Grassland Species Characterization for Plant Family Discrimination by Image Processing

Mohamed Abadi[1], Anne-Sophie Capelle-Laizé[1], Majdi Khoudeir[1], Didier Combes[2], and Serge Carré[2].

[1] Université de Poitiers, XLIM-SIC, BP 30179, 86962 Futuroscope-Chasseneuil
{abadi,khoudeir,capelle}@sic.sp2mi.univ-poitiers.fr
[2] INRA Poitou-Charentes, UR4 P3F, BP 6, 86600 Lusignan, France
{dcombes,scarre}@lusignan.inra.fr

Abstract. Pasture species belonging to poaceae and fabaceae families constitute of essential elements to maintain natural and cultivated regions. Their balance and productivity are key factors for good functioning of the grassland ecosystems. The study is based on a process of image processing. First of all an individual signature is defined while considering geometric characteristics of each family. Then, this signature is used to discriminate between these families.

Our approach focuses on the use of shape features in different situations. Specifically, the approach is based on cutting the representative leaves of each plant family. After cutting, we obtain leaves sections of different sizes and random geometry. Then, the shape features are calculated. Principal component analysis is used to select the most discriminatory features.

The results will be used to optimize the acquisition conditions. We have a discrimination rate of more than 90% for the experiments carried out in a controlled environment. Experiments are being carried out to extend this study in natural environments.

Keywords: shape features, plant classification, leaf recognition, pasture, poaceae and fabaceae family, image processing.

1 Introduction

The areas occupied by temporary or permanent grasslands are now declining due to the increasing woodland and their transformation into arable land [1]. The balance and productivity of these pasture species and their productivity are key outputs for functioning of our grassland ecosystems. They provide a specific and significant contribution to biodiversity and play a major environmental role. Identification of each family allows to measure the proportion of each one in order to understand the spatial evolution of these families. This allows a certain balance between families which is essential to sustainability and agricultural value of temporary grassland. Moreover, studying the effects of light competition of different components of these vegetation cover has been the subject of many studies [2–5].

A. Elmoataz et al. (Eds.): ICISP 2010, LNCS 6134, pp. 173–181, 2010.

In a general case of characterization and identification of grassland species, the main objective of this study is specifically to discriminate between grassland families (poaceae and fabaceae). Thus to understand the evolution of these species within populations and assist enduring environment management, National Institute for Agricultural Research (INRA) of Lusignan (France) has established an experimental plots combination representative species of these two families. The idea is to have indicators to quantify the rate of different families within mixed grassland. Image analysis is effective and less costly in terms of the manpower than classical procedures. Species were cultivated separately. This allows to achieve acquisitions in a controlled environment, thus facilitating the processing and analysis phase. Samples of five species are photographed (white clover, alfalfa, orchardgrass, tall fescue and ryegrass).

This work allows individual characterization of each species based on the study of their geometric shape. It depends on acquisition quality and performed pre-processing [6]. We explain our choice by the fact that the humans easily identify objects by their shape [7] and the features that describe shape are widely used to classify plants by characterizing their leaves [8]. For example, Im et al. [9] used polygon approximation method to describe the geometric shape of leaves which helps to recognize the Acer family. Wang et al. [10] have developed a method that combines different shape features computed on leaves to identify plants in an image database.

The rest of the paper is organized in the following manner. Section 2 describes our acquisition process and pre-processing developed to extract the region of interest. The adopted procedure processing to estimate the limits of shape features and their robustness. The techniques to characterize the shape features and classification algorithms are briefly mentioned in Section 3. In Section 4, we describe the results obtained for poaceae and fabaceae plant families. Section 5 presents the conclusions and perspectives of this study.

2 Image Characteristics and Pre-processing

In this section, we present the leaves to describe poaceae and fabaceae families on which we apply our approach.

2.1 Acquisition Process

For the construction of image database of these two families, a number of sets composing of different cultures were cultivated at INRA of Lusignan under controlled conditions. These sets are composed of several samples. One sample represents each isolation plant (Figure 1 left). The leaf images were acquired for each species in XLIM-SIC laboratory. These acquisitions are effectuated in visible field and under controlled lighting. The acquisitions capture the detailed leaves of these species with NIKON D100 camera (Figure 1 center). This camera allows obtaining images in raw format (i.e. NEF: format without pre-treatment). This format allows exploring all color components (R, $g1$, $g2$ and B) unlike other formats (i.e. tiff, jpeg, . . .).

Our database includes 5 species (white clover, alfalfa, orchardgrass, tall fescue and ryegrass) belonging to poaceae and fabaceae families (Figure 2). For each species, we took 20 samples. So in total we have 180 images. 60 leaves represent the poaceae family (orchardgrass, tall fescue and rye-grass) and 120 leaves represent the fabaceae family (white clover and alfalfa). Each leaf belonging to fabaceae family is composed of 3 leaflets. We applied specific treatments based on morphological operators (erosion followed by dilation) to separate them. This allowed an increase in the number of processed leaves and consequently a better robustness of the searched signature.

2.2 Images Pre-processing

Through these acquisitions, we obtain images at 4 color components whose resolution is 1012×1518. Many authors have shown that effective processing applied on color images, specifically segmentation is determined by using an adequate color space [11–14].

The goal of our pre-processing is to extract the area of interest. For this, we propose to determine, for a given image and for each class, the optimal color component combination.

We used a criterion which, for every iteration, allows maximizing the separation between pixels for every iteration belonging to the two classes. The general scheme of our algorithm is presented in Figure 1 (right).

The optimal combination is based on an iterative procedure. At each iteration the procedure selects a new color combination that is associated with two classes previously identified by a thresholding method. The goal is to define the most discriminating hybrid color combination. For this we use the higher discriminant power. Corresponding regions are then labelled. Then it is used in the next iteration to extract color component data for all other regions. Finally, the procedure is stopped after the defined iteration n by the expert.

The advantage of this approach is that it provides the optimal color combination for each class identified in the image. It provides an effective solution to the problem of correlation between color components and ensures a better discriminant power compared to treatments used on the classical and/or hybrid color spaces. These pre-processing were used to get an intermediate version of the images (binary images). They showed the leaves were well localized and now the results can be used for further processing (i.e. shape features extraction).

3 Discriminative Signatures for Plant Families

To characterize the two families, images with high resolution are now available. To treat them, we developed a procedure consisting of three steps:

– Extract region of interest (leaves localization) (Section 2.2),
– these regions are characterized by computing shape features allowing to describe the geometry,
– Classification algorithms are applied on these features to automatically discriminate between families.

3.1 Features Used

The morphological characteristics of the leaves of these plants are a relevant recognition parameter which is used in plant taxonomy. Indeed, some recent work has focused on the characterization of the leaves for plant recognition. Here we are particularly interested on approaches that explore the geometric shape. We choose to compute the digital morphological features (DMF) because they are most conventional, widely used and are considered most important to characterize the leaves of a plant. Moreover, the geometric shape of the leaf is sub-oval (fabaceae) or highly elongated (poaceae).

Several features have been studied, for reasons of features robustness we have chosen to present only the results using the DMF. They are extracted from the contours of the leaf. These features generally include geometrical features (GF) [15, 16] and invariable moment features (MF) [7, 15, 17]. These are robust and invariant under translation, rotation and scaling.

3.2 Classification Algorithm

Classification is the final step of our processing. The selected discriminant features make the classification method choice relatively less important. In this study, we apply the LDC (Normal Densities based Linear (multi-class) Classifier), QDC (normal Densities Based Quadratic (multi-class) Classifier) and UDC (Uncorrelated normal Densities Based Quadratic Classifier) algorithms. These algorithms are defined in PRTools toolbox version 3.0 developed by Duin [18]. These methods were chosen for their fast convergence, their supervised mode and their adaptation to features having a Gaussian probability density. In this paper the samples are partitioned into training and test sets. On the training set a pattern recognition based on these 3 classifiers is defined. These will serve to estimate the error over the test set.

3.3 Treatment Approach

In order to study the limits and robustness of the techniques to characterize the geometry of the leaf of the plant, we proceed by cutting the leaf images in regions of decreasing size. This choice is justified by the fact that this study fits into the recognition of species within a family's problem. Indeed, in a real environment, the plants are mixed and access to whole leaf is not the most frequent case. Therefore, leaf cutting allows better evaluated quality and capacity of discrimination of features selected and simulating real cases.

The size and geometry of these regions change with the used cutting. For each leaf (Figure 3), we applied 3 cuttings (Figures 4, 5 and 6) composed 3 sets of data.

- Set 1: the leaves are cut into 4 regions (Figure 4),
- Set 2: the leaves are cut into 6 regions (Figure 5),
- Set 3: the leaves are cut into 8 regions (Figure 6).

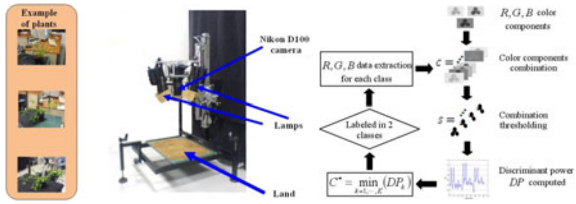

Fig. 1. Left acquisition system and example of plants used. Right general scheme for proposed leaves extraction.

Fig. 2. Examples of plant leaf representative of 5 species studied in this paper. Left to right poaceae family (orchardgrass, tall fescue and rye-grass) and fabaceae family (white clover and alfalfa).

Fig. 3. Leaves after pre-treatment. Left to right: white clover (437×398) and orchard-grass (1024×617)).

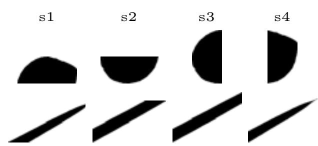

Fig. 4. Leaves cutting in 4 sections. Top to bottom Left to right: white clover (437×199) and orchardgrass (1024×308)).

Fig. 5. Leaves cutting in 6 sections. Top to bottom Left to right: white clover (437×132) and orchardgrass (1024×205)).

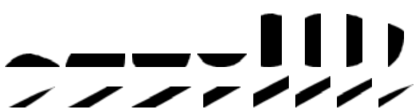

Fig. 6. Leaves cutting in 8 sections. Top to bottom Left to right: white clover (437×99) and orchardgrass (1024×154)).

Table 1. Hu invariant moments values competed on samples of Figure 3

	White clover	orchardgrass
ϕ_1	0.2190	1.3724
ϕ_2	0.1024	0.2276
ϕ_3	0.0353	0.1586
ϕ_4	0.0220	0.1840
ϕ_5	0.0247	-0.1745
ϕ_6	0.0307	0.1940
ϕ_7	-0.0178	-0.1530

Table 2. Hu invariant moments values competed on samples of Figure 4

	White clover				orchardgrass			
	s1	s2	s3	s4	s1	s2	s3	s4
ϕ_1	1.5786	1.5673	1.3937	1.2854	0.2714	0.2778	0.2441	0.2311
ϕ_2	1.4275	1.4054	1.2271	1.0876	0.1752	0.1490	0.0824	0.0977
ϕ_3	0.7799	0.7896	0.6520	0.6889	0.0467	0.0889	0.0191	0.0544
ϕ_4	0.7209	0.6971	0.5733	0.6679	0.0226	0.0681	0.0145	0.0376
ϕ_5	0.7312	0.7160	0.5867	0.6704	0.0224	0.0727	0.0151	0.0413
ϕ_6	0.8308	0.8085	0.6631	0.7336	0.0314	0.0737	0.0223	0.0458
ϕ_7	-0.5849	-0.5620	0.4899	0.5210	0.0269	-0.0479	0.0140	-0.0240

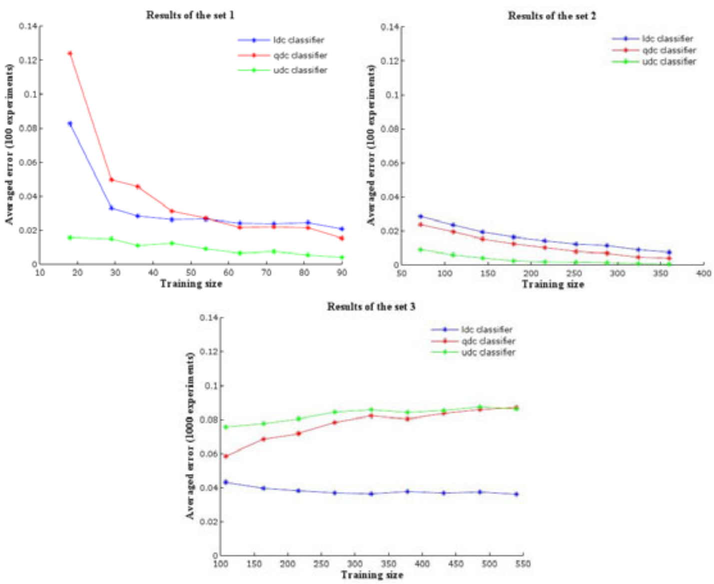

Fig. 7. Classification error for three examples of cutting and classifiers

In each area, geometrical features (14 features) and those based on Hu invariant moments (7 features) are computed. The principal component analysis is then used to select only the most discriminative features. These will constitute the data processed by classification algorithms previously defined.

4 Results

Here we present the results describing the three cuttings that were used. Figure 7 presents classification results on data sets constituting the test set for three classifiers. Note that the *LDC* classifier returns a low error rate classification (0.04) for set 3 unlike to other classifiers. This rate remained stable despite the change in the size of the training set (samples $\geq$100). Similarly, the result obtained by the same classifier is very good for the other sets despite small sizes of training set (samples $\geq$30). They showed a lot of potential to discriminate leaves.

In general, the developments effectuated show that the shape features allowed to discriminate between families where discrimination rate remains higher 90% for cutting ≤ 3. Without the cutting of the leaves, the rate varies between 85% and 100% depending on features selected, classification algorithms and size of training set. In the case where the cutting is > 3, these features face difficulties. This is explained by the fact that the selection of the obtained leaves become smaller. So the geometry is reduced to rectangles with a few variations on the edges. Precisely the rate of discrimination is 73% where leaves are cut into 10 regions.

Tables 1 and 2 give an idea about the Hu invariant moments values variation before normalization. These values are computed on representative plant leaf samples showed in Figure 3 and 4. So in light objectives, we believe that shape features based on digital morphological features give a best separation between the two families (fabaceae and poaceae). Thus the maximum cutting is 3. So the width of the section of the leaf must be greater than the pixels for the two families.

5 Conclusions

In this paper, we have presented an approach for measuring limits and robustness of techniques to characterize the shape of objects. This approach has been tried to separate fabaceae and poaceae families. This separation is based on plant leaf species recognition extracted from high definition images. Recognition is performed by a supervised classification process based on shape leaf description using digital morphological features. These leaves are representative of grassland species studied by INRA in Lusignan (white clover, alfalfa, orchardgrass, tall fescue and rye-grass). The *DMF* computed after applied a segmentation process to locate the leaves and separate them from the ground. The results are satisfactory given the divisions employed.

This study shows that the shape features can separate the two families. The difficulty is that using only this feature can not identify the species composing these families. This problem is more complex to address because the strong resemblance of geometrical shapes of leaves within each family. Further studies are exploring texture and color features.

Acknowledgements. The authors would like to thank Poitou-Charentes region for financing this study.

References

1. Huygue, C., Bournoville, R., Couteaudier, Y., Duru, M.: Prairies et cultures fourragres en France : entre logiques de production et enjeux territoriaux. INRA (2005)
2. Leconte, D., Simon, J.: Diversit floristique des prairies Normandes Un patrimoine, un rle fonctionnel. Prairiales Normandie, Colloque du Robillard (2002)
3. Zaller, J.: Ecology and non-chemical-control of rumex crispus and r. obtusifolius (polygonaceae): a review. Weed research 44(6), 414–432 (2004)
4. Gebhardt, S., Schellberg, J., Lock, R., Khbauch, W.: Identification of broad-leaved dock (rumex obtusifolius l.) on grassland by means of digital image processing. Precision Agriculture 7(3), 165–178 (2006)
5. Himstedt, M., Fricke, T., Wachendorf, M.: Determining the contribution of legumes in legume-grass mixtures using digital image analysis. CROP Science 49, 1910–1916 (2009)
6. Neto, N., Meyer, G., Jones, D., Samal, A.K.: Plant species identification using elliptic fourier leaf shape analysis. Computers and electronics in agriculture 50(2), 121–134 (2006)
7. Wang, X., Huang, D., Du, J., Huan, X., Heutte, L.: Classification of plant leaf images with complicated background. Applied mathematics and computation 205(2), 916–926 (2008)
8. Du, J.-X., Wang, X.-F., Gu, X.: Shape matching and recognition base on genetic algorithm and application to plant species identification. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 282–290. Springer, Heidelberg (2005)
9. Im, C., Nishida, H., Kunii, T.: Recognizing plant species by leaf shapes-a case study of the acer family. In: ICPR '98: Proceedings of the 14th International Conference on Pattern Recognition, Washington, DC, USA, vol. 2, pp. 1171–1173. IEEE Computer Society Press, Los Alamitos (1998)
10. Wang, Z., Chi, Z., Feng, D.: Fuzzy integral for leaf image retrieval. In: FUZZ-IEEE'02. Proceedings of the IEEE International Conference on Fuzzy Systems, Honolulu, HI, USA, vol. 1, pp. 372–377. IEEE Computer Society Press, Los Alamitos (2002)
11. Plataniotis, K., Venetsanopoulos, A.: Color Image Processing and Applications. Springer, Heidelberg (2000)
12. Cheng, H., Jiang, X., Sun, Y.: Color image segmentation: Advances and prospects. Pattern Recognition 34(12), 2259–2281 (2001)
13. Vandenbroucke, N., Macaire, L., Postaire, J.: Color image segmentation by pixel classification in an adapted hybrid color space. Application to soccer image analysis. Computer Vision and Image Understanding 90(2), 190–216 (2003)

14. Yang, L., Yang, X.: Multi-object segmentation based on curve evolving and region division. Chin. J. Comput. 27(3), 420–425 (2004)
15. Du, J., Wang, X., Zhang, G.: Leaf shape based plant species recognition. Applied mathematics and computation 185(2), 883–893 (2007)
16. Wu, S., Bao, F., You Xu, E., Wang, Y., Chang, Y., Xiang, Q.: A leaf recognition algorithm for plant classification using probabilistic neural network. Computer Science Artificial Intelligence (2007)
17. Wang, X., Du, J., Zhang, G.: Recognition of leaf images based on shape features using a hypersphere classifier. In: Huang, D.-S., Zhang, X.-P., Huang, G.-B. (eds.) ICIC 2005. LNCS, vol. 3644, pp. 87–96. Springer, Heidelberg (2005)
18. Duin, R.: A Matlab toolbox for pattern recognition Prtools version 3.0 a edn. (January 2000), http://www.ph.tn.tudelft.nl/prtools