Low-cost image annotation for supervised machine learning. Application to the detection of weeds in dense culture.

Salma Samiei , Ali Ahmad, Pejman Rasti, Etienne Belin, David Rousseau

LARIS, UMR INRA IRHS, Université d'Angers,
62 Avenue de Notre Dame du Lac, 49000 Angers France
david.rousseau@univ-angers.fr.

Abstract

An open problem in robotized agriculture is to detect weeds in dense culture. This problem can be addressed with computer vision and machine learning. The bottleneck of supervised approaches lay in the manual annotation of training images. We propose two different approaches for detecting weeds position to speed up this process. The first approach is using synthetic images and eye-tracking to annotated images [4] which is at least 30 times faster than manual annotation by an expert, the second approach is based on real RGB and depth images collected via Kinect v2 sensor.

We generated a data set of 150 synthetic images which weeds were randomly positioned on it. Images were gazed by two observers. Eye tracker sampled eye position during the execution of this task [5, 6]. Area of interest was recorded as rectangular patches. A patch is considered as including weeds if the average fixation time in this patch exceeds 1.04 seconds.

The quality of visual annotation by eye-tracking is assessed by two ways. First, direct comparison of visual annotation with ground-truth which is shown an average 94.7% of all fixations on an image which fell within ground-truth bounding-boxes. Second, as shown in fig.1 eye-tracked annotated data is used as a training data set in four machine learning approaches and compare the recognition rate with the ground-truth.

These four machine learning methods are tested in order to assess the quality of the visual annotation. These methods correspond to handcrafted features adapted to texture characterization. They are followed by a linear support vector machine binary classifier. The table 1 gives the average accuracy and standard deviation. Experimental results prove that visual eye-tracked annotated data are almost the same as in-silico ground-truth and performances of supervised machine learning on eye-tracked annotated data are very close to the one obtained with ground-truth.

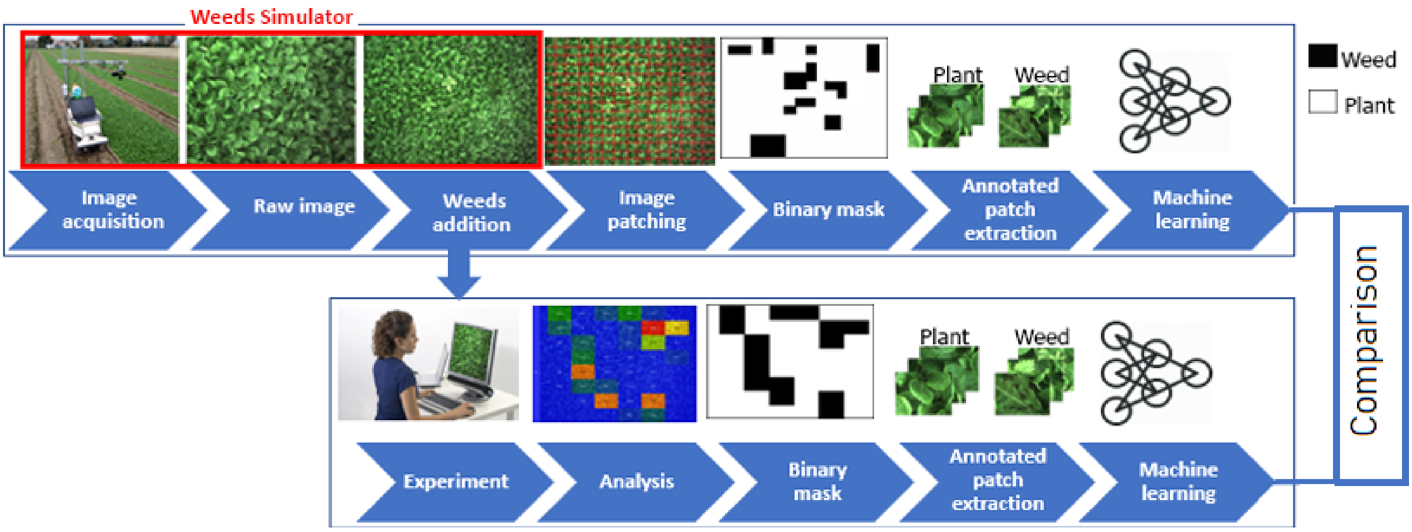

Figure 1: General pipeline of comparison of eye-tracked annotated data with ground-truth.

Table 1: Comparison of in-silico ground-truth annotated data with eye-tracked annotated data for weed detection.

Methods	Recognition Eye-tracking First observer (%)	Recognition Eye-tracking Second observer (%)	Recognition Ground-truth (%)
Scatter Transform [1]	64.8 std: 0.99	64.49 std: 0.72	68.1 std: 0.14
Local binary pattern [3]	70.1 std: 0.69	71.33 std:0.27	73.1 std: 0.11
Haralick coefficients	65.7 std: 0.75	66.19 std:0.59	67.41 std: 0.22
Gabor wavelet filters	64.7 std: 0.41	62.37 std:0.77	67.2 std: 0.54

Another fast automatic annotation method consists in creating for the training stage a multimodal acquisition. This can be for instance the coupling of RGB images with another imaging modality, more expensive than RGB but providing such high contrast that the segmentation in this second modality is trivial. Here we couple RGB and depth imaging [2]. We exploit the fact that weeds grow faster than plants. In-depth images of weeds, therefore, appear in smaller distance to the camera than crop plants and the weed can be segmented in depth with a simple threshold. After registration of RGB and depth images, it is thus possible to automatically annotate the RGB images as shown in Fig. 2. A comparison of manually annotated weeds shows a precision of 77% at the pixel level by this bimodal annotation.

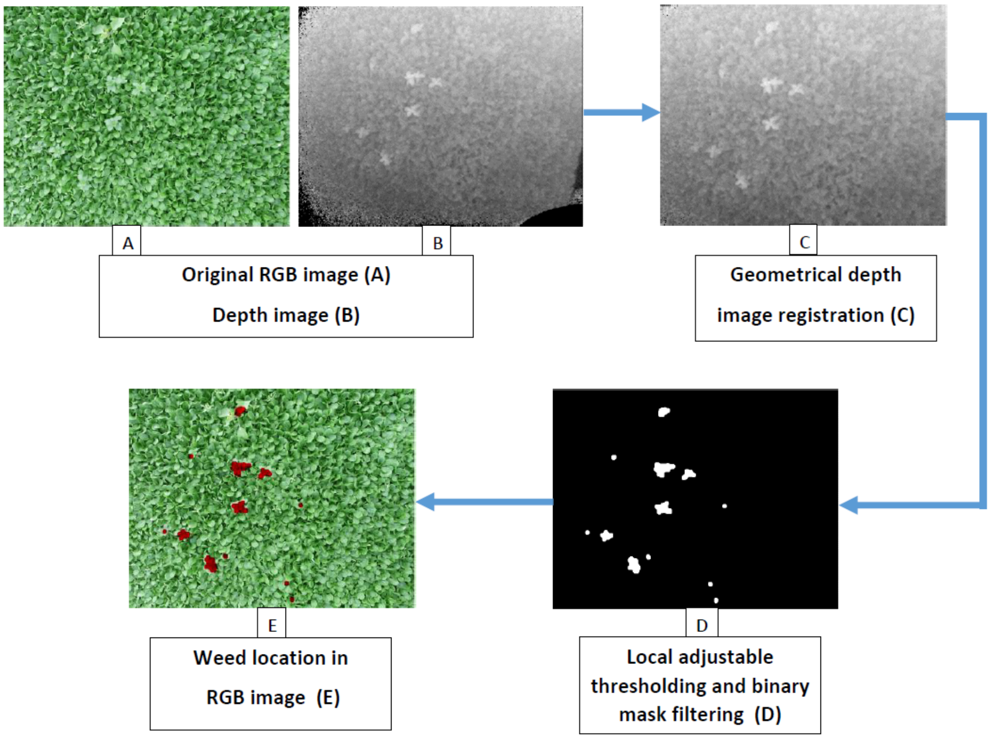

Figure 2: Weed automatic annotation diagram using bimodal RGB and depth images

[1] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transaction on pattern analysis and machine intelligence*, pages 1872–1886, 2013.

[2] Yann Chéné, David Rousseau, Philippe Lucidarme, Jessica Bertheloot, Valérie Caffier, Philippe Morel, Étienne Belin, and François Chapeau-Blondeau. On the use of depth camera for 3d phenotyping of entire plants. *Computers and Electronics in Agriculture*, 82:122–127, 2012.

[3] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[4] Dim P Papadopoulos, Alasdair DF Clarke, Frank Keller, and Vittorio Ferrari. Training object class detectors from eye tracking data. In *European conference on computer vision*, pages 361–376. Springer, 2014.

[5] Tina Walber, Ansgar Scherp, and Steffen Staab. Can you see it? two novel eye-tracking-based measures for assigning tags to image regions. In *International Conference on Multimedia Modeling*, pages 36–46. Springer, 2013.

[6] Kiwon Yun, Yifan Peng, Dimitris Samaras, Gregory J Zelinsky, and Tamara L Berg. Studying relationships between human gaze, description, and computer vision. In *Computer vision and pattern recognition (cvpr), 2013 ieee conference on*, pages 739–746. IEEE, 2013.