# Extraction of 3D Unfoliaged Trees from Image Sequences via a Generative Statistical Approach

Hai Huang and Helmut Mayer

Institute of Photogrammetry and Cartography, Bundeswehr University Munich
Hai.Huang|Helmut.Mayer@unibw.de
http://www.unibw.de/ipk

**Abstract.** In this paper we propose a generative statistical approach for the three dimensional (3D) extraction of the branching structure of unfoliaged deciduous trees from urban image sequences. The trees are generatively modeled in 3D by means of L-systems. A statistical approach, namely Markov Chain Monte Carlo – MCMC is employed together with cross correlation for extraction. Thereby we overcome the complexity and uncertainty of extracting and matching branches in several images due to weak contrast, background clutter, and particularly the varying order of branches when projected into different images. First results show the potential of the approach.

## 1 Introduction

Trees are an essential component of three dimensional (3D) urban information. They add a natural touch and influence the character of an urban scene considerably. Because of their difficult and thus costly acquisition, they are often neglected in 3D urban data sets. This is particularly true if their partly very distinctive shape and texture is to be represented.

Our basic goal is to extract and 3D reconstruct individual unfoliaged deciduous trees from image sequences. Deciduous trees are popular in cities worldwide further away from the equator, as they provide shadow in summer and yet let through most of the light in winter. Thus, they often form the majority of trees in urban areas. From a practical point of view images for data acquisition in cities will often be taken when the trees are unfoliaged as facades etc. are then more readily visible. For us this has the big advantage that one can directly see the branching structure which one can only guess from the foliaged tree.

From a scientific point of view extracting the branching structure in 3D by matching from multiple images is a difficult problem nobody to our knowledge has ever even tried to solve. Extraction and matching of branches is difficult because of bad contrast, clutter by background objects, and because the order of the branches even in neighboring images can vary considerably due to the pronounced 3D structure of trees.

Former work has mostly dealt with tree extraction in aerial images and particularly recently laser scanner data. Much work focuses on forests. The only approach we mention here is [1] as they also use a statistical (Reversible Jump) Markov Chain Monte Carlo – (RJ)MCMC [2] approach to model trees, in their case by a spatial point process.

Work for terrestrial urban images is more limited. In [3] foliaged deciduous trees are segmented in color images based on their texture. There is neither a segmentation of individual trees nor any 3D interpretation. Also [4] focuses only on a two dimensional (2D) interpretation, yet for individual trees. The model is based on the particular symmetries of coniferous trees. [5] is mostly concerned with the animation of trees. For their 3D extraction first a volume is generated by intersecting the view cones resulting from the tree silhouettes in multiple images. The voxels of the volume are colored with the average brightness from the rays from the different images. A branching process is started at 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. The most sophisticated approach today is [6]. 3D volumes are generated as in [5]. From the volumes 3D medial axes are constructed. The medial axes are constrained to "botanical fidelity of the branching pattern and the leaf distribution" [6] via an open Lindenmayer-, or in short L-system [7]. Again, a lot of manual interaction is employed to generate results which are good in terms of visualization.

In this paper we show how by means of generative statistical modeling it becomes feasible, to match branches in wide-baseline image sequences taken unconstrained with a standard consumer camera in spite of the problems stated above. We assume, that we can orient images highly precisely by an automatic orientation procedure in the spirit of [8], yet making use of calibration via the five-point-algorithm [9], determining matches by a least-squares procedure highly precisely and bundle-adjusting everything [10]. We note, that our modeling should be useful to find trees in much more explicit laser-scanner data, though the latter is linked to more effort for data acquisition.

In Section 2 the basic idea of generative statistical extraction employing L-systems for the modeling of the 3D characteristics of trees is described. We use statistical sampling in the form of MCMC to generate the parameters of an L-system which comply with the data as described in terms of likelihood.

The generation of 3D hypotheses, their 2D projection and verification are described in Sections 3 and 4. Hypotheses for trunks are generated from vertical lines matched in several images. For the branches suitable prior distributions for the parameters are discussed particularly focusing on issues with the branching angles. The verification of new hypothesis is currently done using the (normalized) cross correlation coefficient ($CCC$) as (a substitute for) likelihood. After presenting fist results which show the potential of the approach in Section 5, the paper ends up with conclusions.

## 2 Generative Statistical Extraction Using L-Systems

Branches of trees are difficult to extract from terrestrial urban image sequences due to their weak contrast and background clutter from other objects, e.g., facades or other trees. As we want to construct 3D models of trees, we need to match the branches. Because of the complex 3D structure of trees, 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, often employed to guide matching, is often not valid even for images taken close to each other. All this means that the bottom-up extraction of branches and matching them in 3D does not seem promising and suitable constraints describing the structure of trees are needed for their 3D reconstruction.

In our case the structure of trees is described in terms of their growing, or more particularly branching, by an L-system [7]. It is a parallel string rewriting system representing branching structures in terms of bracketed strings of symbols with associated numerical parameters, called modules. The simulation of branching starts with an initial string (axiom) and proceeds in a sequence of steps. By means of productions all modules in the predecessor string are substituted by successor modules. Whether a production is applicable can depend on a predecessor's context, values of parameters, and like in our case on random factors (also called stochastic L-systems). By means of context-sensitive L-systems interactions between plant parts can be represented. We do not use this for our simple first prove-of-concept implementation described in this paper, although it would certainly be helpful. By recursively using the same productions, L-systems represent self-similarity, an important biological characteristic of plants.

The modeling with L-systems results in tree-like structures. Yet, L-systems alone only give means to generate and also visualize trees. For their extraction from images, they need to be linked to a means for extraction. We decided to employ a generative statistical approach based on MCMC and L-systems, where likely candidates of branches are generated by stochastical sampling and are verified by comparing simulated and real images.

Figure 1 presents the basic idea of our approach. After extracting the trunk as described below, branches are grown randomly guided by appropriate prior distributions and are projected into the images via the given highly precisely known orientation parameters. The simulated images are then matched to the given images. As model for the background clutter we use Gaussian noise.

Linking stochastic sampling, L-systems, and likelihood from the images renders it possible to find a tree structure very similar to that of the real tree. I.e., while L-system and MCMC alone can produce a typical tree, e.g., a beech, the link with the likelihood in the images results into a beech with the particular characteristics that can be seen in the images.

## 3 3D Hypotheses Generation

While we focus on the branches, the basic part of many trees and particularly those we are interested in is the trunk. For it we extract lines, assuming that trunks correspond to thick, mostly vertical lines. The vertical direction is presumed to be known approximately by basically taking images horizontally. It usually can be improved by computing the vertical vanishing point from the

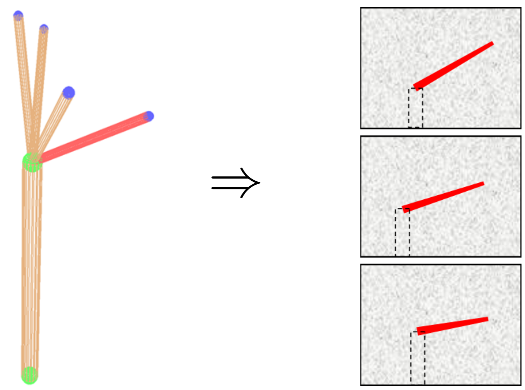

**Fig. 1.** left: Stochastical sampling based on an L-system results in a 3D tree hypothesis – left / right: Projection of a new branch (red) into three empty images with randomly textured background

vertical edges of trunks or on facades as we focus on urban scenes. Found vertical lines, i.e., hypotheses for trunks, are verified by matching in several images. We use the trifocal tensor [11] derived from our highly precisely known orientation parameters to predict from lines in two images a hypotheses for a line in a third image. We further assume that the position of the tree is determined by the trunk.

The scope of this proof-of-concept paper is limited to the first several levels of branches. We right now assume, that the upper stages of branches with very thin twigs might be grown stochastically to just match the image density, but it is to be seen if and on which level of branching this is a valid assumption.

A new branch is modeled in 3D object-space as a cylinder with known begin and the following parameters (the vertical direction is assumed to be approximately known (cf. above) and the x- and y-axis are taken from the local coordinate system of the first camera after aligning it with the vertical direction):

- Azimuth: angle with x-axis of branch projected into horizontal plane
- Inclination: angle between branch and horizontal plane
- Length
- Diameter

MCMC should basically sample the azimuth with a uniform distribution between 0° and 360° (cf. blue horizontal dashed line in Fig. 2, right). However, if a limited number of images is taken forming an acute angle together with the trunk as, e.g., the blue area for the three images in Fig. 2, left, accepted branches trend to concentrate in the area of the acute angle. The reason for this is, that in the center of the tree there is usually the vertical trunk, or there are at least only very few thicker vertical branches. All branches with whatever inclination generated in a vertical plane inside the acute angle or close to it will be more likely accepted as they are projected onto the trunk in all given views and are not disambiguated from another viewing angle. One solution that we have devised for this consists in a modified prior distribution for the azimuth as given as red line in Fig. 2, right.

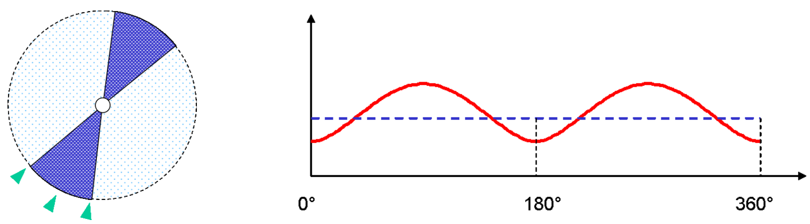


**Fig. 2.** left: Area (blue) where for a smaller number of closely-spaced cameras a concentration of hypotheses can occur as hypotheses are not disambiguated from a larger viewing angle – right: Prior distribution (red) for azimuth (with 0° at central camera) helps to reduce the concentration – original uniform distribution as blue dashed line

In combination with the azimuth only a half circle is needed for the inclination (cf. Fig. 3). Moreover, for most types of trees, the majority of branches look upwards. We thus have devised a prior distribution for the inclination shown in Fig. 3 with darker color denoting higher probabilities.

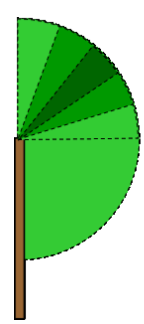

**Fig. 3.** Prior distribution of inclination – Darker color means higher probability

For length and diameter normal distributions are considered. Our first experiments were conducted with a mean of 1 meter for the length for the first level of branches. The diameter is set to a fixed value. For the higher levels of branches we use contraction coefficients.

## 4 2D Projection and Evaluation

The generated hypotheses are projected into the 2D images, to be evaluated there by comparing the simulated images constructed from the projected hypotheses with the given images. The projection of 3D cylinders entails a larger computational effort. As we do many of these projections in MCMC, we decided for the proof-of-concept prototype, where we did not want to use a graphical processing unit (GPU) due to missing experience with its programming, to use a simple and efficient 2D representation derived from the 3D representation. Another reason for this is that the projection of the branches results into patches of nearly constant brightness anyhow. The chosen 2D representation consists of trapezoids. The color is taken as average of the trunk. A trapezoid is described by the following four parameters:

- Direction: angle with x-axis
- Length
- Width of begin
- Width of end

The parameters of the trapezoid are obtained in the following way: The centers of the begin and the end are obtained by projecting the centers of the circles, i.e., the end points of the axis, delimiting the cylinder on both sides, into the image via

$$\mathbf{x}' = \mathbf{P}\mathbf{X}$$

with (homogeneous) 3D points $\mathbf{X}$, image points $\mathbf{x}'$, and the projection matrix $\mathbf{P}$ [11]. To compute reasonable approximations for the widths, we connect the end points of the axis of the cylinder with the camera center and determine one of the normals to this vector. The distance between the projections of the end point of the axis and of the intersection point of the normal with the cylinder surface equals half the width in the image.

The projection of a hypothesis is compared with the corresponding original image $i$ by means of the cross correlation coefficient $CCC_i$ for the intensities computed by HSI color transformation. To compare different hypotheses, the whole images have to be compared with the projections of the complete 3D models. As MCMC sampling consists of 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 been changed.

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

$$CCC = \prod_{i=1}^{n} CCC_i \quad .$$

Multiplication is used as we interpret the $CCC_i$ values as likelihoods and we assume independence of the images given the 3D model. Additionally, we found empirically that this conservative combination helps to sort out bad hypotheses early. We are aware that the actual size of $CCC_i$ values can be far from correct likelihoods. Yet, our experiments give evidence to assume that they are proportional to correct likelihoods. A function linking raw $CCC_i$ values and likelihoods could be obtained by determining a statistics of $CCC_i$ values for a larger number of known correct and incorrect hypotheses for branches at a certain level. This is subject of further work.

We found empirically that it is not useful to sample all parameters of a branch at the same time. Thus, the MCMC sampling of the parameters is done sequentially. First, only azimuth and inclination are jointly varied over 1000 iterations while the length is kept fixed. The latter is optimized only afterwards with 500 iterations. In future work we plan to relax the sequential sampling via conditional probabilities controlling which parameter to sample next.

## 5 Results

Fig. 4 and 5 show first results. The input data consists of an image triplet for the former and an image quadruple for the latter, both taken unconstrained with a hand-held 5 Megapixel camera. As output we obtain a VRML (virtual reality modeling language) model describing the trunk and the first two levels of the main branching system of the trees.

The scenes, taken on different continents under very different lighting conditions demonstrate, that we can basically determine branches on the first two levels. Yet, we note that our proof-of-concept implementation still misses branches and reaches only a limited accuracy. We assume that with more experience and particularly by using more levels and RJMCMC, the latter to dynamically generate and delete competing hypotheses via the jumps, we will be able to drop most wrong hypotheses as they will be substituted by better fitting hypotheses.

## 6 Conclusions

We have proposed a generative statistical approach for the extraction of the branching system of unfoliaged trees. By combining the descriptive power for trees of L-systems with statistical sampling by means of MCMC and simple cross correlation we are able to extract partly occluded branches with possibly weak contrast from image sequences as shown by our first results.

Concerning future work, we first want to generalize the implemented L-system in the direction of open L-systems [7]. We might need to change the parameterization away from the vertically centered azimuth and inclination angles to a more local representation based on branching angles. Yet, we note that generative statistical modeling is not confined to L-systems. We basically just need a means to construct realistic trees that can be efficiently controlled. For this, e.g., also [12] could be a good basis.

Parameters such as contraction rates or branching angles could be learned by extracting a larger number of trees leading to priors probably conditional to the branching level. As already noted above, by correlating against trees

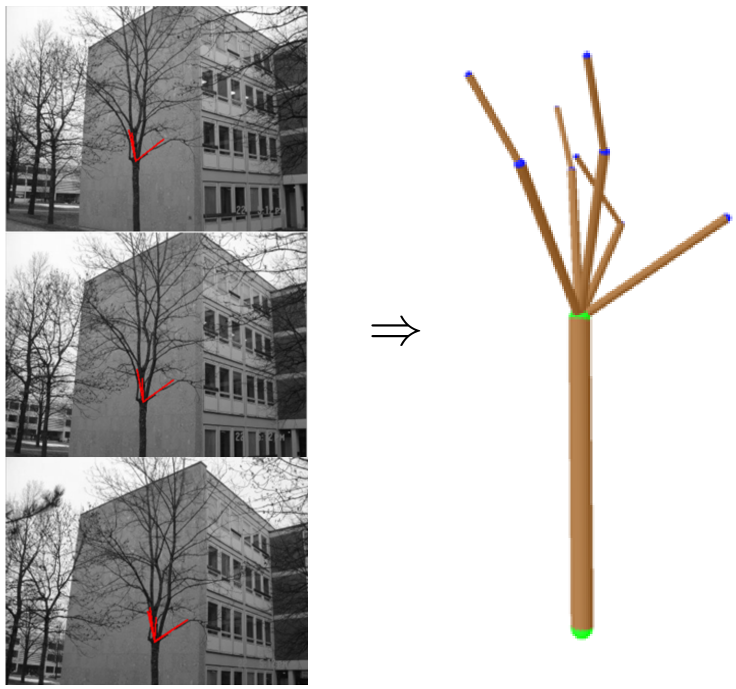

**Fig. 4.** Extraction from an image triplet limited to the the trunk and the first two levels of branches – intermediate stage of processing projected into images (left) and final result (right)

and representative samples of the background, a function to upgrade correlation coefficients to likelihoods could also be learned.

An important question will be to decide, how many branches are to be formed on a level and how many levels are appropriate for the tree, i.e., to control the complexity. E.g., for our proof-of-concept implementation, if we had not limited the branching level, a small tree could keep growing, even though new branches are just hallucinated into the background, as there is no obvious way of stopping. This leads to the issue of model selection. The idea is to balance the complexity of a hypothesis, i.e., the size of the tree or more particularly the number of parameters, against its likelihood according to the data. For this, the theory developed for compositional systems [13] might prove helpful, possibly also in conjunction with RJMCMC, to dynamically add and delete hypotheses, the latter, if better solutions evolve.

## Acknowledgment

We thank Deutsche Forschungsgemeinschaft for supporting Hai Huang under grant MA 1651/11-1.

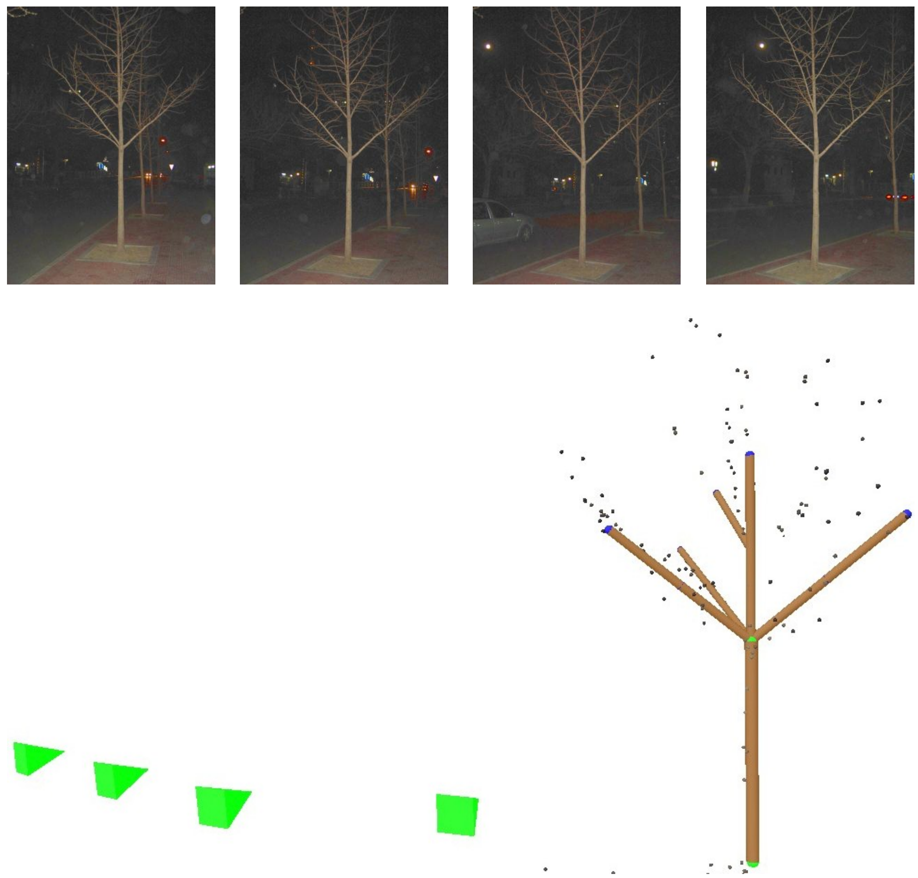

**Fig. 5.** Extraction from an image quadruple limited to the trunk and the first two levels of branches – images (top) and result (bottom) showing the tree, the cameras as (green) pyramids and points used by the orientation procedure

## References


1. Andersen, H.E., Reutebuch, S., Schreuder, G.: Bayesian Object Recognition for the Analysis of Complex Forest Scenes in Airborne Laser Scanner Data. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Volume (34) 3A. (2002) 35–41
2. Green, P.: Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. Biometrika **82** (1995) 711–732
3. Haering, N., Myles, Z., Vitoria, N.: Locating Deciduous Trees. In: IEEE Workshop on Content Based Access of Image and Video Libraries. (1997) 18–25
4. Forsyth, D., Malik, J., Fleck, M., Greenspan, H., Leung, T., Belongie, S., Carson, C., Bregler, C.: Finding Pictures of Objects in Large Collections of Images. In:

Object Representation in Computer Vision II, Berlin, Germany, Springer-Verlag (1996) 335–360
5. Sakaguchi, T., Ohya, J.: Modeling and Animation of Botanical Trees for Interactive Environments. In: Symposium on Virtual Reality Software and Technology. (1999) 139–146
6. Shlyakhter, I., Rozenoer, M., Dorsey, J., Teller, S.: Reconstructing 3D Tree Models from Instrumented Photographs. IEEE Computer Graphics and Applications **21**(3) (2001) 53–61
7. Měch, Prusinkiewicz, P.: Visual Models of Plants Interacting with Their Environment. In: SIGGRAPH '96. (1996) 397–410
8. Pollefeys, M., Van Gool, L., Vergauwen, M., Verbiest, F., Cornelis, K., Tops, J.: Visual Modeling with a Hand-Held Camera. International Journal of Computer Vision **59**(3) (2004) 207–232
9. Nistér, D.: An Efficient Solution to the Five-Point Relative Pose Problem. IEEE Transactions on Pattern Analysis and Machine Intelligence **26**(6) (2004) 756–770
10. Mayer, H.: 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. (2005) 1–6
11. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision – Second Edition. Cambridge University Press, Cambridge, UK (2003)
12. Lintermann, B., Deussen, O.: Interactive Modeling of Plants. IEEE Computer Graphics and Applications **19**(1) (1999) 2–11
13. Geman, S., Potter, D., Chi, Z.: Composition Systems. Quarterly of Applied Mathematics **LX** (2002) 707–736