Leaf Segmentation under Loosely Controlled Conditions

Simone Buoncompagni¹ simone.buoncompagni2@unibo.it Dario Maio¹ dario.maio@unibo.it Vincent Lepetit² lepetit@icg.tugraz.at

Extracting accurately the shape of a leaf is a crucial step in image-based plant identification systems. The partial or total absence of textures on leaf surface and the high color variability of leaves belonging to same species make shape as the main recognition element. For such reason, leaf segmentation plays a decisive role in the leaf recognition process.

Even though many general segmentation methods have been proposed in the last decades, leaf segmentation presents specific challenges. In particular, a pixel-level precision is required in order to highlight fine scale boundary structures and discriminate similar global shapes. Moreover, even if the input image can typically be taken in controlled conditions, where the leaf is the only visible object over a white background, the user taking the picture is not necessarily an expert and the conditions are often not ideal: the leaf exhibits specular reflections, casts shadows, the background is never exactly white and is usually non-uniform, and the image can be blurry.

Recently, a solution for the problem at hand has been proposed in [1] where leaf segmentation is carried out by estimating the probability distribution of foreground and background pixels. However, several drawbacks appear in this formulation due to challenging leaves like pine needles, false positives detection related to shadows and false negatives detection related to specularities. Prior distributions and post-processing operations are employed to tackle such problem, with the risk of hurting the final leaf shape.

In this paper we introduce a new solution by training a pixel-wise classifier [3] that learns filter responses associated to background and foreground regions in images of leaves. Our classifier is trained by selecting positive (leaf) and negative (non-leaf) feature samples that lie on the neighboorhod of the leaf boundary thus focusing learning only on "sensitive" pixels. Such classifier is then applied to each pixel location of a given unknown test image. This provides a score map that we then threshold using two different thresholds to detect pixels that belong to foreground and background with a high level of probability. With these pixels at hand we initialize an EM algorithm with a good initial estimate of foreground and background cluster parameters in the saturation-value color space, differently from [1] which has to initialize the EM segmentation with the same values for all the images. The other difference with [1] is that we can consider as unlabeled data only the pixels that are in the neighborhood of the detected leaf boundary. This allows to keep focusing on segmenting correctly the pixels around the leaf boundary, and in practice it is enough to get a good segmentation of the other pixels, which are easier to classify.

For evaluation we use the Leafsnap *Field* image dataset publicly available online [1] where different leaves of different species are acquired against solid background and variable light conditions thus simulating typical images that a user could provide for plant recognition.

To train our pixel-wise classifier we randomly select one image for each species and we manually produce segmentation and thicker contours to discriminate between positive and negative training samples placed in the neighborhood of boundary. Since segmentation ground truth is not available and its manual production for thousands of images would require an inestimable amount of time, we considered a subset of the original *Field* dataset. Our testing set is made of 300 images: 150 images for which the EM approach of [1] performs already well thus producing faithful segmentation in accordance with the leaf shape plus 150 more challenging images for which EM partially or totally fails.

The general behavior of different methods can be qualitatively appreciated looking at Fig. 1 where results returned by Leafsnap, Leafsnap without post-processing (marked with *), GrabCut and our method are reported. As the reader can see comparing ground truth details with real segmentations, it is confirmed that post-processing hurts quality of

- ¹ Department of Computer Science and Engineering University of Bologna Bologna, Italy
- ² Institute for Computer Graphics and Vision Graz University of Technology Graz, Austria



(a) Leaf image



(b) Ground truth

(d) Leafsnap [1]

(f) Ours



(c) Leafsnap* [1]



(e) GrabCut [2]

Figure 1: Leaves segmentation under loosely controlled conditions with different methods. False positives in red, false negatives in orange (best viewed in color).

boundaries and should be avoided. On the other hand, GrabCut tends to return round contours. With our method some errors still remain, due to those background pixels that look strongly similar to leaf and vice-versa. However, our method represents a good trade-off since we do not use post-processing but at the same time we assure a good robustness to false positives. Furthermore, our contours are much closer to the ground truth.

- N. Kumar, P. N. Belhumeur, A. Biswas, D. W. Jacobs, W. J. Kress, I. C. Lopez, and J. V. Soares. Leafsnap: A Computer Vision System for Automatic Plant Species Identification. In *ECCV*, pages 502–516. Springer, 2012.
- [2] C. Rother, V. Kolmogorov, and A. Blake. Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts. *TOG*, 23(3):309–314, 2004.
- [3] A. Sironi, V. Lepetit, and P. Fua. Multiscale Centerline Detection by Learning a Scale-Space Distance Transform. In *CVPR*, pages 2697– 2704. 2014.