

# TERRESTRIAL IMAGE BASED 3D EXTRACTION OF URBAN UNFOLIAGED TREES OF DIFFERENT BRANCHING TYPES

Hai Huang

Institute of Photogrammetry and Cartography  
Bundeswehr University Munich, 85577 Neubiberg, Germany  
hai.huang@unibw.de, www.unibw.de/ipk

Commission III, ThS-7

**KEY WORDS:** Image Understanding, 3-D Feature Extraction, Computer Vision, Feature Extraction, Urban Planning, Vegetation, Three-dimensional, Statistics

## ABSTRACT:

In this paper we propose extensions to a generative statistical approach for three-dimensional (3D) extraction of urban unfoliated trees of different branching types from terrestrial wide-baseline image sequences. Unfoliated trees are difficult to extract from images due to their weak contrast, background clutter, and particularly the possibly varying order of branches in different images. By combining generative modeling by L-systems and statistical sampling one can reconstruct the main branching structure of trees in 3D based on image sequences in spite of these problems. Here, we particularly classify trees into different branching types and specific L-systems are applied for each type for a more plausible description. We combine Monte Carlo (MC) with subsequential Markov Chain Monte Carlo (MCMC) to robustly and efficiently deal with the sparse distributions of the branching parameters. First results show the potential of the extended approach.

## 1 INTRODUCTION

The extraction of trees can help to enhance 3D urban geoinformation. Trees are an essential component adding a natural touch to and for a realistic visualization of city models.

In this paper we aim at extracting the main branching structure of individual unfoliated deciduous trees from wide-baseline image sequences. We basically follow our approach presented in (Huang and Mayer, 2007), extending it with the following novelties:

- Classification of branching types and refined modeling
- Combined MC and MCMC sampling for robust and efficient search

Deciduous trees are popular in cities as they provide shadow in summer while letting the sunlight through in winter. Thus, they often form the majority of trees in urban areas. From a practical point of view images for data acquisition are often taken when trees are unfoliated, as facades, etc. are then more readily visible. From a scientific but also a practical point of view unfoliated trees have the advantage, that they explicitly show the branches. Yet, it is hard to extract unfoliated trees because of their geometric complexity, weak contrast, background clutter, and the varying order of the branches when projected into different images even when the images have been taken very close to each other.

Former work has mostly dealt with tree extraction in aerial images and especially recently laser scanner data. Much work focuses on forests. (Hyypä et al., 2005) aims in their empirical study on the accuracy of the estimation of tree volumes from aerial data. They estimate a segment for each tree from image data, but found, that the height is much more reliably determined from laser-scanner data than by image matching. (Persson et al., 2004) uses laser-scanner data for the determination of the outline

as well as the height. Image data, also in the infrared is used to mostly reliably differentiate pine, spruce, and deciduous trees.

Aerial images are used in (Cheng et al., 2006), which like our work employs a statistical (Bayesian) framework consisting also of a generative component. It comprises segmentation, stereo, and 3D fitting and is demonstrated by extracting individual trees. Both data-driven (inverse modeling, 2D data to 3D geometries) and generative – model-driven (3D models to 2D images) components are integrated.

(Pfeifer et al., 2004) and (Gorte and Pfeifer, 2004) extract detailed models for trees from terrestrial laser scanner data. (Pfeifer et al., 2004) aims at fitting cylinders to the trunk and the thicker branches. A possibly better solution is provided by (Gorte and Pfeifer, 2004). The laser points are rasterized in 3D voxel space and operations such as closing from mathematical morphology and thinning are used to obtain a connected 3D skeleton. We note that for both approaches a relatively high density of points is needed to avoid gaps in the skeleton, or even more critical, to fit cylinders at all.

In the work of (Sakaguchi and Ohya, 1999), which employs like our work terrestrial images, a volume is carved out by intersecting the view cones generated from the tree silhouettes in multiple images. The voxels of the volume are colored with the average brightness of the rays from the different images. A branching process is started on the ground extending into dark areas assumed to correspond to the trunk or branches. The given results are plausible, but there is much human intervention involved. A more sophisticated automatic approach is (Shlyakhter et al., 2001). 3D volumes are generated as in (Sakaguchi and Ohya, 1999). From the volumes 3D medial axes are constructed. The medial axes are constrained to the “botanical fidelity of the branching pattern and the leaf distribution” (Shlyakhter et al., 2001) via an open Lindenmayer-, or in short L-system (Měch and Prusinkiewicz, 1996). Again, manual interaction is employed to generate results which are good in terms of visualization.

Recent work includes (Tan et al., 2007), which focuses on the semi-automatic construction of realistic looking tree models from images. It is based on high quality structure from motion and dense depth estimation and it employs shape patterns of visible branches to reconstruct missing parts of the tree.

In this paper we show how generative statistical modeling based on L-systems and MCMC makes it feasible, to extract branches in wide-baseline image sequences taken unconstrained with standard consumer cameras in spite of the problems with clutter and occlusions stated above. Our basis is a highly precise structure from motion procedure (Mayer, 2005) making use of calibration via the five-point-algorithm (Nistér, 2004). Corresponding points are obtained with high precision by least-squares matching and bundle adjustment is used after every step.

In Section 2 the generative modeling of the tree structure by means of L-systems is described for three different branching types. The generation of 3D hypotheses, their 2D projection, and evaluation are presented in Sections 3 and 4 comprising also the classification of the branching types based on results for the first level of branches. Hypotheses for trunks are generated by line extraction and image matching while those for branches are produced by MC/MCMC sampling of their parameters in 3D object-space with given prior distributions based on the classified branching type. The evaluation of new hypotheses is conducted using the normalized cross correlation coefficient ( $CCC$ ) as a substitute for likelihood. After demonstrating the potential of the extended approach with preliminary results in Section 5, the paper ends up with conclusions.

## 2 L-SYSTEM FOR DIFFERENT BRANCHING TYPES

As stated above, the branching structure of trees is difficult to extract from terrestrial wide-baseline urban image sequences. To construct 3D models of trees bottom-up/data-driven from the images, we need to match the branches. Often, the ordering constraint, i.e., a point left of another point on an epipolar line in one image is also left of the corresponding point on the epipolar line in the other image, is employed to guide matching. Yet, because of the complex 3D structure of trees, the ordering constraint is often not valid for branches even for images taken close to each other. All this means that the bottom-up extraction and matching of branches does not seem promising and suitable constraints describing the structure of trees are essential for their 3D reconstruction.

We therefore decided to model the tree structure generatively, i.e., top-down/model-driven by means of L-systems (Měch and Prusinkiewicz, 1996). The latter are widely used in computer graphics to simulate the structure and growth of vegetation. They are parallel string rewriting systems representing branching structures in terms of bracketed strings of symbols possibly with associated numerical parameters. The simulation of branching starts with an initial string (axiom). By means of productions substrings of the predecessor string are substituted by successor strings according to specific rules. Models produced by L-systems are able to maintain basic botanic features such as hierarchical structure and self-similarity.

### 2.1 Branching types

Basically, branching structures of trees can be divided into two main groups: “monopodial” and “sympodial”, cf., e.g., (Deussen and Lintermann, 2005), for which different Production Rules have to be used. The monopodial branching system (cf. Fig. 1 (m)) has a prominent main axis, which is stronger and longer than the

side branches. The side branches are again stronger and longer than their side branches of the second order, etc. Because of the dominant axes, monopodial branching structures have a radially symmetric crown.

Fig. 1 (sd) and (sm) show the two main types of sympodial branching. “Sympodial-dichasium” – sd branching means, that two buds of a branch sprout and grow synchronously. For this kind of tree trunk and crown are clearly separated. The most common branching structure for trees is “sympodial-monochasium” – sm, where one of the secondary branches has approximately the same direction as the original branch. Sympodial-monochasium branching results into only partially symmetric branching structures, which will still often appear very similar to monopodial branching.

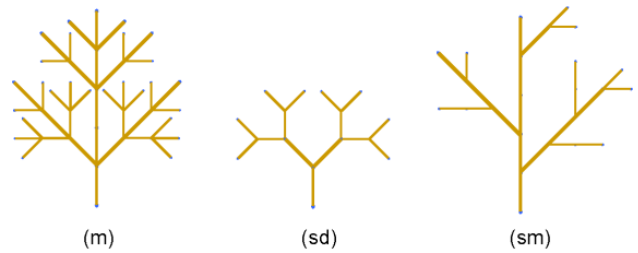


Figure 1: Branching types: (m) monopodial; (sd) sympodial-dichasium; (sm) sympodial-monochasium

### 2.2 L-systems

We have devised L-systems for the different branching types with predefined basic Production Rules for each type. An example for how L-systems work is given for monopodial – m trees. An L-system as shown in Fig. 1 (m) is, e.g., defined as follows:

$$G_{(m)} = (V, S, \omega, P)$$

with

$V_{(m)}$  (Variable): F

$S_{(m)}$  (Constants): +, -, <, >, [, ]

$\omega_{(m)}$  (Initial State): F

$P_{(m)}$  (Production Rule):  $F = F[+>F][-<F]F$

in which the Variable F corresponds to growth, i.e., a new piece of cylinder, while the Constants describe rotation around certain axes (“+” and “-” indicate turn left and right – inclinations; “<” and “>” indicate roll left and right – azimuths) and the creation of new branches (enclosed by brackets “[” and “]”). The Production Rule instructs to replace F with the specific string in the next iteration. Different Production Rules lead to different branching structures.

The L-systems for sympodial-dichasium – sd and sympodial-monochasium – sm (cf. Fig. 1 (sd) and (sm)) branching are defined in the same way, but with specific Production Rules:

$$P_{(sd)} : F = F[+<F][-<F]$$

$$P_{(sm)} : F = F[+>F]F[-<F]F$$

Variable and Constants are parameterized, i.e., the values of the size of the cylinder and the angles of rotation are not fixed. In contrast to graphical modeling with L-systems, the parameters in our approach are determined by statistical sampling to fit the real scene.

### 3 GENERATION OF 3D HYPOTHESES

The L-system provides a reasonable structure for the branches. Yet, the L-system alone only gives means to generate and also visualize trees. Thus, after the trunk is located by line extraction and image matching, the branches are extracted via a generative statistical approach: MC and MCMC are employed to link the 3D model to the scene, i.e., to find a plausible structure for the tree visible in the images. Likely candidates for branches are generated by stochastic sampling and are verified by comparing simulated and real images. Via classification the branching type is determined from the first level of branches.

#### 3.1 Extraction of the trunk

The trunk is a basic part of many trees we are interested in. For its extraction we assume that it corresponds to a thick, mostly vertical line and it defines the lower part of the main axis outside the crown. The vertical direction is presumed to be known approximately by basically taking images horizontally. It can often be improved by computing the vertical vanishing point from the vertical edges of trunks or on facades as we focus on urban scenes. Vertical lines, i.e., hypotheses for trunks, are verified by matching in several images. We use the trifocal tensor (Hartley and Zisserman, 2003) derived from the known orientation parameters to predict from lines in two images hypotheses for lines representing the trunk in other images. For the remainder of the paper we assume that the 3D position of the tree is determined by the trunk.

#### 3.2 Hypotheses generation for branches

A branch in 3D object-space is modeled as a cylinder with known begin. As parameters azimuth (angle with x-axis of branch projected into horizontal plane), inclination (angle between branch and horizontal plane), length, and diameter are used. New hypotheses are generated by sampling the space of the parameters.

In Fig. 2 the basic idea for the generation of branches is presented. Based on the trunk, branches are grown by sampling their parameters statistically, guided by appropriate prior distributions. I.e., the inclination is presumed to be often about 45 upwards and relatively seldom downwards. A newly generated hypothesis is projected into the images via the given highly precisely known orientation parameters. The hereby generated simulated images are matched to the given images (cf. Section 4). As model for the background clutter we use Gaussian noise.

Our current experiments are limited to the major branching structure, which includes the longer and thicker branches in the first few levels. Small twigs are ignored in modeling and search.

#### 3.3 Combined MC and MCMC sampling

Combined Monte Carlo plus subsequential Markov Chain Monte Carlo (MC+MCMC) is used for the statistical sampling. Since the parameters, especially the branching angles, are distributed sparsely in a relatively large space, plain Monte Carlo, i.e., random numbers are drawn from the given prior distributions, is used in the first phase for a coarse sampling and refined in the second phase by means of MCMC. MCMC, cf. (Neal, 1993), is characterized by the Markov property: a sampling step depends only on the previous step, i.e., the space of parameters is sampled locally.

Particularly, the best samples, e.g., ten from one hundred, from MC are taken as candidates for a refined search using ten MCMC iterations for each. The number of candidates is reduced to three

after the first round of refinement and the best one is finally found after another twenty MCMC iterations for each of them.

The basic MC search and the Metropolis-Hastings algorithm (Neal, 1993), integrated into our MCMC, both help to avoid local minima while still allowing to find solutions quickly. Preliminary prior distributions for the parameters have been devised from experience and observations, resulting in an efficient search, particularly after the first level of branching.

#### 3.4 Classification of branching types

The classification of the branching types is done after the determination of the first level of branches, i.e., the branches, which directly grow from the trunk. We employ a particularly flexible search for this level by adding the vertical Z-coordinate of the begin point of the branch, i.e., its joint position along the trunk, as an additional parameter. This parameter is sampled together with the branching angles. Thus, branches do not have to start exactly at the end point of trunk, but can be distributed in its vicinity.

The classification is based on the distribution of the determined branches: trees are supposed to be of the monopodial – m type if most of the joint positions of the extracted branches are concentrated in a relatively small area. If there is no branch maintaining the direction of the trunk, trees are classified as sympodial-dichasium – sd. If none of the above holds, the branches are assumed to be more suitably described as sympodial-monochasium – sm trees. The appropriate Production Rule of the corresponding L-system (cf. Section 2.2) is then employed for the further levels.

## 4 EVALUATION OF HYPOTHESES

### 4.1 2D projection

The 3D hypotheses generated above are projected into 2D resulting into simulated images. They are evaluated by comparing them with the given images. We use a simple and efficient 2D representation derived from the 3D representation instead of the actual projection of the 3D cylinder, as the latter entails a larger computational effort and for statistical sampling many projections are needed. Another reason for doing so is that the projection of the branches results into patches of nearly constant brightness anyhow.

The chosen 2D representation consists of trapezoids. A trapezoid is described by its direction (angle with x-axis), length, width of begin, and width of end. We determine the centers of the ends of a trapezoid by projecting the end points of the 3D axis of the cylinder into the image. Reasonable approximations for the widths are obtained from the projection of points on the normals to the lines from the end points to the projection center with the distance radius of the cylinder.

### 4.2 Evaluation by cross correlation coefficient

The cross correlation coefficient ( $CCC$ ) is used for the evaluation of the 2D-projections of the 3D branch hypotheses based on the given images. Hypotheses are projected with the trunk color, Gaussian noise simulates background clutter, and both together compose simulated images.

Each simulated image  $i$  and its corresponding given image are compared based on the intensities computed by the Hue-Saturation-Intensity (HSI) color transformation and results in an independent  $CCC_i$ . To be able to compare different hypotheses for the whole tree, the matching is done against the projections of the convex

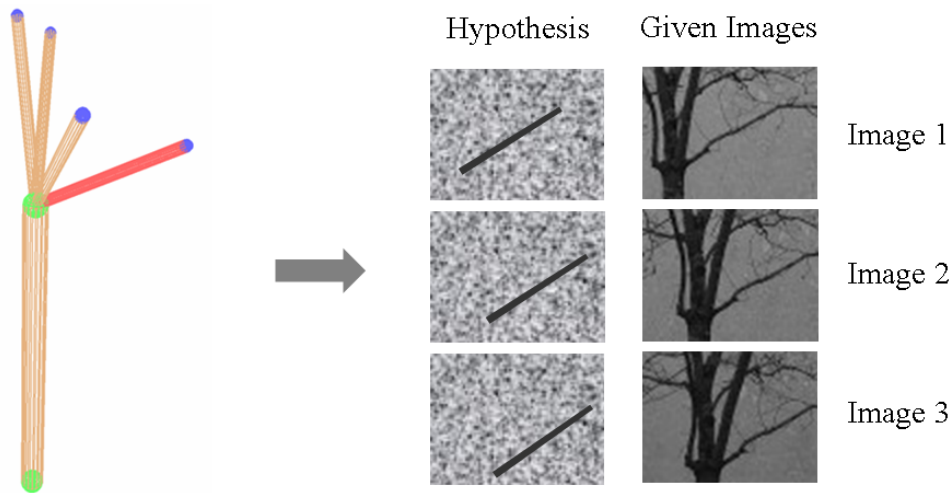


Figure 2: Stochastic sampling based on an L-system results in a 3D tree hypothesis (left). Projection of a new branch (red) into three empty images in the form of a trapezoid with Gaussian noise as background (center) and given image data (right). For the sake of clarity only the projection of the new branch is shown.

3D hull of all hypotheses. As MCMC sampling usually entails a larger number of iterations, the comparison has to be efficient. This is done by an incremental update of only those parts of the 2D projection and the corresponding variances and covariances, which have changed.

$CCC_i$  values for the  $n$  individual images are combined via multiplication into a global  $CCC$  value

$$CCC_{global} = \prod_{i=1}^n CCC_i. \quad (1)$$

We use multiplication because we assume that the  $CCC_i$  values are proportional to likelihoods and we assume independence of the images given the 3D model. Moreover, we found empirically, that this conservative combination helps to sort out wrong hypotheses early. We are aware that the actual size of the  $CCC_i$  values can be far from correct likelihoods. Yet, our experiments give evidence to assume that they are reasonably proportional to likelihoods.

The  $CCC$  values of all accepted hypotheses are normalized and result into a criterion, which hypotheses have to fulfill, to avoid implausible results. This is especially helpful for irregular trees and artificial pruning.

By means of experiments we found that it is not useful to sample all parameters of a branch at the same time. Thus, sampling of the parameters is conducted sequentially. Firstly, only azimuth and inclination (and additionally joint position for the first level, cf. Section 3.4) are jointly sampled by means of MC+MCMC as described in Section 3.3 while the length is kept fixed. The length is optimized only by MCMC subsequently. For future research we plan to relax the sequential sampling via conditional probabilities controlling which parameter to sample next.

## 5 RESULTS

The input data for our experiments consists of wide-baseline image sequences taken unconstrained with a hand-held consumer

digital camera. As output, the major branching structure of the target tree is represented in the form of a VRML (Virtual Reality Modeling Language) model. Focusing on urban scenes, we mainly analyze monopodial –  $m$  and sympodial-monochasium –  $sm$  trees, which form the majority of urban trees.

A result for an image triplet depicting an  $sm$  tree is presented in Fig. 3. A snapshot taken at the end of the search process shows the extracted branches. The search is limited to the thicker branches, which are modeled by the L-system and stops when no more hypotheses are accepted. Fig. 4 and 5 show an  $sm$  and an  $m$  tree, the latter taken under very different lighting conditions. The VRML models are presented from positions corresponding to the imaging positions to allow for an easy comparison. Since we model both the trunk and the branches as cylinders, the arc-shape of the trunk and the main branches of the tree in Fig. 4 is roughly represented as consecutive cylinders.

We basically have determined the main structure on the first several branching levels of the trees even for partially occluded branches as, e.g., the lowest left branch of Fig. 5. However, thinner branches and twigs are still missing, because they are not yet described by the L-system.

## 6 CONCLUSIONS

We have presented extensions to our approach (Huang and Mayer, 2007) for the 3D extraction of unfoliated trees from image sequences. The approach combines the descriptive power for trees of L-systems with statistical sampling and cross correlation into a generative statistical approach. A more sophisticated modeling by L-systems is devised for an improved interpretation of trees. Branching structures of trees are classified into different types after the extraction of the first level of branches and specific L-systems are applied for refined modeling. For the statistical sampling of the sparsely distributed branching parameters we combine MC and MCMC together with the Metropolis-Hastings algorithm for a more robust and efficient sampling. With the approach we are able to extract major branches to represent the basic structure of individual trees even when they are partly occluded and the images are taken under different lighting conditions. Preliminary results show the feasibility and potential of the approach.

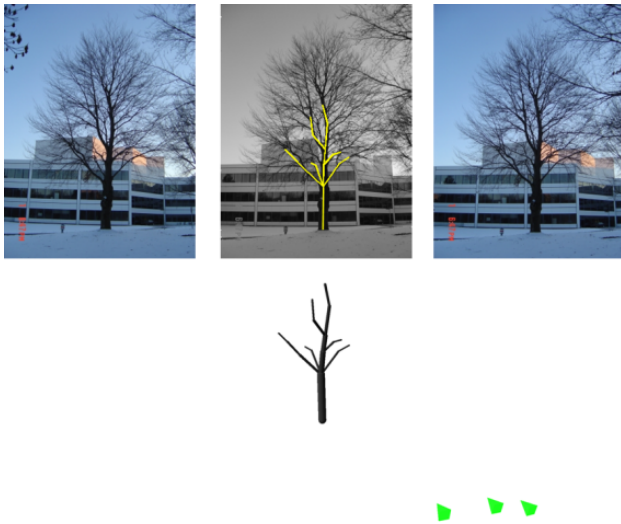


Figure 3: Result for a sympodial-monochasium tree limited to the trunk and the first several levels of branches – image triplet (top), a snapshot of the search process (top-center), and the result as a VRML model (bottom). The cameras are shown as (green) pyramids.



Figure 4: Result for a sympodial-monochasium tree limited to the trunk and the first several levels of branches – image triplet (left) as well as the VRML model (right) seen from the corresponding camera positions.

Our future research aims at more complex, e.g., context-sensitive, L-systems based on a more detailed geometrical modeling with generalized cylinders for the reconstruction of the branching struc-

ture of trees. The estimation of empirical distributions for the parameters from a larger number of examples is expected to lead to better results and more efficient modeling. Those parameters could include contraction rates for length, diameters and branching angles.

By correlating against trees and representative samples of the background, a function to upgrade correlation coefficients to correct likelihoods could be determined. Reversible Jumps - RJ (Green, 1995) are to be integrated in conjunction with model selection into MCMC. Search is thus optimized by an informed selection of competing hypotheses while at the same time avoiding overfitting.

We finally note that generative statistical modeling is not confined to L-systems. One basically just needs a means to construct realistic looking trees that can be efficiently controlled. For this, e.g., also (Lintermann and Deussen, 1999) could be basis. As shown above, the basic branching type of the object tree is classified during the extraction of main branches. The Production Rules of L-system could also be refined according to the already found structure. We assume that the upper stages of branches with very thin twigs might be grown stochastically with the derived Production Rules and parameters to match the image density.

#### ACKNOWLEDGEMENTS

We thank Deutsche Forschungsgemeinschaft for supporting Hai Huang under grant MA 1651/12-1 and the anonymous reviewers for their helpful comments.

#### REFERENCES

- Cheng, L., Caelli, T. and Sanchez-Azofeifa, A., 2006. Component Optimization for Image Understanding: A Bayesian Approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 28(5), pp. 684–693.
- Deussen, O. and Lintermann, B., 2005. Digital Design of Nature. Springer, Berlin, Germany.
- Gorte, B. and Pfeifer, N., 2004. Structuring Laser-scanned Trees Using 3d Mathematical Morphology. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 35(B5), pp. 929–933.
- Green, P., 1995. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika 82, pp. 711–732.
- Hartley, R. and Zisserman, A., 2003. Multiple View Geometry in Computer Vision – Second Edition. Cambridge University Press, Cambridge, UK.
- Huang, H. and Mayer, H., 2007. Extraction of the 3d Branching Structure of Unfoliated Deciduous Trees from Image Sequences. Photogrammetrie – Fernerkundung – Geoinformation (6/2007), pp. 429–436.
- Hyypä, J., Mielonen, T., Hyypä, H., Maltamo, M., Yu, X., Honkavaara, E. and Kaartinen, H., 2005. Using Individual Tree Crown Approach for Forest Volume Extraction with Aerial Images and Laser Point Clouds. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 36 (3/W19), pp. 144–149.

Lintermann, B. and Deussen, O., 1999. Interactive Modeling of Plants. IEEE Computer Graphics and Applications 19(1), pp. 2–11.

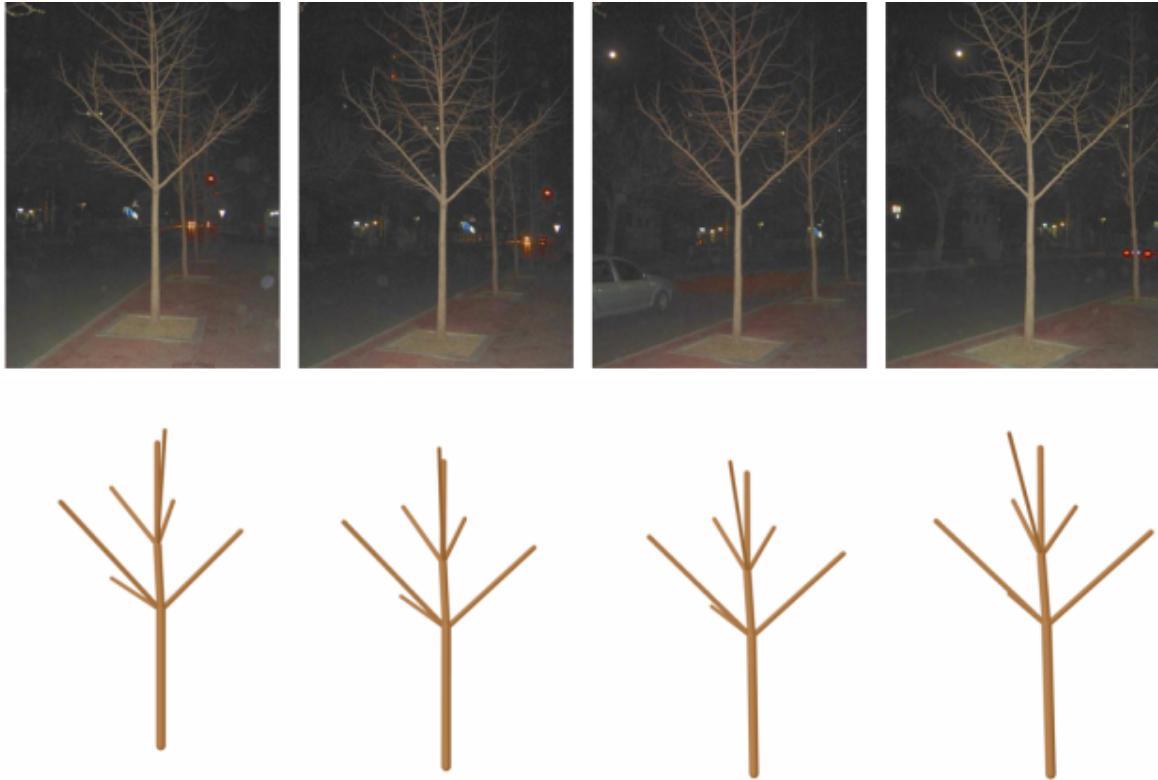


Figure 5: Result for a monopodial tree limited to the trunk and the first several levels of branches – image quadruple taken at night (top) as well as the VRML model (bottom). Please note the occluded branch on the lower left.

Mayer, H., 2005. Robust Least-Squares Adjustment Based Orientation and Auto-Calibration of Wide-Baseline Image Sequences. In: ISPRS Workshop in conjunction with ICCV 2005 “Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images” (BenCos), Beijing, China, pp. 1–6.

Měch and Prusinkiewicz, P., 1996. Visual Models of Plants Interacting with Their Environment. In: SIGGRAPH '96, pp. 397–410.

Neal, R., 1993. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto.

Nistér, D., 2004. An Efficient Solution to the Five-Point Relative Pose Problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(6), pp. 756–770.

Persson, Å., Holmgren, J., Söderman, U. and Olsson, H., 2004. Tree Species Classification of Individual Trees in Sweden by Combining High Resolution Laser Data with High Resolution Near-infrared Digital Images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 (8/W2), pp. 204–207.

Pfeifer, N., Gorte, B. and Winterhalder, D., 2004. Automatic Reconstruction of Single Trees from Terrestrial Laser Scanner Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 35 (B5), pp. 114–119.

Sakaguchi, T. and Ohya, J., 1999. Modeling and Animation of Botanical Trees for Interactive Environments. In: Symposium on Virtual Reality Software and Technology, pp. 139–146.

Shlyakhter, I., Rozenoer, M., Dorsey, J. and Teller, S., 2001. Reconstructing 3D Tree Models from Instrumented Photographs. *IEEE Computer Graphics and Applications* 21(3), pp. 53–61.

Tan, P., Zeng, G., Wang, J., Kang, S. and Quan, L., 2007. Image-based Tree Modeling. *ACM Transaction on Graphics*.