Original papers

Semantic labeling and reconstruction of grape bunches from 3D range data using a new RGB-D feature descriptor

Jennifer Mack[a,*], Frank Schindler[a], Florian Rist[b], Katja Herzog[b], Reinhard Töpfer[b], Volker Steinhage[a]

[a] Department of Computer Science IV, University of Bonn, Endenicher Allee 19A, 53115 Bonn, Germany
[b] Julius Kühn-Institut, Federal Research Centre of Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof, 76833, Germany

ABSTRACT

In the context of grapevine breeding, high precision and automated phenotyping plays an important role in order to screen breeding material (e.g. seedlings) or to characterize genetic repositories with high-throughput. Grape bunches hereby reveal a large variability regarding size, shape, compactness and color. We design and evaluate a new RGB-D descriptor for the semantic labeling of grape bunches. For this, we examine RGB and HSI color spaces and combine them with Fast-Point-Feature Histograms. With the best combination of FPFHs and the hue channel we achieve an average F-value of 88.61%, outperforming classical descriptors like PFHRGB and SHOTColor by at least 8%. Additionally, we show a new method for the derivation of parametric reconstructions of the elliptical berries based on a least squares fitting, yielding Pearson correlation coefficients of 0.8 and 0.9 for the main diameters of the berries.

1. Introduction

1.1. Motivation

Today, plant phenotyping, especially of crop plants, plays an important role with regard to plant breeding and precision agriculture. For grapevine breeding purposes, an objective pre-selection of seedlings as well as a comprehensive screening and phenotyping of genetic repositories are some of the most important needs. In general, phenotyping is done by visual estimations following the OIV list (Internationale de la Vigne et du Vin, 2009). The determination of phenotypic data is very labor-intensive, leading to the so-called "phenotyping bottleneck" (Furbank and Tester, 2011). Further, the achieved data are subjective and inaccurate. Thus, the development of automated methods suitable for high-throughput phenotyping is a high-priority task.

The color and shape of grape berries are important traits with regard to table grape breeding and production, the characterization of genetic repositories or in order to evaluate health status and visible quality traits within breeding material. To make this more efficient, developing automated methods for the phenotyping of grape bunches is an important goal. It is also challenging, as different grapevine cultivars show a high variability in shapes and colors of berries. With shape and color being the only features present, the only way to deal with this variability is to find a descriptor that is able to generalize over the different cultivars, but can still distinguish between the types of plant organs.

Previous work (Mack et al., 2017a,b; Paulus et al., 2013; Wahabzada et al., 2015) shows that Fast-Point-Feature Histograms (Rusu et al., 2009) provide a suitable description of points to label them as part of the surface of a berry or the stem structure, based on the local shape. This step is called the semantic labeling of the point cloud. In this contribution, we evaluate whether the addition of color information improves the results even in spite of the high variability of grape bunches in this context. We compare our resulting descriptor to the most common combinations of color and shape descriptors, namely PFHRGB (Rusu et al., 2008) and SHOTColor (Tombari et al., 2011). Examples for other descriptors applicable to RGB-D images are BRAND (Nascimento et al., 2012) and RISAS (Wu et al., 2017).

A method to derive exact parametric representations of spherical berries based on a RANSAC approach is presented by Mack et al. (2017a). While ellipsoid berries could be derived in the same way, tests show that this requires very high computation times due to the many parameters that have to be estimated. Therefore, we present an alternative approach, dividing the point cloud consisting of berry surface points into regions including only one berry each. A least squares fitting

---

* Corresponding author.
  E-mail address: mack@cs.uni-bonn.de (J. Mack).

as described by Li and Griffiths (2004) for ellipsoids is then used on the points of each region to provide the parametric description.

The goal of this work is to design a new RGB-D descriptor based on an evaluation of color and shape histograms regarding their suitability for the semantic labeling of points. We assume that adjusting the descriptor to the needs of our task (e.g. few differences in texture between the classes and variations regarding color and shape for each class) will lead to a better result than when using a more general descriptor. The classification is done unsupervised based on k-means with χ^2-distance, which has shown to be more reliable than the Euclidean distance when applied to histograms. Additionally, we derive a parametric representation for the ellipsoid berries that finally provides the possibility to compute the desired phenotypes, like the average radii of the berries.

1.2. Related work

Recently, different methods were presented for the high-throughput derivation of plant phenotypes. They can be categorized based on their input.

Some image-based approaches derive attributes, like bunch length, width and compactness (Cubero et al., 2015; Kicherer et al., 2015; Tello et al., 2016), the number (Aquino et al., 2017) and sizes of berries per bunch (Roscher et al., 2014). While in those publications the results have to be extrapolated to full grape bunches, as the 2D images only show a part, we work on full 3D scans, created by moving a 3D scanner (e.g., laser scanner, optical 3D scanner) around the grape bunch, offering more complete and exact results. Alternatively, it is possible to derive 3D data from images employing stereovision approaches (Klodt and Cremers, 2014; Ivorra et al., 2015; Rose et al., 2016), or combine depth data provided by a Time-of-Flight (TOF) camera with color information (Fernández et al., 2014). Still, these approaches are computationally expensive.

Earlier publications relied on scans created with expensive 3D laser scanners (Paulus et al., 2013; Mack et al., 2017a). New work by Rist et al. (2018) shows that the comparably cheap high-resolution handheld 3D Artec Spider Scanner that we also used here delivers results of sufficiently high quality. Furthermore, this scanner provides RGB data that was not included in earlier publications.

Schöler and Steinhage (2015) presented a first approach to the derivation of the full stem skeleton of a grape bunch. Due to the inner skeleton being completely occluded, they base their method on the visible parts of the grape bunch, namely berries and peduncle. While they use a RANSAC-method similar to Mack et al. (2017a) for the derivation of exact 3D berry detections and reconstructions, in both cases the berries are assumed to be spherical. Grape bunches with elliptical berries are not taken into account.

We use the Point Cloud Library (PCL, Rusu and Cousins, 2011) implementations of FPFHs, PFHRGBs, SHOTColor, region growing and RANSAC and the PCL-Visualizer to provide screenshots.

2. Materials and methods

The workflow is depicted in Fig. 1. The grape bunch (Fig. 1(a)) is fixed on a rotating hook and scanned from the front, generating a 3D point cloud (Fig. 1(b)). This point cloud is segmented into points that are part of a berry surface and points that are part of the stem skeleton (Fig. 1(c)). For this, we design a new descriptor as a combination of shape and color features. Afterwards, berry hypotheses are fitted into the berry point cloud using a least squares fitting method (Fig. 1(d)).

2.1. Plant material and sensor

We generated scans from 11 table grape bunches of the cultivars Sultana, Sugraone, Sugarthirteen, Ruby Seedless, and Prime, acquiring 2 independent scans per grape, yielding a total of 22 scans. To provide a representative data set, grape bunches of different colors and shapes

were chosen: 5 grape bunches with green berries (1 Sultana, 2 Sugraone, 2 Prime) and 6 red table grape cultivars (3 Sugarthirteen, 3 Ruby Seedless) with varying ellipticity. Fig. 2 shows examples that demonstrate the differences in shape and colors between the cultivars.

All point clouds were generated using the Artec 3D Spider scanner (Artec 3D, 2017) with a resolution of 0.1 mm and an accuracy of up to 0.05 mm. The grape bunch was fixed on a hook and rotated by 360° while scanning from the front side. This process yields a complete scan. Additionally to depth information, the scanner provides RGB values for each point. The acquisition of 3D point clouds is done under standardized lab conditions as described by Rist et al. (2018).

As reference, we use a manually created ground truth for the semantic labeling. The 3D reconstruction of the elliptical berries is evaluated with respect to statistical reference data for the number of berries and their average radii that was obtained from manual measurements of the grapes.

2.2. Color descriptor

We combine the descriptor based on shape and color information. For shape, we rely on the Fast-Point-Feature Histogram (FPFH) descriptor (Rusu et al., 2009) that has been shown to have superior performance to other 3D descriptors in phenotyping applications (Mack et al., 2017b; Paulus et al., 2013). However, so far there were no tests done combining color with shape information. To integrate color information, we evaluate different color channels and their combinations. As the scanner provides RGB values, we include the Red, Green, and Blue channel. For grape bunches with berries of the same color and clearly distinct stem skeleton, a semantic labeling of the data based on their RGB-values alone might be possible. But, as our data set contains grape bunches with red and green berries, we have to cover a more general spectrum. The HSI model has higher similarity to human perception of colors and the transformation of colors from one to the other is more intuitive. Because of that, we include it as another possibility for the labeling of our data, yielding the additional channels Hue, Saturation, and Intensity. Each of the channels is binned into a histogram. We set the number of bins for both FPFHs and color histograms fixed to 33, respectively, as the evaluation by Mack et al. (2017b) showed that increasing this value does not significantly change the results.

For the classification, the points are divided into two classes using a k-means approach with $k = 2$ and χ^2-distance. As usually more points in a point cloud are part of berry surfaces than the stem system, the cluster containing more points is assigned the label "berry points" and the cluster with fewer points the label "stem points".

2.3. Reconstruction of elliptic berries

In Mack et al. (2017a), a region growing approach is used to combine coherent points into regions. Berries are fitted into these regions with a RANSAC-based procedure. RANSAC can be used to fit different parametric primitives (Schnabel et al., 2007), but the computation time rises with the number of parameters that have to be optimized. For spheres, the optimization of 4 parameters (center point and radius) can be done relatively fast, but for ellipsoids, this number rises to 9 parameters and becomes highly time consuming.

A study by Rist et al. (2018) showed that, on the one hand, with a spherical RANSAC, the number of elliptical berries can be detected reliably, as they are fitted in the center or lower part of the berry. On the other hand, the length of the berries deviates strongly from the reference values, as spheres are not a sufficient representation of ellipsoid-shaped berries. The radii are an important feature for the phenotyping of berries, thus, it is necessary to improve this result.

To avoid the computational expensive RANSAC optimization, but still represent the berries as ellipsoids instead of spheres, we use a least squares fitting of ellipsoids that works faster than RANSAC. However, least squares fitting is less stable with respect to regions containing