A calculation method of plant similarity giving consideration to different plant features

Wei-long Ding [a,b,*], Shui-sheng Wu [a], Nelson Max [b], Fu-li Wu [a,b], Li-feng Xu [a]

[a] College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China
[b] Department of Computer Science, University of California, Davis 95616, USA

HIGHLIGHTS

- We report a strategy to compute the similarity between different plants.
- Topology structure, peripheral contour and geometry details of a plant are considered.
- By combining these different measures we get the similarity between different plants.
- The experimental results explain effectiveness of the proposed method.

ARTICLE INFO

ABSTRACT

A method to compute the similarity between different plants is proposed, using features of a plant's topological structure and peripheral contour, as well as its geometry. The topological structures are described using tree graphs, and their similarity can be calculated based on the edit distance of these graphs. The peripheral contour of a plant is abstracted by its three-dimensional convex hull, which is projected in several directions. The similarity of the different projections is calculated by an algorithm to compute the similarity of two-dimensional shapes. The similarity of the geometrical detail is computed by considering the geometrical properties of different level branches. Finally the overall similarity between different plants is calculated by combining these different similarity measures. The validity of proposed method is evaluated by detailed experiments.

1. Introduction

Morphology is a basis for classifying plants into different types (Ma and Yao, 2007; Zheng et al., 2011). The similarities of the morphologies, structures, and habits between different plant species is related to their closeness or distance in phenotype relationships. Thus the discrimination between different plants is a critical step in the classification and retrieval of vegetation. Traditional discrimination methods rely mainly on manual operation, and thus may be subjective, labor intensive, and unsuited for rapid classification and retrieval of vegetation

A plant-structure simulation model which can accurately describe the geometry and topology of a plant has significance in scientific research on evapotranspiration, the ideal plant type design of crops, and the optimization of cultivation measures (Guo and Li, 2001). However, in judging whether a model is precise or not, one must calculate the degree of similarity between the reconstructed 3D model and the real plant. Therefore, the definition and calculation of the similarity between different plants has important theoretical significance and practical value.

Compared to research on the similarity of three-dimensional models (Schneider and Behr, 2006; Pan et al., 2009), the comparison of DNA and protein sequences (Peng and Hu, 2006), and the similarity of malicious code (Yang et al., 2011), research on plant structure similarity is considerably weaker. Related research on plant structure similarity includes three aspects: 1) similarity of plant architectures (Ferraro, 2000; Godin and Caraglio, 1998; Ferraro et al., 2004]; 2) plant species identification based on blade similarity (Nam et al., 2008; Caballero et al., 2010; Du et al., 2007; Goeau et al., 2011; and 3) similarity of tree-structured data (Zhang, 1993; Zhang, 1995; Yang et al., 2005). The methods to calculate the similarity between two plant architectures include global comparison methods, analytical comparison methods, and tree-graph-based comparison methods (Ferraro and Godin, 2000). In global comparison methods, the similarity of two plant architectures is

* Corresponding author at: College of Computer Science & Technology, Zhejiang University of Technology, Hangzhou 310023, China. Tel.: +86 571 85290527.
E-mail address: wlding@zjut.edu.cn (W.-l. Ding).

calculated by using parameters such as fruit production, stem diameter, and crown size. This method can roughly compare the similarity of global structure, but cannot compare in detail the plant topology and organ arrangement geometry. The analytical comparison methods first statistically analyze the topology and the spatial distributions of organs of a plant, and then use these features to compare the similarity between two plants. The tree-graph based comparison method employs edit distance (Zhang, 1993) to describe the similarity between two plants. As it requires complex mathematical calculation and frequent operations, this method is somewhat complicated. Some proposed methods to identify the type of species of plants are based on the similarity of the blade (Nam et al., 2008; Caballero, Aranda; Du et al., 2007; Goeau et al., 2011). However, those studies only focus on the leaves, but neglect the similarity of plant topology. The study of similarity of tree-structured data (Zhang, 1993; Zhang, 1995; Yang et al., 2005) also pays attention only to the abstract tree graph, but there are significant differences with the real tree structures. In summary, plant structure similarity needs further study.

The objective of this paper is to present a method to compute the similarity between different plants which comprehensively considers plant topology, 2D projections of peripheral contour features, and inner details. The experimental results show that the proposed method can effectively calculate the similarity between different real plant architectures, and can distinguish different plant species, families, or genera based on their similarity.

2. Definition of the similarity of plantmorphology

We consider the similarity of three aspects of plantmorphology: 1) topology, which describethe structural relationship between various organs, 2) the peripheral outlines of a plant and the contour of each branch, and 3) the inner features, which describe the geometric characteristics, such as branching angles and diameters of the different organs. Assume that the similarities of plants in n different aspects have been calculated as the feature vector $s = (s_0, s_1, \ldots, s_{n-1})^T$ with each $s_i \in [0, 1]$. Also assume some empirical weightingfactors(in the range(0,1]) are assignedtoeach feature respectively, in the row vector $w = (w_0, w_1, \ldots, w_{n-1})$. Then theweightedaverage similarityis calculated as:

$$\overline{S} = \frac{\sum_{i=0}^{n-1} w_i s_i}{\sum_{i=0}^{n-1} w_i} \tag{1}$$

Let $S_m = \max\{s_0, s_1, \ldots, s_{n-1}\}$. Then our formula for calculating the similarity of plantsis:

$$S = b\overline{S} + (1-b)S_m \tag{2}$$

where b is an empirical constant. Note that S and $\overline{S}$ are both in [0, 1].

To calculate the similarity of plants, we consider seven features. The first two are the similarity S_t of topological structure and the similarity S_{3g} of peripheral outline. The other five are similarities of different inner features, which include: the average value of the angle between the branches and the stem (S_a), the diameter ratio of lateral branches to the stem (S_d), the width-to-height ratio of the peripheral outline (S_{wh}), the average value of the angle between second-order branches and first-order branches (S_{a1}), and the cross-sectional area ratio of lateral branches to the stem (S_{s1}).

3. Similarity of two plant topologies

3.1. The description of plant topology

The topological structure of a plant is determined by the relationships and distribution of internodes and nodes, usually represented by a tree graph (Ferraro and Godin, 2000; Zhang, 1993). A tree graph G is defined as a collection of vertices V and directed edges E, denoted as $G = \{V, E\}$. A vertex corresponds to a node. An edge corresponds to an internode connecting two nodes, and is represented by an ordered pair (v_i, v_j) (where v_i and v_j respectively represents the vertices). Edges are separated into two classes according to the geometry of the plant. An axial edge from v_i to v_j is one that continues in the direction from v_i's parent to v_i. Other edges are non-axial. We consider only trees with a single edge starting at the root, and this edge is also axial. In contrast to the articles (Ferraro and Godin, 2000; Zhang, 1993), in this paper the vertices only represent the nodes, not including leaves, flowers, fruits, andotherorgans. Unless otherwise specified, anode in the following sections also indicatesthevertexofatree graph. Two types of edges between nodes are used to identify the different axes on a given plant: a precedent relationship(denoted by "$<$") and a branching relationship (denoted by "$+$")(Ferraro and Godin, 2000). For example, in Fig. 1, v_6 is a child of node v_1 and the edge (v_1, v_6) is axial, so their relationship is precedent, denoted as $v_1 < v_6$; v_6 is a node of the main stem, and v_5 is a node of a branch, so the edge (v_5, v_6) is non-axial and their relationship is branching, denoted as $v_6 + v_5$. Similarly, we have $v_6 < v_7$, $v_7 < v_2$, $v_7 + v_8$. Extending the axial relationship across multiple internodes, we have $v_1 < v_2$, $v_7 < v_4$, and $v_1 + v_4$.

$T[v]$ indicates the full sub-tree whose root is node v (the collection of nodes that includes node v and all its descendant nodes), and $|T|$ indicates the number of nodes of the tree graph T. If v is not the root and the edge to v from its parent is axial, then the sub-tree $T(v)$ is called an axial sub-tree, and if this edge is non-axial, then $T(v)$ is called a branch sub-tree. The number of internodes along the growth direction of the terminal buds of a branch sub-tree is called the depth of a branch sub-tree. For example, in Fig. 2, the depths of branch sub-trees B, C , B', and C' are 4, 3, 3, and 2 respectively.

3.2. Branch degradation of tree graphs

In a simplified tree graph, a plant's topology is defined as the relation and distribution of the connections between the nodes. Therefore, we only consider the branching difference when comparing the similarity of plant topologies. To give a larger weight to the branching, we do branch degradation on the tree graphs. Branch degradation is performed by repeatedly removing any nodes v which have a parent node, and exactly one child node
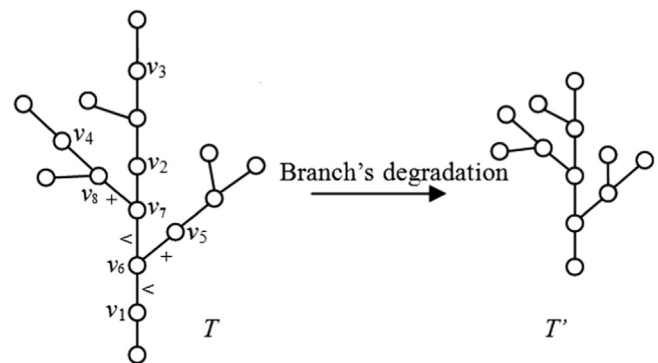

Fig. 1. Branch's degradation of tree graphs.