Extraction of the 3D Branching Structure of Unfoliaged Deciduous Trees from Image Sequences

HAI HUANG & HELMUT MAYER, Neubiberg

Keywords: 3D Tree Extraction, Branching Structure of Trees, Statistical Generative Models, L-Systems, Markov Chain Monte Carlo – MCMC

Summary: In this paper we propose an approach for the three-dimensional (3D) extraction of the branching structure of unfoliaged deciduous trees from urban wide-baseline 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 the extraction of the branches. With this generative statistical approach we avoid 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.

Zusammenfassung: *Extraktion der 3D Verzweigungsstruktur unbelaubter Laubbäume aus Bildsequenzen.* Dieses Papier stellt einen Ansatz für die Extraktion der drei-dimensionalen (3D) Verzweigungsstruktur unbelaubter Laubbäume aus städtischen Bildsequenzen mit langer Basis vor. Die Bäume werden in 3D mittels L-Systemen modelliert. Markoff Ketten Monte Carlo (MCMC) wird zusammen mit Kreuzkorrelation für die Extraktion der Äste genutzt. Mit diesem generativen statistischen Ansatz wird die Komplexität und Unsicherheit der Extraktion und Zuordnung von Ästen in mehreren Bildern wegen schwachem Kontrast, Störobjekten im Hintergrund und insbesondere der z.T. unterschiedlichen Ordnung der Äste nach Projektion in verschiedene Bilder vermieden. Erste Ergebnisse zeigen das Potential des Ansatzes.

1 Introduction

In our environment trees play an essential role. This is particularly true in urban areas, where they are too often the only prominent representatives of nature. Because of their complex structure their acquisition is costly. Thus, they are often neglected, or at least only acquired in a very simplified form for geoinformation systems (GIS), especially for three-dimensional (3D) city models. The distinctive shape and texture of some of the trees that influence the appearance of their whole environment is only represented for very locally limited architectural models.

In this paper we aim at extracting the 3D branching structure of individual unfoliaged deciduous trees from wide-baseline image sequences. Deciduous trees are popular in cities worldwide 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.

From a scientific point of view extracting the 3D branching structure of unfoliaged trees by matching in multiple images is a difficult problem nobody to our knowledge has tried to solve yet. 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 especially recently laser scanner data. Much work focuses on forests. Work for tree extraction from terrestrial urban images is scarcer. HAERING et al. (1997) segment groups of foliaged deciduous trees in color images based on texture without any 3D interpretation. Also FORSYTH et al. (1996) focus only on a two-dimensional (2D) interpretation, yet for individual trees. They particularly model the symmetries of coniferous trees. (SAKAGUCHI & OHYA 1999) is mostly dealing with the animation of trees, but is one of the few papers actually concerned with the 3D extraction of trees. 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 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 automatic approach today is arguably (SHLYAKHTER et al. 2001). 3D volumes are generated as in (SAKAGUCHI & 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 & PRUSINKIEWICZ 1996). Again, manual interaction is employed to generate results which are good in terms of visualization.

We show how generative statistical modeling based on L-systems and Markov Chain Monte Carlo – MCMC makes it feasible, to match branches in wide-baseline image sequences taken unconstrained with a standard consumer camera in spite of the problems with clutter and occlusions stated above. Our basis is a procedure (MAYER 2005) for the highly precise automatic determination of the orientation of images 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 basic idea of generative statistical extraction employing L-systems and MCMC is described. We use statistical sampling in the form of MCMC to generate the parameters of an L-System modeling the 3D characteristics of trees which comply with the data evaluated in terms of likelihood. The generation of 3D hypotheses, their 2D projection, and evaluation are described in Sections 3 and 4. Hypotheses for trunks are generated from approximately vertical lines matched in several images. For the branches suitable prior distributions for the parameters are discussed particularly focusing on the branching angles. The evaluation of new hypotheses is conducted using the (normalized) cross correlation coefficient (*CCC*) as a substitute for likelihood. After presenting first results demonstrating the potential of the approach in Section 5, the paper ends up with conclusions.

2 Generative Statistical Extraction Using L-systems and MCMC

The branching structure of trees is difficult to extract from terrestrial wide-baseline urban image sequences because of possibly weak contrast and background clutter from other objects, e.g., facades or other trees. To construct 3D models of trees, 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 even for images taken close to each other. All this means that the bottom-up / data-driven extraction of branches and matching them in 3D does not seem promising and suitable constraints describing the structure of trees are essential for their 3D reconstruction.

We describe the structure of trees in terms of their growth, or more particularly branching, by a Lindenmayer-, or in short L-system (MĚCH & PRUSINKIEWICZ 1996). 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). 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 termed stochastic L-systems). By means of context-sensitive L-systems interactions between plant parts can be represented. We do not use this for our first proof-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.

Basically, branching structures of trees can be divided into two main groups for which different production rules have to be used: monopodial and sympodial (DEUSSEN & LINTERMANN 2005). The monopodial branching system (cf. Figure 1 (a)) has a prominent main axis, which is stronger and longer than the side branches. The side branches are again stronger and longer than the side branches of the second order, etc. Because of the dominant axes monopodial branching structures have a radially symmetric crown.

Figure 1 (b) and (c) show the two main types of sympodial branching. Sympodial, dichasium 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 branching, 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.

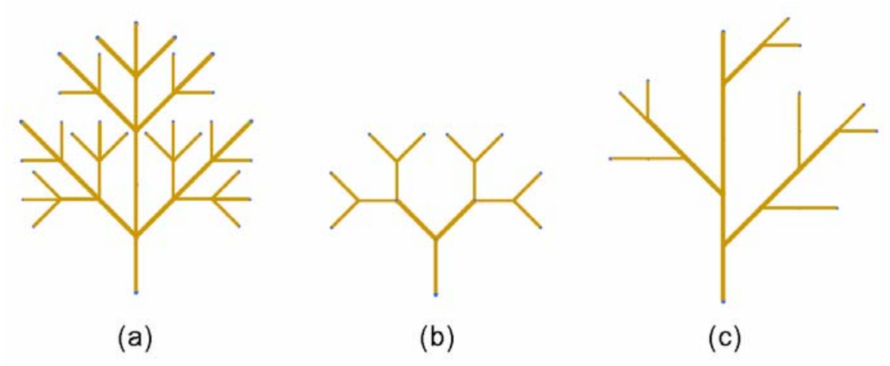

Fig. 1: Types of branching structures: (a) monopodial (b) sympodial, dichasium (c) sympodial, monochasium.

For the sake for flexibility, we employ at the moment a mixture of all three sorts of branching: We let the branches sprout at the end of the trunk or a branch in all possible directions, yet preferring inclinations around 45° via a prior function (cf. below).

Modeling with L-systems enforces tree-like branching 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, where likely candidates of branches are generated by stochastical sampling and are verified by comparing simulated and real images.

In Figure 2 the basic idea of our approach is presented. After extracting the trunk as described below, branches are grown randomly guided by appropriate prior distributions and are then projected into the images via the given highly precisely known orientation parameters. The hereby generated simulated images are matched to the given images. As model for the background clutter we use Gaussian noise.

By linking stochastic sampling, L-Systems, and likelihood from the images we find a tree structure very similar to that of the real tree. While L-system and MCMC can produce a typical tree, e.g., a beech, the link with the likelihood generated by matching with the images results into a beech with the particular characteristics that can be seen in the given images.

3 Generation of 3D Hypotheses

While we focus on the branching structure, a basic part of many trees we are interested in is the trunk. For it we extract straight 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 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 & ZISSERMAN 2003) derived from the known orientation parameters to predict from lines in two images hypotheses for lines in other images. For the remainder of the paper we assume that the position of the tree is determined by the trunk.

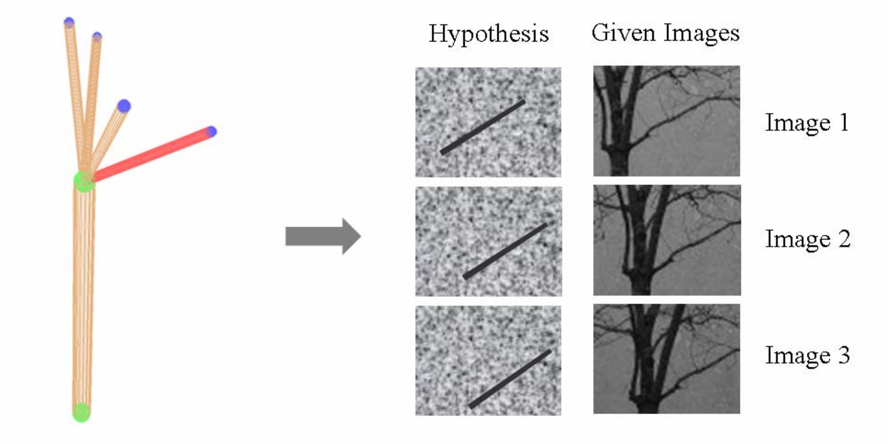

Fig. 2: Stochastical sampling based on an L-system results in a 3D tree hypothesis (left). Projection of a new branch (red) into three empty images with randomly textured background (center) and given image data (right). For the sake of clarity only the projection of the new branch is shown.

This proof-of-concept paper is limited to the first two levels of 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.

We assume that the vertical direction is approximately known (cf. above). The x- and y-axis are taken from the local coordinate system of the first camera after aligning it with the vertical direction. The azimuth is sampled by MCMC with a uniform distribution between 0° and 360°. Because of the resulting symmetry, only a half circle is needed for the inclination (cf. Fig. 3). For most types of trees the majority of branches points upwards. We thus have empirically devised a prior distribution for the inclination with highest probability around 45°.

For length and diameter normal distributions are considered. Our first experiments were conducted with means between 0.7m and 1.5m for the length for the first level of branches. The diameter is set to a fixed value.

4 Projection and Evaluation

The hypotheses generated above are projected into 2D resulting into simulated images. They are evaluated by comparing the simulated images with the given images. The projection of 3D cylinders entails a larger computational effort. For MCMC many of these projections are needed. We thus decided for the proof-of-concept prototype, where we did not want to use a graphics 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 doing so 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 color. A trapezoid is described by the parameters direction (angle with x-axis), length, width of begin, and width of end.

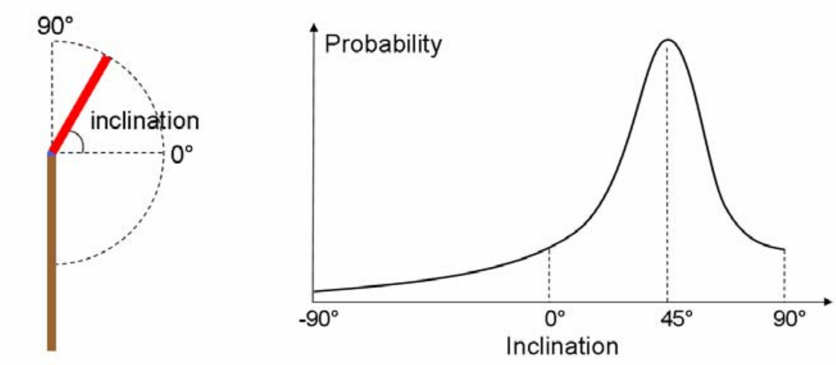

Fig. 3: left: Inclination of branch (red) – right: Empirically determined prior distribution.

We determine the parameters of the trapezoid as follows: 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 **X**, image points **x'**, and the projection matrix **P** (HARTLEY & ZISSERMAN 2003). To compute reasonable approximations for the widths, we connect each end point of the axis of the cylinder with the camera center and determine a normal to this vector. The distance between the projections of the end point of the axis and of a point on the normal with distance radius of the cylinder from the axis equals half the width in the image.

The projected 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 be able to compare different hypotheses, the matching is done against the projections of the convex 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 been changed.

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

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

We use multiplication because we interpret the CCC_i values as 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 give a reasonable approximation to correct likelihoods. A function linking raw CCC_i values and likelihoods could be obtained by determining the statistics of CCC_i values for a larger number of known correct and incorrect hypotheses for branches at a certain level.

By means of experiments we found that it is not useful to sample all parameters of a branch at the same time. Thus, MCMC sampling of the parameters is conducted sequentially. First, only azimuth and inclination are jointly varied over 1000 iterations while the length is kept fixed. The length is optimized only afterwards with 500 iterations. For future research we plan to relax the sequential sampling via conditional probabilities controling which parameter to sample next.

5 Results

Figures 4 and 5 present first results. The input data consists of an image quadruple for the first and an image triplet for the second, 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.

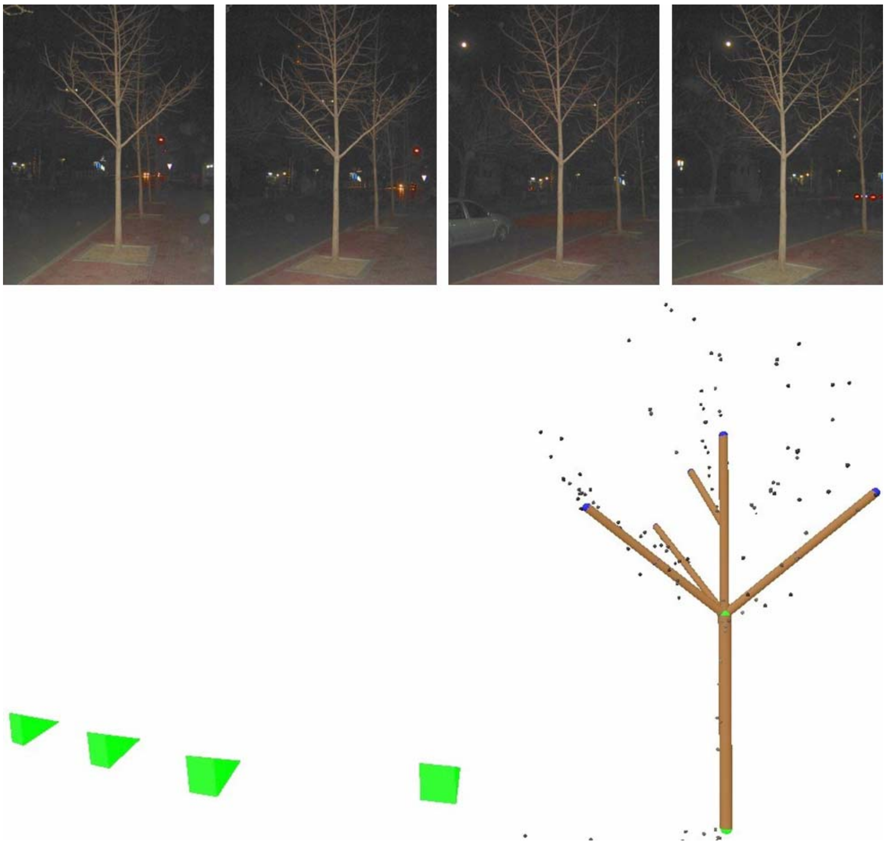

Fig. 4: Result for a Chinese image quadruple limited to the trunk and the first two levels of branches – Images (top) as well as the 3D tree, the cameras as (green) pyramids, and points used by the orientation procedure (bottom).

The scenes, taken in China and Germany under very different lighting conditions demonstrate, that we can basically determine the branching structure on the first two levels even though branches are partially occluded as, e.g., one of the left branches of the Chinese quadruple. Yet, we note that our proof-of-concept implementation still misses branches and reaches only a limited accuracy.

6 Conclusions and Discussion

We have proposed an approach for the extraction of the branching system of unfoliaged trees from wide-baseline image sequences. It combines the descriptive power for trees of L-Systems with statistical sampling by means of MCMC and cross correlation into a statistical generative approach. Using the approach we are able to extract branches even when they are partly occluded as demonstrated by our first results. The envisaged final result of our approach is the basic branching structure of a particular tree. It will allow very realistic visualizations, e.g., for movies, and one could add leaves with different colors for different seasons. The branches could even be animated by simulating the forces of wind on them. When analyzing many trees, the resulting statistics for the parameters could be used for ecological applications or for simulating the interaction with radio waves for synthetic aperture radar (SAR).

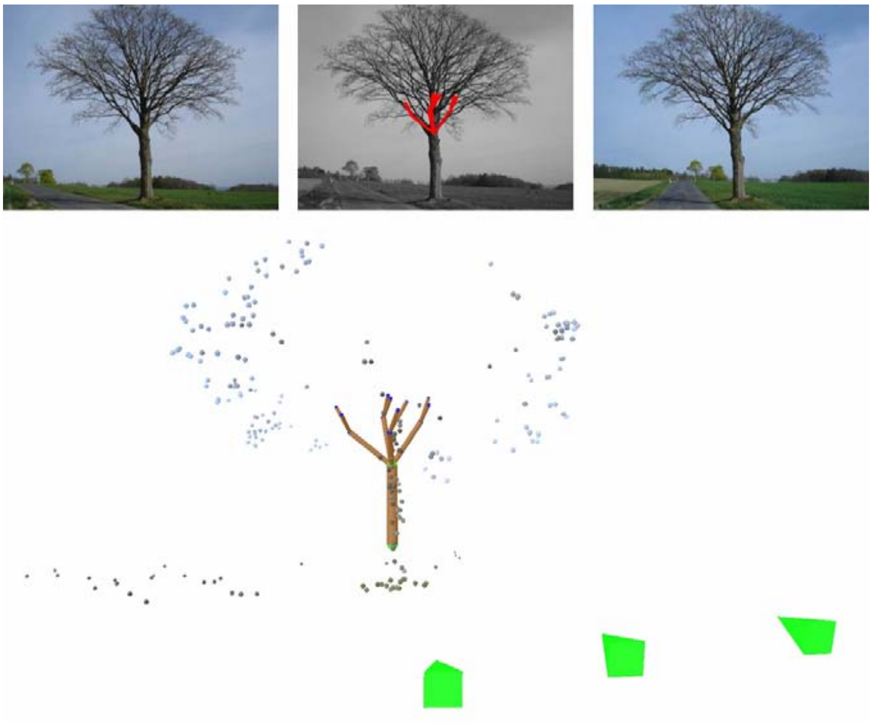

Fig. 5: Result for an image triplet limited to the trunk and the first two levels of branches – branches projected into images (top) as well as the 3D tree, the cameras as (green) pyramids, and points used by the orientation procedure (bottom).

Concerning future research, we first want to generalize the implemented L-system in the direction of open L-systems (MĚCH & PRUSINKIEWICZ 1996) and distinguish between different types of trees (monopodial, sympodial; cf. above). We might need to change the parameterization away from the vertically centered azimuth and inclination angles to a more local, context-based representation using branching angles.

We also 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 (LINTERMANN & DEUSSEN 1999) could be a basis. 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 has to be seen if and on which level of branching this is a valid assumption.

It should be possible to learn parameters such as contraction rates for lengths and diameters or branching angles by extracting a larger number of trees leading to priors probably conditional to the branching level. As noted above, by correlating against trees and representative samples of the background, a function to upgrade correlation coefficients to likelihoods could also be learned.

One question which arises is, 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. If there are other trees or facades with strong linear textures in the background, there will be a strong tendency, that too many and thus too dense branches will be estimated and that they also extend beyond the perimeter of the tree. We want to tackle this issue by means 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 (GEMAN et al. 2002) might prove helpful, possibly also in conjunction with reversible jump (RJ) MCMC (GREEN 1995), to dynamically add better and delete worse hypotheses, the latter, if better solutions evolve.

Finally, 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.

Acknowledgements

We thank Deutsche Forschungsgemeinschaft for supporting Hai Huang under grant Ma 1651/11-1 and the anonymous reviewers for their helpful comments.

References

DEUSSEN, O. & LINTERMANN, B., 2005: Digital Design of Nature. – Springer. Berlin, Germany.

FORSYTH, D., MALIK, J., FLECK, M., GREENSPAN, H., LEUNG, T., BELONGIE, S., CARSON, C. & BREGLER, C., 1996: Finding Pictures of Objects in Large Collections of Images. – Object Representation in Computer Vision II. Springer-Verlag, Berlin, Germany, 335–360.

GEMAN, S., POTTER, D. & CHI, Z., 2002: Composition Systems. – Quarterly of Applied Mathematics **LX**: 707–736.

GREEN, P., 1995: Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. – Biometrika **82**: 711–732.

HAERING, N., MYLES, Z. & VITORIA, N., 1997: Locating Deciduous Trees. – IEEE Workshop on Content Based Access of Image and Video Libraries, 18–25.

HARTLEY, R. & ZISSERMAN, A., 2003: Multiple View Geometry in Computer Vision – Second Edition. – Cambridge University Press. Cambridge, UK.

LINTERMANN, B. & DEUSSEN, O., 1999: Interactive Modeling of Plants. – IEEE Computer Graphics and Applications **19** (1): 2–11.

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

MĚCH & PRUSINKIEWICZ, P., 1996: Visual Models of Plants Interacting with Their Environment. – SIGGRAPH '96, 397–410.

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

SAKAGUCHI, T. & OHYA, J., 1999: Modeling and Animation of Botanical Trees for Interactive Environments. – Symposium on Virtual Reality Software and Technology, 139–146.

SHLYAKHTER, I., ROZENOER, M., DORSEY, J. & TELLER, S., 2001: Reconstructing 3D Tree Models from Instrumented Photographs. – IEEE Computer Graphics and Applications **21** (3): 53–61.

Address of the Authors

Dipl.-Ing. HAI HUANG, Prof. Dr.-Ing. HELMUT MAYER, Bundeswehr University Munich, Institute of Photogrammetry and Cartography, D-85577 Neubiberg, Tel.: +49-89-6004-3429, Fax: +49-89-6004-4090, e-mail: Hai.Huang@unibw.de, Helmut.Mayer@unibw.de

Manuskript eingereicht: Mai 2007
Angenommen: Juni 2007