). Bruzzone and Marconcini (2009) use SVM in their proposed Domain Adaptation SVM (DASVM) for multispectral images. They assume that both, source data and target data, are available at training time and adjust the separating hyperplane iteratively by substituting source domain training data with pseudo-labelled target domain data. Items are replaced from the source domain with the highest distance to the hyperplane by pseudo-labelled data from the target domain closest to the hyperplane. For evaluation, they present a circular validation strategy, which subsequently adapts the resulting classification model for the target domain back to the source domain. The problems, difficulties and limits of DASVM are discussed in Bruzzone and

Marconcini (2010). Especially, they deduce the complexity of the problem from the Jensen-Shannon-Divergence. Instead of SVM, Paul et al. (2016) use logistic regression to improve the training and classification performance.

However, the aforementioned approaches are not suitable to our problem due to specific characteristics:

- Only unlabelled target data is available at training time.
- Our source model contains multiple linear predictors.
- We make no assumptions about the distance between the probability distributions of source and target domain.

Instead, we propose a transformation of the target domain features which enables a direct application of the source model. Firstly, we perform a z-score transformation which translates and scales the target features rawly. Secondly, a cubic polynomial is used for a precise adjustment. The second transformation is evaluated by an objective function including, additional prior knowledge about classification and physiological drought processes in plants.

3. Experimental procedure

In this section, we present three data sets of hyperspectral images which are captured from two different plants species in different environments under different conditions (Section 3.1). For each of the data sets we derive a source model (Section 3.2) which will be used in Section 4.

3.1. Experimental setup and hyerspectral images

Three sets of hyperspectral images from different experiments have been used for this study (Fig. 2). Two sets were measured in drought stress experiments on single barley (*Hordeum vulgare*) plants in laboratory, while the third set was measured on maize (*Zea mays*), grown up in the field under different treatments (untreated, irrigated, fertilized, and a combination of irrigated and fertilized).

The first barley experiment was set up in 2010 (Römer et al., 2012). Twelve plants of cultivar "Scarlett" were cultivated in single pots under controlled conditions in the greenhouse. At developmental stage BBCH31[1], which corresponds to the end of tillering and the beginning of the main shoot, the plants were divided into two groups, six plants each. While the first group was well watered, the second one was stressed by a reduced irrigation. The second barley experiment was conducted in 2011 under the same experimental conditions (Römer et al., 2012). Now, twelve plants of cultivars "Wiebke" and "Barke" were grown. At developmental stage BBCH31 they were divided into three groups, each of four plants. Again, one group continued growing under optimal conditions, while the others were drought stressed (reduced watered and unwatered).

In both barley experiments, the plants were observed with the hyperspectral imager SOC-700 from Surface Optics (Ocean Optics, San Diego, CA, USA). The imager has a spatial resolution of 640×640 pixels and a spectral resolution of approximately 4 nm. It measures 120 equally distributed bands in the range of 430–890 nm. The images were recorded in the laboratory under controlled illumination, provided by six 400 W halogen lamps. The sensor and the lamps were arranged above the imaging position of a single pot.

In the 2010 barley experiment, the measurements started two days after water reduction. In nine sessions within the next 30 days a total of 108 images were recorded. In the 2011 barley experiment, the measurements started one day after water reduction and were continued daily for the next 20 days. The observation of the unwatered plants was stopped after eleven days, when the drought symptoms differentiating unwatered plants from the controls. In total, 204 images were recorded in 2011.

The third data set was obtained using a different sensor on a different species (maize) under different illumination conditions in a field experiment in Northern Italy. The maize plants were grown in plots of $15\ m \times 16.5\ m$ under four different treatments: untreated, irrigated, fertilized, and a combination of irrigated and fertilized. The hyperspectral images were acquired July 22nd 2010, 44 days after seeding, from four meters above the canopy; each treatment was observed by three images of different areas of a corresponding plot. In order to attain similar illumination conditions, the measurements were conducted between 10.00 and 14.00 h local time under cloudless sky. The images were recorded by a PS V10E sensor (Spectral Imaging Ltd, Oulu, Finland). They have a spatial resolution of 1392×840 pixels and a spectral resolution of 1040 bands in the range of 400–1000 nm. This instrument was mounted in nadir position in a boom lift at 4 m from top-of-canopy level and moved horizontally (Römer et al., 2012). In total, 12 images were recorded.

The water status of the analysed barley plants was estimated by measuring the water content and water potential at the time point when the hyperspectral images were taken. Variations in leaf water potentials (L, MPa) were measured with a Scholander pressure chamber (Scholander et al., 1965), following the methodological procedures described by Turner (1988). The plant water Chlorophyll concentrations were measured with a SPAD Chlorophyll meter (SPAD 502, Konica Minolta Sensing Europe B.V., MN Nieuwegein, NL) and estimating by using the mean value of 3 measurements per leaf.

Hyperspectral images were radiometrically calibrated by subtracting the dark frame and calculating the relative reflectance by using the ratio to a white reference panel, which was available in each image.

The separation of background pixels and plant pixels was done by learning a SVM model, which has been derived using Active Learning (Lewis and Gale, 1994). Active Learning starts with a small set of training data from what a SVM classifier is learned. The classifier is applied to each hyperspectral signature to estimate the classification probabilities via Platt scaling (Platt, 2000). In each Active Learning iteration the hyperspectral signatures, which are most unlikely, were clustered by using the k-Means algorithm. The resulting centroids ensure a wide range of hyperspectral signatures which could not be reliable assigned to a class, wherefore they were manually labelled by the user. The process is terminated by the user if the classifier is evaluated as confident. This iterative procedure converges quickly and is user friendly since only few data, which obviously differ due to the clustering, has to be labelled (Schmitter et al., 2015).

3.2. Source model

We use an ordinal SVM model as source model. It is based on the previous knowledge that drought induced stress forms an ordinal order, mainly related to chlorophyll degradation (Merzlyak et al., 1999). The model classifies each pixel into a drought class, the entity of classes follows an ordinal scale from vital to stressed.

The required labels were generated by performing the unsupervised clustering algorithm k-Means (MacQueen et al., 1967) on the hyperspectral signatures. The centroids of the resulting clusters are sorted in ascending order from 1 to k (vital to stressed) to represent the order of the drought stress process (Behmann et al., 2014b). For the parameter k, we chose $k = 10$.

[1] The BBCH (Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie) scale is an uniform coding of the growth stages of mono- and dicotyledonous plants (Hess et al., 1997).

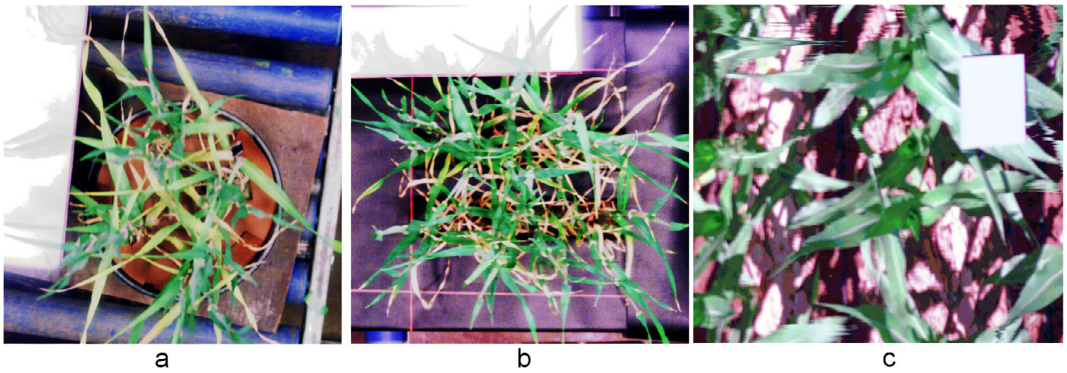

Fig. 2. Representative top-view images of the plants of the three experiments that were used to develop the classification approach. (a) and (b) Potted barley plants that were grown in the greenhouse in 2010 (a) and 2011 (b). (c) Top-of-canopy view of the maize experiment in Italy. Images were acquired at clear days under stable sunlight illumination. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

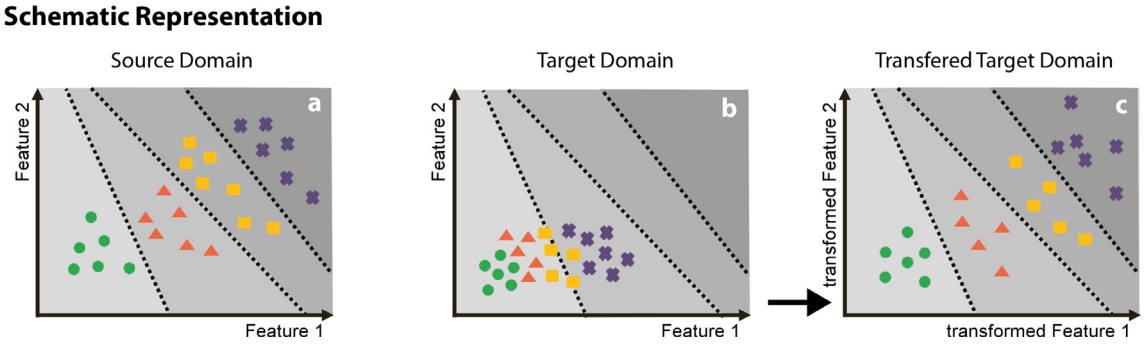

Fig. 3. Schematic representation of the proposed feature transformation approach. Instances of four classes are represented by the shapes. A correct classification is reached if the shapes are correctly separated by the dotted line. Application of the source model to source domain (a), target domain (b) and transformed target domain (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Subsequently, the feature space was reduced by a feature extraction in which each observed spectrum was represented by a set of commonly used Vegetation Indices (Table 1).

On the basis of the vegetation indices and the ranked labels the ordinal SVM was learned. The ordinal structure was accomplished by dividing the multi-class problem in n − 1 binary classification problems (Frank and Hall, 2001). Based on the prior knowledge about the ordinal structure, only the discriminant between neighbouring classes had to be learned.

A schematic representation of a result ordinal classification model is depicted in Fig. 3a.

4. Method

Here we present the Transfer Learning approach for the early detection of drought stress in hyperspectral images. As source model we assume an ordinal SVM model which has been used for the early detection of drought stress of a specific plant in a specific year under a specific setting (Section 3). The source model has to be applied to hyperspectral images from a different year and/or a different plant species in a different setting. Thus, the source task and the target task are the same ($\mathcal{T}_\mathcal{S} = \mathcal{T}_\mathcal{F}$), whereas the source domain and the target domain are different ($\mathcal{D}_\mathcal{S} \neq \mathcal{D}_\mathcal{F}$). The difference is caused either by different feature spaces ($\mathcal{X}_\mathcal{S} \neq \mathcal{X}_\mathcal{F}$) or by different feature distributions ($P(X_S) \neq P(X_T)$). In order to ensure the same feature spaces $\mathcal{X}_\mathcal{S} = \mathcal{X}_\mathcal{F}$, the set of Vegetation Indices (Section 3.2) is also derived for the hyperspectral images of the target domain. However, a straightforward application of the source model to the target domain leads to an insufficient classification result.

We overcome the differences of $P(X_S)$ and $P(X_T)$ by transforming the target domain features (Section 4.1). This procedure is schematically illustrated in Fig. 3. The source domain is successfully classified by the source model (Fig. 3a), however an application of the source model to the target domain is not satisfying. A transformation of the target domain features enables a successfully classification by the source model (Fig. 3c). The transformation parameters are found by an objective function (Section 4.2) in a global optimization procedure (Section 4.3). The results are evaluated by different validation criteria (Section 4.4).

4.1. Feature transformation

The feature transformation consist of two steps. In a first step, the target domain features are rawly translated and scaled to decrease the differences between $P(X_S)$ and $P(X_T)$. We used the z-score transformation to adapt the means and standard deviations of the target domain features to the means and standard deviations of the source domain features. The feature distributions $P(X_S)$ and $P(X_T)$ are still not identical but the difference between them is significantly reduced. The resulting transformed target domain features are the input for the optimization of the feature transformation in the next section. In a second step, a refined transformation by a quadratic polynomial is used for a precise adjustment. The adjustment was calculated for each Vegetation Index separately due to their different calculation formulas and

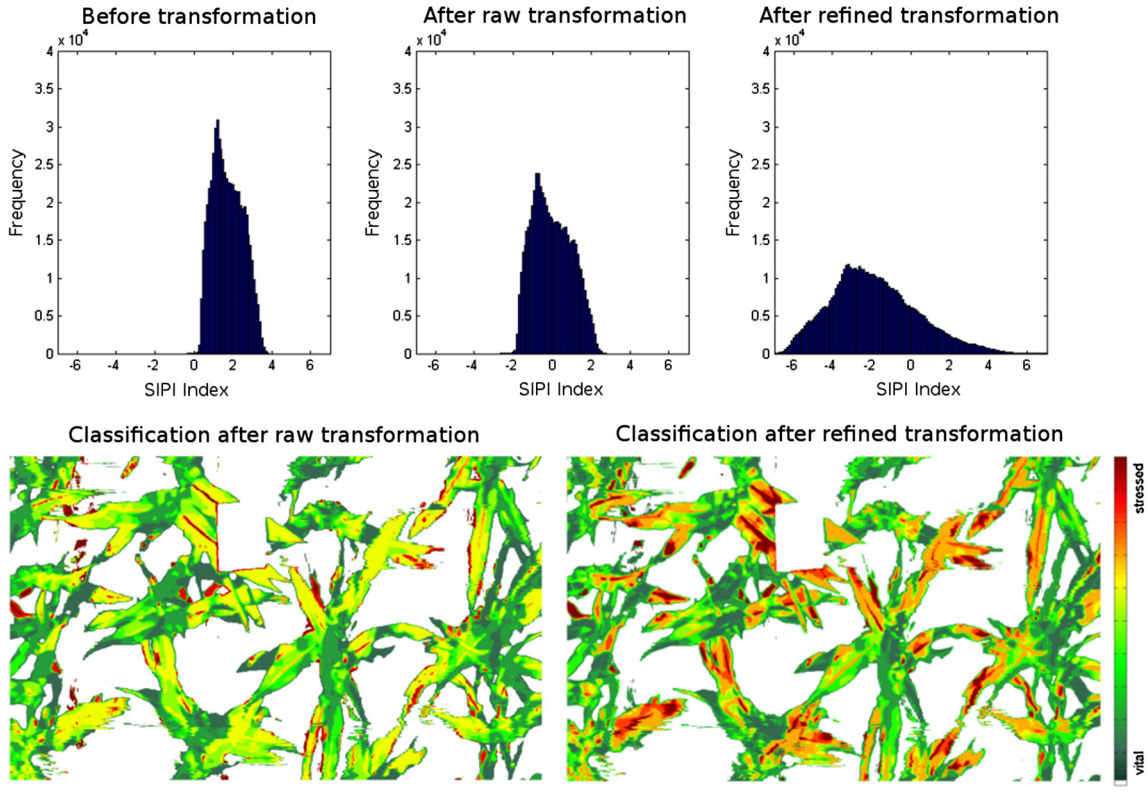

Fig. 4. Impact of the raw (z-score) and refined transformation to the feature distribution using the example of the Structure Insensitive Pigment Index (top row) and the corresponding classification results (bottom row). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the different ranges of the hyperspectral signal from which they are composed.

4.1.1. Refined transformation

Generally, the features do not follow a Gaussian distribution but a skew-symmetric or a multi-modal distribution (Fig. 4). Thus, the remaining differences between source and target domain are adjusted by adapting the target features $\bar{x}_T$, using a cubic transformation:

$$\hat{x}_T = t_2\bar{x}_T^3 + t_1\bar{x}_T + t_0 \tag{1}$$

We have chosen the cubic polynomial as (i) it provides a sufficient degree of freedom to transform the target domain features, and (ii) regards negative values. The transformation parameters $t = \{t_0, t_1, t_2\}$ are determined by an optimisation approach. The objective function is described in Section 4.2, and the optimisation in Section 4.3.

4.2. Objective function

The objective function evaluates the transformation parameters t in order to enable a directed search for the optimal transformation parameters. Common validation strategies for classification and Transfer Learning are cross-validation (Kohavi et al., 1995) or the confusion matrix (Congalton, 1991). These strategies require labelled data to estimate the optimal classifier. However, the proposed setting deals with non-existent labels in the target domain, wherefore these strategies are inapplicable. Instead, we propose an objective function that evaluates previous knowledge about *similarities* in the occurrence of the same process from different plants in different environments under different conditions. The knowledge is represented by four criteria:

- *Mixing* (*Mix*) evaluates a basic characteristic of *classification*: the diversity of the features within a class should be as small as possible.
- *Deviation* (*D*) evaluates a basic characteristic of *Domain Adaptation*: the difference between the distribution of the transformed target features and the distribution of the source features, which should be as small as possible.
- *Smoothness* (*S*): evaluates biological knowledge about the *spatial propagation* of the drought stress process: drought induced stress proceeds smoothly and continuously in every plant. Thus, neighbouring pixels should be assigned to the same or to a neighbouring drought class.
- *Mightiness* (*M*) evaluates an assumption about *similar states* of the drought stress: the distributions of the drought classes in source and target domain should not be too far apart.

The combination of these criteria summarise the prior knowledge and integrates it into the optimisation procedure. Each criterion is assigned with a weight w_i:

$$\min Z(t) = w_1 Mix + w_2 D + w_3 S + w_4 M, \tag{2}$$

where $0 < w_i < 1,\ i = 1,\ldots,4$ and $\sum_{i=1}^{4} w_i = 1$. Whereas distinct values for the different weights would be possible our experiments have suggested that same values ($w_i = 0.25,\ i = 1,\ldots,4$) are both appropriate and robust.

The evaluation criteria are estimated with the transformed target features $\hat{x}_T$ and their feature distribution $P(\hat{X}_T)$. The minimisation of the objective function leads to optimal transformation parameters.

4.2.1. Mixing

Classification aims to group similar data points by minimising the differences of instances within a class and maximising the dif-

ference to instances in other classes. *Mixing* validates the scattering of the instances within a class. In contrast to *smoothness*, which validates the *spatial* distribution of the labels, mixing validates the *diversity* of the features.

For estimating the differences within a class, a k-means clustering is performed for each class. The similarity between the centroids c within a class is estimated by the (Pearson) correlation coefficient r (Murphy, 2012). These coefficients are averaged for each class to evaluate the mixing within a class. The average of all classes is added to the objective function:

$$Mix = \frac{1}{4!k}\sum_{l=1}^{k}\sum_{i=1}^{4}\sum_{j=i+1}^{5} r_{i,j}^{(l)} \tag{3}$$

where $r_{i,j}^{(l)}$ represents the correlation coefficient between the i-th centroid and the j-th centroid within the l-th drought class. In each of the drought classes, the k-Means builds five clusters. The choice of five clusters bases on an empirical determination of the data within the drought classes.

4.2.2. Deviation

A successful classification is based on the assumption that the distribution of train and test data do not differ. Thus, the *deviation* measures the distance between $P(X_S)$ and $P(\widehat{X}_T)$, which has to be minimal.

In order to estimate the deviation, we use the Jensen-Shannon-Divergence (JSD), which is a non-parametric measure to compare distributions. The JSD bases on the Kullback-Leibler-Divergence $D_{KL}(P\|Q)$ (Murphy, 2012):

$$JSD(P\|Q) = \frac{1}{2}D_{KL}(P\|M) + \frac{1}{2}D_{KL}(Q\|M) \tag{4}$$

with

$$M = \frac{1}{2}(P+Q) \tag{5}$$

$D_{KL}(P\|Q)$ is the Kullback-Leibler-Divergence of two discrete distributions P and Q:

$$D_{KL}(P\|Q) = \sum_i log_2\left(\frac{P(i)}{Q(i)}\right)P(i) \tag{6}$$

The deviation D between $P(X_S)$ and $P(\widehat{X}_T)$ is determined by the mean JSD of all distributions $P(x_S^i)$ and $P(\hat{x}_T^i)$:

$$D = \frac{1}{n}\sum_{i=1}^{n} JSD(P(x_S^i)||P(\hat{x}_T^i)) \tag{7}$$

where $P(x^i)$ is the distribution of the i-th feature.

4.2.3. Smoothness

According to the prior knowledge about the propagation of drought induced stress, neighbouring pixels likely refer to the same stage of drought. Smoothness introduces this spatial information into the objective function.

The smoothness is derived from a label image Y_T which is generated by assigning a label to each plant pixel. The labels are estimated by applying the source model to the transformed target features. Y_T should exhibit an ordinal class structure which corresponds to the gradual drought stress propagation. The position and magnitude of transitions between classes within the label images are derived by calculating the gradient image ΔY_T, which detects local changes in an image (Canny, 1986). Background pixels, which were removed during preprocessing, occur in the label image as zeros, to obtain a complete image. Thus the gradient image contains irrelevant transitions from background to plant, which are removed in each iteration of the optimization procedure.

The evaluation of the spatial smoothness is based on the relative frequencies $h(\Delta Y_T)$ of the magnitudes in the gradient image. Transition between classes which differ widely are not desirable for what reason each relative frequency is weighted by its strength i:

$$h_w(\Delta Y_{T_i}) = h(\Delta Y_{T_i})i,\ i = 1,\ldots,s \tag{8}$$

where s is the maximum magnitude. Thus, small magnitudes contribute less to the criterion than higher magnitudes. Enabling a comparable criterion, the weighted frequencies are normalised:

$$\bar{h}_w(\Delta Y_T) = \frac{h_w(\Delta Y_T)}{|h_w(\Delta Y_T)|} \tag{9}$$

The knowledge about the drought stress process expects small magnitudes. Thus, the spacial smoothness S is integrated into the objective function by only regarding magnitudes greater than two:

$$S = \sum_{i=3}^{s}\bar{h}_w(\Delta Y_{T_i}) \tag{10}$$

4.2.4. Mightiness

The propagation of the drought induced stress and the drought stress status is similar albeit not equal in both domains. Therefore, the distribution of the drought classes should be distributed similarly in both domains. Care must be taken to reflect this assumption in the formula appropriately. If there is a substantial difference between source and target domain, as in the case of barley and maize, hyperspectral responses to drought stress are rather different, caused by the different pigment structures in different species. Even after refined transformation as described in 4.1.1, labels tend to cluster in one class rather than distributing over all classes. The task of the criterion mightiness is to enforce this spreading. Although it seems counter-intuitive at first glance this is best achieved by assuming that the class distributions are similar in source and target domain, as long as no other information (that for instance one species or variety might be more resistant then the other) is available.

The similarity of the class distributions is assessed by the relative class frequencies of source $h(y_S)$ and target domain $h(y_T)$. The deviation of these two class distributions is used to evaluate the mightiness M:

$$M = |h(y_T) - h(y_S)| + \epsilon \tag{11}$$

All classes are supposed to occur in the classification and a sufficient number of items should be assigned to each class. Small or empty classes should be avoided, wherefore a penalty term ϵ is added for each class which contains less than 1% of the instances. Furthermore, the penalty is added, if a class contains a significant larger number (40%) of instances than the other classes. These thresholds for the penalty term were chosen on the basis of a ten class problem ($k = 10$). The penalty term raises for each class which violates the mentioned regulations. This penalty term also ensures that the objective function convergences to a solution, in which all classes occur.

All terms, except the penalty term, are normalised to $[0, 1]$. This allows a comparison of the influences of the different criteria to the objective function.

4.3. Optimization algorithm

The optimal parameters for the transformation of the target domain features are found by minimising the proposed objective function (Eq. (2)). For optimising the parameters an optimisation method has to be chosen according to characteristics of the objective function. In our case, the features follow skew-symmetric or

multi-modal distributions and the terms of the objective function have conflicting goals. Furthermore, the objective function is discontinuous (and not differentiable), as it is updated by using a classification within each optimization iteration. Thus, no Hessian-matrix is computable, wherefore a gradient-free optimization method is required. Another requirement is that three transformation parameters are needed for each feature, which leads to a high number of parameters. Since no assumptions about a unique local optimum are present, a global optimization procedure is desirable.

Algorithm 1. Optimization of the transformation parameters.

Input: source domain data x_S, target domain data x_T, source model m, annealing schedule $Temp(x)$
Output: transformation parameters t
(1) Similarity transformation
(2) $t \leftarrow t^0$ Initial transformation parameter
while $Z_{new} > 0.01$ & $k < k_{max}$ **do**
(3) $t^{new} \leftarrow$ select transformation parameters
(a) $\hat{x}_T = t_2^{new} x_T^2 + t_1^{new} x_T + t_0^{new} \leftarrow$ transform target features
(b) Apply classification model m to transformed features $\hat{x}_T$
(c) Evaluate the transformation parameter via the objective function
$Z_{new}(t^{new}) = w_1 D + w_2 S + w_3 M + w_4 Mix$
(4) $\Delta Z = Z_{new}(t^{new}) - Z(t)$
(5) $T \leftarrow$ schedule $Temp(k/k_{max})$
if $\Delta Z > 0$ **then**
(6) $t \leftarrow t^{new}$ accept new parameters
else
(7) $t \leftarrow t^{new}$ accept new parameters with probability $e^{\Delta Z/T}$
end if
end while

We used the well known global optimization algorithm *Simulated Annealing* (Kirkpatrick et al., 1983), which minimises a given objective function without the need of derivations. Besides, this method is likely to find near-optimal values (local optima not far from the global optimum) for a large amount ($23 \times 3 = 69$ parameters) of parameters under adversary conditions. The algorithm adapts the controlled cooling of heated material to achieve a low-energy state. In each step, the algorithm samples new transformation parameters by altering the actual parameters uniformly with a given length. The decrease of the length during the optimization process, enables large alternation at the beginning and smaller alternation at the end. The objective function is evaluated for each trial of new parameters and their evaluation value is compared to the evaluation value of the current parameters. If the new parameters have a lower evaluation value they are always accepted. Otherwise, the new parameters are only accepted with a certain probability which can exceed local minima. The probability of accepting alternated parameters decreases during the optimization process like the decreasing temperature of the cooling of heated material. The decreasing probability of accepting parameter with higher evaluation values and the shrinking range where possible new parameters are sampled enables the optimization procedure to find and converge to a global minimum.

4.4. Evaluation of ordinal classification

After the optimization of the transformation parameters the resulting classification, of the transformed target domain by the source model, needs to be evaluated. The evaluation has to cover aspects of the classification and the ordinal class structure. Common metrics to assess the performance of a classification model (e.g. overall accuracy) do not cover all aspects of ordinal classification. The overall accuracy validates the quality of a classification in which every misclassification is treated equally. However, the ordinal structure of the model and the distances between "true" and "predicted" classes are not considered. These two criteria are included into the evaluation, since the drought states are sorted and the state "vital" is assigned to the first class and the drought state "stressed" is assigned to the last class. The following criteria cover these aspects and were used to validate the results of the proposed method:

- Overall Accuracy (O.A.)
- Root-mean-squared error (RMSE)
- Spearman's rank correlation coefficient (Spearman, 1904)

The RMSE is a common measure to estimate the mean difference between the "true" and "predicted" classes. The class ranking is also used for the validation of the ordinal structure by the Spearman's rank correlation coefficient $-1 \leqslant r_S \leqslant 1$ (Cardoso and Sousa, 2011). The results of the proposed method are evaluated by all three criteria, since they evaluate different aspects: O.A. evaluates the correct classified data, MSE evaluates the mean difference and r_S evaluates the ordinal structure.

5. Results and discussion

We demonstrate the applicability of the proposed transformations by using the three data sets described in Section 3. From each of the data sets, we have selected a representative hyperspectral image out of the reduced watered group, in which the advanced stages of drought effects were already visible.

For each of the three selected images, a source model (Section 3.2) has been learned on the basis of randomly selected labelled pixels (unsupervised labelling in Section 3.2). Subsequently, each source model was applied to each of the two remaining hyperspectral images (target domains) by using our approach described in Sections 4.1 and 4.2. To reveal the improvement of the proposed method the source model has also been applied to the untransformed features of the target domains.

For the evaluation of our method each source model has been applied to its own domain, the experiment the model is learned on, to obtain labels. These labels are compared (i) to the results of the application of the source model to the untransformed features, and (ii) to the results of the proposed method. These comparisons are based on the criteria for validating ordinal classification results introduced in (Section 4.4). As all labels are estimations which are based on a discretization of a continuous process, we primary consider improvements which are achieved by the approach.

Besides, we applied the optimised transformation parameters to the series of images and analysed the target domains with the source models in order to examine if an early detection of drought stress is possible (Section 5.2).

5.1. Domain adaptation

In Fig. 5, we exemplify the results from $\mathscr{D}_{\mathscr{S}} = Barley2010$ and $\mathscr{D}_{\mathscr{T}} = Maize$. This is the most challenging adaptation since the measurement setting and the plants greatly differ in the experiments. Due to these differences, the application of the source model to the untransformed target features leads to a classification result where most of the instances have been assigned to a healthy class, while some inter-classes are under-represented (Fig. 5b). In

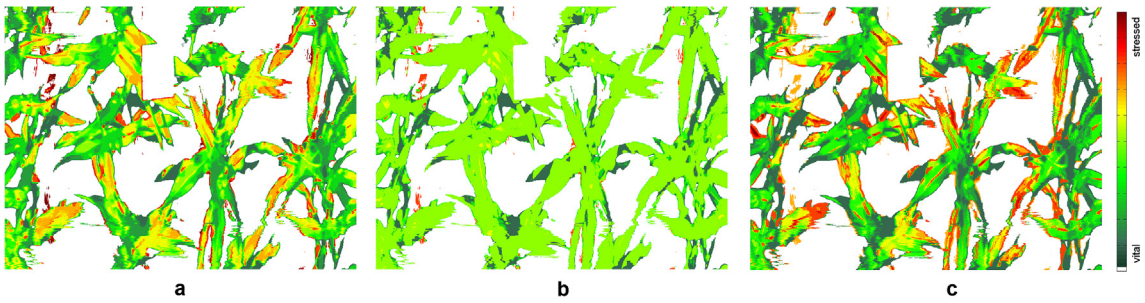

Fig. 5. Classification result of the maize data with the maize model (a). Results of the application of the barley model for 2010 to the untransformed maize features (b) and the transformed features (c). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Results of the model transfer: application of the source model to the original target features data and the transformed target features.

Domain		Untransformed			Transformed		
Source	Target	O.A.	r_S	RMSE	O.A.	r_S	RMSE
Barley2010	Barley2011	0.25	0.78	2.14	0.36	0.89	1.36
Barley2010	Maize	0.23	0.60	1.75	0.43	0.86	1.39
Barley2011	Barley2010	0.23	0.49	3.05	0.31	0.65	2.25
Barley2011	Maize	0.13	0.89	2.31	0.26	0.77	1.87
Maize	Barley2010	0.20	0.40	3.47	0.28	0.45	2.49
Maize	Barley2011	0.20	0.83	2.80	0.40	0.88	1.32

general, the classification result of the transformed features (Fig. 5c) corresponds to the biological characteristics. Thus, the feature transformation improves the classification result significantly, which is rather astonishing taking the substantial differences between the different scenarios (barley vs. maize, Germany vs. Italy, foliar tunnels vs. open field) into account.

We could confirm the visual impression by estimating the listed evaluation criteria for ordinal classification (Section 4.4). The values for the comparison of the untransformed classification result with the results of the source model are shown in Table 2. The RSME of the depicted label image (Fig. 5) is 1.75 and decreases to 1.39 if we apply the proposed method (Table 2, row 2). Since this metric evaluates the values of the differences between expected and actual classification result, the classification could be improved. Furthermore the ordinal structure of this example could be improved, since r_S and the overall accuracy could be increased by transforming the features.

The comparison of all transfer examples in Table 2 shows that the overall accuracy could be increased in all cases. Furthermore, the RMSE is reduced for all examples. Spearman's rank correlation coefficient r_S increases in all cases except the transfer from $\mathscr{D}_{\mathscr{S}} = Barley2010$ to $\mathscr{D}_{\mathscr{T}} = Maize$ (Table 2, rows 4). However, these measures show that our approach improves the classification result w.r.t. the ordinal structure.

5.2. Early detection of drought stress

In this section, we show the use of results of the for the early detection of drought effects. In order to draw conclusions about this development, the separation of the treatments (e.g. reduced watered and well-watered for barley in 2010) within the *series of hyperspectral images* must be regarded. Therefore, the health states of the individual plants are described by summarising the classification result for each image within a histogram, containing the relative frequencies of classes. For more details see Behmann et al. (2014b).

5.2.1. Drought stress in barley

The relative class frequencies of each image are used to estimate the stress state of the plant. Therefore a linear SVM model was learned on basis of the histograms of a vital plant at the first observation day and a reduced watered plant at the last observation day. The stress status of a plant was estimated by the SVM output, which is the normalised distance to the discrimination function. The mean SVM scores and the standard deviation are shown in Figs. 6 and 7 for each treatment. The normalised distances to the discrimination functions are used for a one-way ANOVA test which checks the separability of the different treatments with a significance level of 5%.

Pigment degradation can originate from drought induced stress or occurs naturally during leaf senescence. Both processes can be identified in Fig. 6. The degradation from senescence causes a rise of the degraded pigments in the reduced watered (red) and the well-watered (green) plants. Drought causes an additional increase of degraded pigments in the reduced watered plants (red in Fig. 6). SPAD measurements confirmed the constant decrease in the chlorophyll content in the droughted plants with faster rate than in the control ones (data not shown). Nevertheless, no inflexion point could be seen in the curve of SPAD value data. In Fig. 7 the increasing amount of degraded pigments is only affected by drought stress, since the amount of degraded pigments does not raise in the well-watered (green) plants.

The source model $\mathscr{D}_{\mathscr{S}} = Barley2011$ enables a separation of the treatments at day 8 after drought induction. The transfer from the $\mathscr{D}_{\mathscr{S}} = Maize$ to $\mathscr{D}_{\mathscr{S}} = Barley2010$ enables a separation of the treatments at day 14 after drought induction (Fig. 6 right). However, in both cases, a detection of drought effects is possible, whereas Table 2 indicates an usage of the source model from barley 2011.

In the 2011 data, source data $\mathscr{D}_{\mathscr{S}} = Barley2010$ and $\mathscr{D}_{\mathscr{S}} = Maize$ have been used (Fig. 7 left and right, respectively). In the first days all treatments develop similar. The separation of the well watered and unwatered treatments is possible from day 9 (barley source model) and day 10 (maize source model) after drought induction (Table 3). A first separation of the reduced-watered and fully watered plants is possible at day 11 with the barley source model. A reliable separation is possible from day 15 (barley source model) and day 16 (maize source model).

The separation of the different treatments in the barley data from 2010 with $\mathscr{D}_{\mathscr{S}} = Barley2011$ is possible on the same day as in Behmann et al. (2014b) and one observation day earlier as in

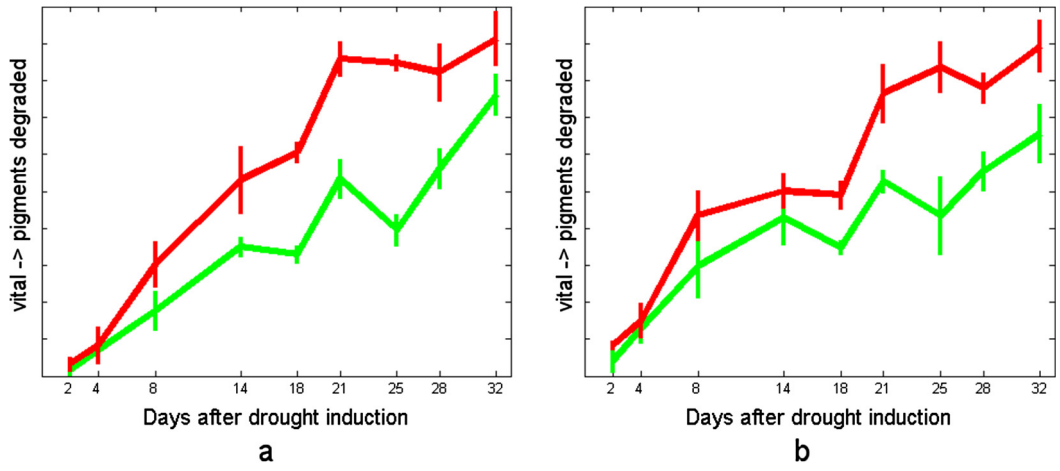

Fig. 6. Result of the histogram classification of the barley data 2010 with the source model from barley 2011 (a) and the maize model (b). Well-watered (green) plants are separable form the reduced watered (red) plants. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

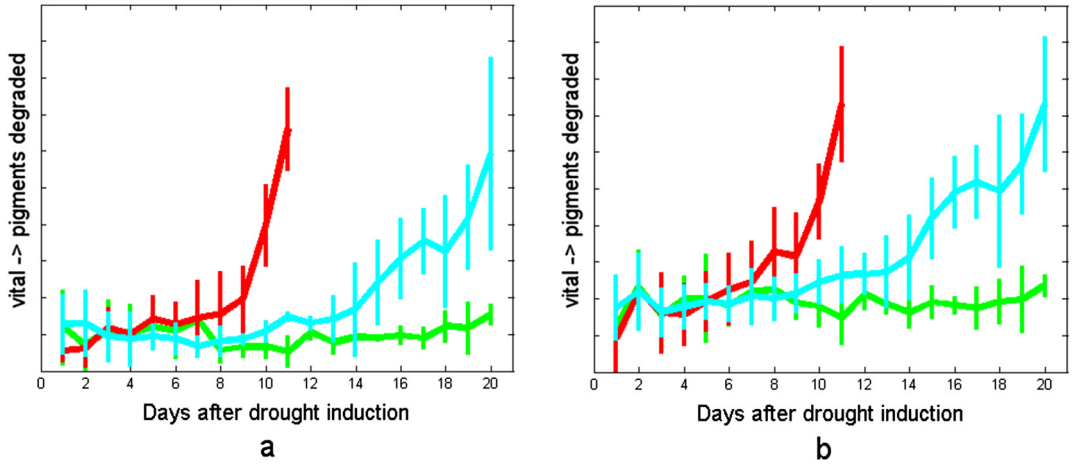

Fig. 7. Result of the histogram classification of the barley data 2011 with the source model from barley 2010 (left) and the maize model (right). Well-watered plants (green) are separable form the reduced watered plants (cyan) and the unwatered plants (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Results of the one-way ANOVA-test for the barley data from 2011. The resulting P-values show whether the treatments are separable or not.

Source domain	Barley2010	Maize	Barley2010	Maize
Day	Reduced-watered/fully watered		No water/fully watered	
7	0.16	0.53	0.83	0.61
8	0.47	0.92	0.15	0.13
9	0.45	0.32	0.02	0.06
10	0.11	0.16	0.00	0.00
11	0.02	0.27	0.00	0.00
12	0.32	0.12	–	–
13	0.19	0.10	–	–
14	0.16	0.04	–	–
15	0.06	0.02	–	–
16	0.00	0.00	–	–

Römer et al. (2012). $\mathscr{D}_{\mathscr{S}} = Maize$ enabled a separation of the barley data from 2010 on the same day than in Römer et al. (2012) and one observation day later than Behmann et al. (2014b). For the barley data set from 2011 the no watered group was separable from the fully watered with $\mathscr{D}_{\mathscr{S}} = Barley2010$ on the same day, and with $\mathscr{D}_{\mathscr{S}} = Maize$ on day later as in Römer et al. (2012). The separation of the treatments, reduced watered and the fully watered, with $\mathscr{D}_{\mathscr{S}} = Maize$ was possible two days later and with $\mathscr{D}_{\mathscr{S}} = Barley2010$ three days later than in Römer et al. (2012).

The analysis of water status in Barley 2011 experiments revealed that the water potential increased at second time-point of measurement for the unwatered variant and at fourth time-

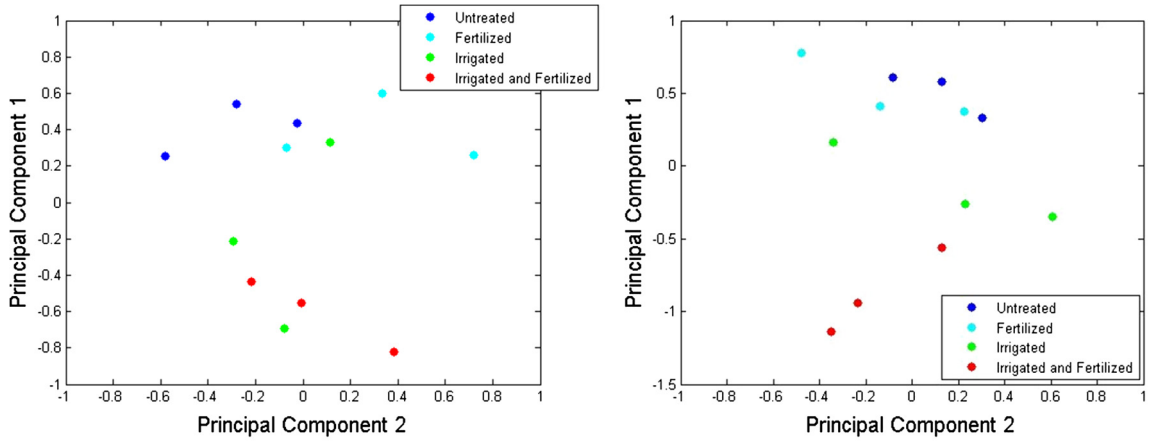

Fig. 8. Resulting first two components of the maize data set grouped by their treatment. The treatments are differentiable by using the barley data from 2010 as source domain (left) and by using the barley data from 2011 as source domain (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

point for the variant with reduced watering, whereas for the control at the fifth day. The measured values of water content showed changes at the same time points but in the opposite direction, reduction at time points 2 and 4, for the unwatered and with reduced watering variants, respectively. Based on these invasive analysis data it can be concluded that the plants were suffering from water lack.

5.2.2. Maize in the field

Due to the small amount of images, which were taken on the same day, the resulting class frequencies of the maize data cannot be analysed like the barley data sets. An intuitive and simple visualisation of the class frequencies of the twelve maize images is not possible, wherefore a dimensional reduction by a Principal Component Analyse (PCA) was performed. The score values of the first two principal components enable a simple visualisation, wherein the different treatments can be depicted (Fig. 8).

As the barley models are based on drought stressed plants, a differentiation between the "untreated" and "irrigated" maize plants was the aim of this classification. Fig. 8 shows that the "irrigated" plants are separable from the "untreated" plants for both source models ($\mathscr{D}_{\mathscr{S}} = Barley2010$ and $\mathscr{D}_{\mathscr{S}} = Barley2011$). Furthermore, "irrigated and fertilised" plants are separable with $\mathscr{D}_{\mathscr{S}} = Barley2011$ from all other treatments.

6. Conclusion

In this article we have described a novel method for unsupervised Domain Adpatation. An unique feature of our approach is that no labels are required in the target domain. We achieve the appropriate adaptation by the careful design of an evaluation function which is based on biological plausibility. We showed that unsupervised Domain Adaptation is able to uncover early stages of drought stress in hyperspectral images by reusing an existing model. A transformation of the target data overcame the differences between source and target domain and enabled the application of the existing model. The lack of labels in the target domain was tackled by introducing an objective function which evaluated the transformation. In the objective function prior knowledge about the characteristics of plants and the propagation of drought induced stress was combined with characteristics from machine learning. The suitability of this transfer was shown by applying it to all combinations of the three proposed data sets and in all directions. The proposed method could improve the classification and overcame even differences in crops and their environments. Especially the transfer from the laboratory to the field reveals the possibilities of Transfer Learning for stress detection.

Acknowledgements

The authors acknowledge the funding of the CROP.SENSe.net project in the context of Ziel 2-Programms NRW 2007–2013 "Regionale Wettbewerbsfähigkeit und Beschäftigung (EFRE)" by the Ministry for Innovation, Science and Research (MIWF) of the state North Rhine Westphalia (NRW) and European Union Funds for regional development (EFRE) (005-1103-0018) while the preparation of the manuscript. The authors thank the staff of DISAT-UNIMIB and Universit cattolica di Piacenza for the support during the Italian field campaigns and the set-up of the maize field experiment.

References

Arnold, A., Nallapati, R., Cohen, W.W., 2007. A comparative study of methods for transductive transfer learning. In: Seventh IEEE International Conference on Data Mining Workshops, 2007. ICDM Workshops 2007. IEEE, pp. 77–82.

Behmann, J., Schmitter, P., Steinrücken, J., Plümer, L., 2014a. Ordinal classification for efficient plant stress prediction in hyperspectral data. Int. Arch. Photogram. Remote Sens. Spatial Inform. Sci. 40 (7), 29.

Behmann, J., Steinrücken, J., Plümer, L., 2014b. Detection of early plant stress responses in hyperspectral images. ISPRS J. Photogram. Remote Sens. 93, 98–111.

Blitzer, J., Kakade, S., Foster, D.P., 2011. Domain adaptation with coupled subspaces. In: International Conference on Artificial Intelligence and Statistics, pp. 173–181.

Bruzzone, L., Marconcini, M., 2009. Toward the automatic updating of land-cover maps by a domain-adaptation svm classifier and a circular validation strategy. IEEE Trans. Geosci. Remote Sens. 47 (4), 1108–1122.

Bruzzone, L., Marconcini, M., 2010. Domain adaptation problems: a dasvm classification technique and a circular validation strategy. IEEE Trans. Pattern Anal. Mach. Intell. 32 (5), 770–787.

Canny, J., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. (6), 679–698

Cardoso, J.S., Sousa, R., 2011. Measuring the performance of ordinal classification. Int. J. Pattern Recogn. Artif. Intell. 25 (08), 1173–1195.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37 (1), 35–46.

Curran, P.J., Dungan, J.L., Peterson, D.L., 2001. Estimating the foliar biochemical concentration of leaves with reflectance spectrometry: testing the kokaly and clark methodologies. Remote Sens. Environ. 76 (3), 349–359.

Datt, B., 1998. Remote sensing of chlorophyll *a*, chlorophyll *b*, chlorophyll *a+b*, and total carotenoid content in eucalyptus leaves. Remote Sens. Environ. 66 (2), 111–121.

Frank, E., Hall, M., 2001. A simple approach to ordinal classification. In: Proceedings of the 12th European Conference on Machine Learning. Springer-Verlag, pp. 145–156.

Gamon, J., Penuelas, J., Field, C., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sens. Environ. 41 (1), 35–44.
Gamon, J., Surfus, J., 1999. Assessing leaf pigment content and activity with a reflectometer. New Phytol. 143 (1), 105–117.
Gitelson, A., Merzlyak, M.N., 1994. Spectral reflectance changes associated with autumn senescence of Aesculus hippocastanum l. and Acer platanoides l. leaves. spectral features and relation to chlorophyll estimation. J. Plant Physiol. 143 (3), 286–292.
Gitelson, A.A., Merzlyak, M.N., Chivkunova, O.B., 2001. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. Photochem. Photobiol. 74 (1), 38–45.
Gitelson, A.A., Zur, Y., Chivkunova, O.B., Merzlyak, M.N., 2002. Assessing carotenoid content in plant leaves with reflectance spectroscopy. Photochem. Photobiol. 75 (3), 272–281.
Gómez-Chova, L., Tuia, D., Moser, G., Camps-Valls, G., 2015. Multimodal classification of remote sensing images: a review and future directions. Proc. IEEE 103 (9), 1560–1584.
Hess, M., Barralis, G., Bleiholder, H., Buhr, L., Eggers, T., Hack, H., Stauss, R., 1997. Use of the extended bbch scale general for the descriptions of the growth stages of mono; and dicotyledonous weed species. Weed Res. 37 (6), 433–441.
Huang, J., Gretton, A., Borgwardt, K.M., Schölkopf, B., Smola, A.J., 2006. Correcting sample selection bias by unlabeled data. In: Advances in Neural Information Processing Systems, pp. 601–608.
Huete, A., Liu, H., Batchily, K., Van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of tm images for eos-modis. Remote Sens. Environ. 59 (3), 440–451.
Jensen, J.R., 2009. Remote Sensing of the Environment: An Earth Resource Perspective. Pearson Prentice Hall.
Karimi, Y., Prasher, S., Patel, R., Kim, S., 2006. Application of support vector machine technology for weed and nitrogen stress detection in corn. Comput. Electron. Agric. 51 (1), 99–109.
Kaufman, Y.J., Tanré, D., 1996. Strategy for direct and indirect methods for correcting the aerosol effect on remote sensing: from avhrr to eos-modis. Remote Sens. Environ. 55 (1), 65–79.
Kirkpatrick, S., Vecchi, M., et al., 1983. Optimization by simmulated annealing. Science 220 (4598), 671–680.
Kohavi, R., et al., 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI, vol. 14. pp. 1137–1145.
Lewis, D.D., Gale, W.A., 1994. A sequential algorithm for training text classifiers. In: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Springer-Verlag New York, Inc, pp. 3–12.
Liu, Y., Li, X., 2014. Domain adaptation for land use classification: a spatio-temporal knowledge reusing method. ISPRS J. Photogram. Remote Sens. 98, 133–144.
MacQueen, J., et al., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1. California, USA, pp. 281–297.
Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. Physiol. Plant. 106 (1), 135–141.
Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogram. Remote Sens. 66 (3), 247–259.
Mucherino, A., Papajorgji, P., Pardalos, P.M., 2009. A survey of data mining techniques applied to agriculture. Oper. Res. 9 (2), 121–140.
Murphy, K.P., 2012. Machine Learning: A Probabilistic Perspective. MIT Press.
Pan, S.J., Yang, Q., 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.
Paul, A., Rottensteiner, F., Heipke, C., 2016. Iterative re-weighted instance transfer for domain adaptation. XXIII ISPRS Congress, Commission III, vol. 3(3). Copernicus GmbH, Göttingen, pp. 339–346.
Pennisi, E., 2008. The blue revolution, drop by drop, gene by gene. Science (New York, NY) 320 (5873), 171.
Penuelas, J., Baret, F., Filella, I., 1995. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. Photosynthetica 31 (2), 221–230.
Platt, J., 2000. Probabilistic output for support vector machines and comparisons to regularize likelihood methods. In: Advanced in Large Margin Classifiers.
Römer, C., Bürling, K., Hunsche, M., Rumpf, T., Noga, G., Plümer, L., 2011. Robust fitting of fluorescence spectra for pre-symptomatic wheat leaf rust detection with support vector machines. Comput. Electron. Agric. 79 (2), 180–188.
Römer, C., Wahabzada, M., Ballvora, A., Pinto, F., Rossini, M., Panigada, C., Behmann, J., Léon, J., Thurau, C., Bauckhage, C., 2012. Early drought stress detection in cereals: simplex volume maximisation for hyperspectral image analysis. Funct. Plant Biol. 39 (11), 878–890.
Rouse, J., Haas, R.H., Schell, J., D.W., D., 1973. Monitoring vegetation systems in the great plains with erts. In: NASA. Goddard Space Flight Center Third ERTS-1 Symp., vol. 1. pp. 309–317.
Schmitter, P., Behmann, J., Steinrücken, J., Mahlein, A.-K., Oerke, E., Plümer, L., 2015. Aktives lernen zur detektion von pflanzenkrankheiten in hyperspektralen bildern. Wissenschaftlich-Technische Jahrestagung der DGPF, Köln, Germany, pp. 398–406.
Scholander, P.F., Hammel, H., Bradstreet, E.D., Hemmingsen, E., 1965. Sap pressure in vascular plants. Science 148 (3668), 339–346.
Sims, D.A., Gamon, J.A., 2002. Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages. Remote Sens. Environ. 81 (2), 337–354.
Spearman, C., 1904. The proof and measurement of association between two things. Am. J. Psychol. 15 (1), 72–101.
Tardieu, F., Schurr, U., 2009. White paper on plant phenotyping. In: EPSO Workshop.
Tuberosa, R., Salvi, S., 2006. Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci. 11 (8), 405–412.
Tuia, D., Volpi, M., Trolliet, M., Camps-Valls, G., 2014. Semisupervised manifold alignment of multimodal remote sensing images. IEEE Trans. Geosci. Remote Sens. 52 (12), 7708–7720.
Turner, N.C., 1988. Measurement of plant water status by the pressure chamber technique. Irrig. Sci. 9 (4), 289–308.
Vogelmann, J., Rock, B., Moss, D., 1993. Red edge spectral measurements from sugar maple leaves. Int. J. Remote Sens. 14 (8), 1563–1575.
Vogt, K., Paul, A., Ostermann, J., Rottensteiner, F., Heipke, C., 2017. Boosted unsupervised multi-source selection for domain adaptation. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 4.