

Parallel Implementation of Multiple Interdependent Lindenmayer Systems

JOSE JAVIER GONZALEZ ORTIZ*

University of Michigan

December 7, 2015

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

1 Introduction

Modern techniques for botanic and arboreal growth simulation rely in specific and dedicated models to correctly recreate the behavior of different types of plants. This creates a segmented and disconnected variety of models and techniques. As A. Lindenmayer and Przemyslaw described in [PL96], almost all botanical structures can be successfully model by Lindenmayer Systems.

However, raw L-Systems are not complex enough for this task and therefore we need to include both a parametric implementation to successfully model these types of structures. Moreover, parametric L-Systems in spite of depending on a number of initial parameters and conditions, they are deterministic. To cope with this limita-

tion, we introduce randomness in the traditional form described in [PL96], by using Stochastic-Parametric L-Systems. This systems will have production rules that will not only rely on several parameters to correctly scalate the complexity of the system but will also apply this rules using a source of randomness to choose between a number of specified rules.

Another aspect we have to properly take into a account is the fact that botanical growth almost never occurs in a isolated fashion, and therefore we will be interested in looking at a framework that can accommodate the simultaneous derivation of multiple L-Systems. Having multiple stochastic L-Systems produces a non-realistic result as we will see later, so to compensate for this caveat we had to define the concept of Interdependent L-System.

*jjgo@umich.edu

In this paper we investigate the use of Inter-dependent L-System to realistically model the growth of a forest and we analyze several parallel implementations of this problem in a shared memory machine using OpenMP. L-Systems are parallel rewriting grammar which initially would seem to simplify the problem, but this derivation step leads to highly uneven amounts of work in a great number of scenarios. Furthermore, as we shall see later, adding the interdependence to the systems serializes the problem since after every timestep each system will need to communicate a number of metric to a variable number of other systems. This problem will produce a scenario experienced by a great amount of local simulation techniques.

We will introduce several parallel algorithms with increasing complexity and with increasing speedup and compare their behavior under different types of datasets, which from a simulation perspective would translate into different ecosystems. The final algorithm will involve calculating the connected components of the forest graph to initially subdivide the problem in equally complex albeit not equally balanced subproblems, and introduce a general approach to tackle this subproblems. To accomplish this, the algorithm will deserialize the time dimension by allowing some L-Systems to carry out further iterations as long as the correctness constraints are met. We will also analyze the bigger bottleneck of the algorithm and the way it tries to cope with the variability of input systems.

Overview: First we will start by providing a background in canonical, parametric and stochastic L-Systems in Section 2. Section 3 will elaborate the model used to realistically model the problem and will discuss the serial implementation that was carried out. The parallel algorithms designed to solve this problem are discussed in Section 4 and their associated results are shown in Section 5 and analyzed in Section 6.

1.1 Previous Work

General L-Systems: Prusinkiewicz and Lindenmayer gave the basic definition to the L-System algorithm and structure in [PL96]. This was further extended by the work carried out in [PM01, PJM94, PMKL01]

Parallelizing L-Systems: Lacz and Hart introduced the use of manually written pixel shaders to compute L-Systems [LH04]. A distributed memory approach making use up to 8 CPUs and the Message Passing Interface (MPI) was described in [YHL⁺07]. An algorithm from the Parallel Generation of L-Systems was introduced by [LWW09]. The approach was appropriate for both GPU and multi-core CPUs, parallelizing both the derivation and interpretation of given L-Systems. The implementation provided was generic and supported parametric, stochastic and context sensitive productions. The work was further extended in [LWW10] to make the algorithm work with multiple L-Systems.

Forest Simulation: L-Systems are within the most popular grammars to satisfactory model botanical structures. The definitions introduced by [PL96] were later improved to more realistic three dimensional trees for [AK84]. Since then, both models have been used in a number of forest growth models.

The model proposed by [KS02] provides a wonderful framework to model the interdependence of L-Systems. Geometrical interpretation of parameters was used to calculate shadow cones and a carbon allocation economy was used to successfully control the growth of the trees. A later paper introduced a quasi-physical simulation of large-scale dynamic forest scenes by introducing a similar growth model and a wind field to account for environmental factors. Finally, [RLP07] introduced a new model by employing a three dimensional version of the space colonization algorithm.

2 Background

2.1 Definition

The work presented in this paper is based on L-systems so we will briefly introduce this modeling structure.

L-Systems: are parallel rewriting systems and a type of formal grammar. L-systems are now commonly referred as *parametric* L-systems, defined as a tuple:

$$\mathbf{G} = (V, \omega, P) \quad (2.1)$$

where

- \mathbf{V} (*alphabet*) is a set of symbols containing the elements in the string that can be replaced
- ω (*axiom*) is a string of symbols from V defining the initial state of the system.
- \mathbf{P} is a set of *production rules* which define the way variables can be replaced with combinations of constants and variables. Each rule is composed by a *predecessor* and a *successor*. The successor will consist of a list of symbols that will replace the predecessor. Usually the predecessor contains only one symbol, if it involves more, the system is called *context sensitive*. If a production rule is not specified for a symbol, the identity production is assumed.

Parametric L-systems: We can further expand the definition of L-system to accommodate parameters. In a parametric grammar, each symbol in the alphabet has a list of parameters associated with it. Parameters are usually real valued but there is not constraint in the mathematical structures that can be used. Production rules need to be extended to deal with parametric symbols. Parameters can be both *global* or *local* depending on if they belong to the current predecessor. Parameters are used in both conditional statements to choose between different production rules and for modifying the parameters of the symbols in the successor. In the

following example (2.2), we can see the behavior, where g is a global parameter.

$$A(x, y) : x < 2 \rightarrow A(g, x + 1)CB(3, y) \quad (2.2)$$

Stochastic L-systems: In order to deal with non-deterministic models, randomness can be introduced in the model by defining for each production rule $p \in P$ a set of production rules $\bar{p} = \{p_1, p_2 \dots p_k\}$ with associated probabilities $\bar{q} = \{q_1, q_2 \dots q_k\}$ and a random variable R that will evaluate the rule p as p_i with probability q_i . This kind of grammars are tremendously useful when generating great amounts of L-systems, because otherwise all the elements in the group would look the same.

2.2 Application

Once defined the specific type of grammar or combination of grammars, to actually generate the geometry described by a L-System we have to execute two stages: the *derivation* of the final symbol string and the *interpretation* of such string to a particular geometry or structure.

Derivation: this phase involves a number of successive interpretation iterations. In each iteration, all symbols in the current state are translated in parallel using the production rules. For each symbol, a production rule with a matching predecessor and condition is searched. Once found, the symbol will be substituted by the parametric symbols in the successor which are evaluated using the global parameters at the current iteration and the local parameters of the symbol that is being substituted. The initial state is the axiom of the system, the state string is updated after each iteration, so the final state is specified by the arbitrary number k of iterations that we have defined.

Interpretation: after the final state string of symbols is produced a set of interpretation rules will have to be applied to extract the information of the system. Most commonly, a geometric representation is generated by using two or three dimensional Turtle Geometry [AD86]. This translates the commands into modifications of the turtle state, which is represented by a position in space and a angle orientation in such space. Usually most modifications can be associated with euclidian affine transformation. However *branching* (or *bracketed*) systems will also use push and pop commands to input the current *turtle state* into a stack for later recovery.

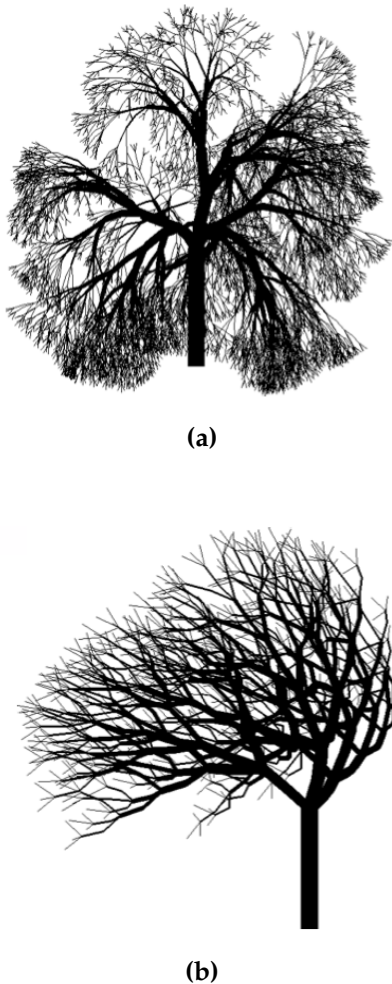


Figure 2.1: Examples of trees with ternary branching^[PL96]

3 Interdependence

Three dimensional parametric and stochastic L-Systems are a really powerful tool that help us model complex botanical structures such as monopodial, sympodial and ternary trees. By varying global parameters we can regulate some fixed parameters such as the way gravity affects the tree or the branching angle that the tree will follow. We can visualize these ideas in the ternary trees shown in Figure 2.1.

However, whereas we need stochastic systems as a source of variability to realistically emulate the structure of forests, there is a enormous caveat using them for multiple independent L-systems. To look into the problem lets consider the size of a state string S as $|S| = L$. If each production rule translate one symbol into two and we only have deterministic production rules, the output will have size $2L$. However, if each production rule only substitutes with a 50% chance then we only expect half of the symbols to change and the expected size of the output string will be $3L/2$. So we can find a proportionality between input and output sizes and this will deem an exponential growth. However, if by chance, one string gets a significantly greater or lower amount of successful production rules this will affect in the next iteration, since having more symbols in the string will allow the system to grow faster and viceversa.

This produces an oddly looking unrealistic system as we can see in Figure 3.1 where one of the plants has, after a few iterations grown bigger, and will keep doing so; whereas other plants have carried average iterations resulting in similar systems. This phenomenon can be justified by the fact that: a) the systems are completely independent of each other and therefore do not share any common parameters, b) we are not taking into account any real model for the shared medium, such as mineral resources.



Figure 3.1: Simulated growth of several simple stochastic L-Systems^[PL96]

In a similar fashion as [KS02] did, we can employ both the global parameters to the system and universal parameters to the whole group of systems to thwart this undesired behavior. These parameters display the relationship between different systems or different parts of the same system, and are usually updated after each derivation step with the information provided by different metrics on the current string state. Depending on the locality of the parameters we can differentiate two types of dependence:

- a) **Selfdependence** - In this case there exists a relationship between different parts of the same system. For example, given a tree whose production rules depend on the amount of light received by each branch would need to update this information after each timestep considering the shadows projected by its own branches.
- b) **Interdependence** - A more complex case that involves relationships between potentially different types of L-Systems, and that reflects a reciprocal interaction between them. We can extend the previous example, given a number of trees if we consider not only the shadow

of a tree to itself, but the shadow projected from other trees we will have introduced a interdependence relationship into the growth simulation.

Independent and selfdependent systems provide a embarrassingly parallel problem, since each system can be analyzed without considering the others they can be treated as individual problems. The load balancing can be achieved by a simple manager-worker scheme using a shared queue. However, interdependent groups of L-Systems can pose a nontrivial parallelization given appropriate constraints. For this paper we have used a local resource economy which translates in the following behavior. A tree will only be affected by the trees within a radius R , which from now onwards will be referred to as his *neighbors*. If a tree is bigger than the average size of their neighbors then it will be harder for it to grow at the next iteration, preventing the overgrowth behavior depicted in Figure 3.1. If a tree is smaller than the average size of its neighbors, then it will more likely for it to grow in the next iteration, compensating the randomness in previous iterations. We need to use this constraints since as we explained before, whereas randomness is a good source of variability for the population, it deeply alters the growth rate of different systems leading to undesired and unrealistic behaviors.

We still need to define what the *size* of a tree means from an analytical viewpoint. We will use the concept of *metrics*, functions when given as an input a list of parametric symbols, output a list of real valued parameters that reflect different aspects of the system. In our implementation the metric employed measured both the number of nodes (points where branching occurs) and leaves (terminal points of the system) to account for the dimensions of the tree.

A serial implementation of the described scenario was carried out¹. A fixed number of different L-Systems composed a given probability distribution of monopodial, sympodial and ternary trees was randomly distributed in a square surface. All these tree systems were slightly modified from the Honda canonical definitions given in [PL96] to accommodate for randomness in the growth. Next, a neighbor look-up list was initialized, since neighboring relations do not change throughout the simulation. For a fixed number k of iterations, each system was applied the derivation procedure and its metric recalculated and stored in a global array. Then, before starting the next iteration, the growth parameter of each system was updated using the metrics from its neighbors.

This implementation allows for huge variability of environments. Modifying the probability distribution of trees, the size of the Area, the number of systems and the number of iterations allow us to model from old, dense boreal forests to a widely spaced savanna.

4 Parallelization

When looking at a decently sized group of L-Systems ($N \approx 100$) and a considerable number of iterations ($k \approx 14$), the serial implementation starts taking seconds on most modern machines. Trying to increase any of the parameters goes to the domain of minutes and even hours. Therefore, the problem can use parallelization techniques in order to reduce the computation time by introducing more processing power. For this analysis we have chosen a shared memory architecture using the OpenMP application programming interface in C++. In this paper we are going to focus in the parallelization of the derivation step for the interdependant system described in Section 3.

The interpretation step of multiple L-Systems is an inherently embarrassingly parallel problem since all of the systems are independent when it comes to interpretation. To further optimize the parallelization of the interpretation step, one can look at parallelizing the interpretation of an individual L-System. This can be easily done for *non-branching* systems by replacing each symbol with its associated affine transformation (4x4 matrix). Since matrix multiplication is an associative operation we can easily solve the problem by using a *reduce* function. *Branching* systems can follow a similar approach but the complexity escapes the scope of the current paper. A more detailed explanation can be found in [LWW10].

4.1 Dynamic scheduling

A initial naive parallelization can be used by just simply parallelizing each one of the derivation iteration steps. Using a dynamic scheduling algorithm, the processors will evenly split the work and solve each iteration in parallel. We consider dynamic scheduling instead of static scheduling to get a better load balancing scheme. Even though we know the amount of subproblems from the start, we cannot evenly divide them into the processors since the complexity of each system will be variable due to the stochastic formulation that the trees are following.

As we shall see later in Section 5, this approach decreases the overall time but is far from perfect. The main disadvantage associated with this approach comes from the fact that a barrier must be set up at the end of each iteration to ensure correctness. This will imply that for p processors and N similarly sized trees, at the end of each iteration approximately $r \equiv N \bmod p$ processors will be still working and $p - r$ will be idle, waiting for the rest to finish and wasting processing power. On

¹All the code for this project can be found in <http://github.com/JJGO/Parallel-LSystem>

average, this means that half of the processors will be idle for approximately $\frac{1}{\lceil N/p \rceil}$ part of the time. As p grows larger this fraction also becomes larger which is undesirable.

Furthermore, the presence of the two implicit *barriers* per iteration in the algorithm incurs in a great overhead since all threads will be synchronized after each iteration. Therefore, if we want to optimize the algorithm we will need to reduce all of this communication to a minimum.

4.2 Lookahead strategy

Given the constraints of the problem we can enhance the parallelization by taking into account two facts:

- a) **Local dependence** - As we explained before, the simulation has only a local dependence from a tree to its neighbors. Therefore, we need not wait for every tree to finish iteration i to start iteration $i + 1$. A given tree t at iteration i only needs the metrics of its neighbors at iteration i to start the derivation phase. By applying this mechanism we can make use of every the processor available all the time; however, we will need to introduce series of constraints into the implementation.

An additional verification step must be performed to test if all the neighbors have in fact reached at least the current iteration and proceed only if the minimum of their iterations is bigger or equal the iteration of the system at hand.

The shared array that previously stored the metrics at each iteration will need another dimension for all the iterations, otherwise, we could potentially be reading a metric from a future iteration of a given neighbor, and correctness would be violated. Therefore, we will need to store for each tree n the metric at every iteration i .

Since no two processors will try to update the same metric or iteration variable during the entire program and reading not yet updated iteration variables does not affect the correctness; we will not need mutual exclusion to modify the metric variables. Accordingly, the communication is clearly lessened to only the neighbor checking and the work scheduling.

- b) **Geometry of the subproblems** - Since the simulation is running in parallel over all the N trees for k iterations we know in advance that $N \cdot k$ derivation steps will need to be done. We can dynamically schedule all this subproblems via an ordered work queue, so that the processors can take one subproblem at a time and if the derivation is possible (the neighbors have reached all the previous iteration) it will proceed, otherwise will return the work to the queue and try the next available subproblem. Since we are using an approach that relies on the connectivity of the trees in the forest, we can express the neighboring relationships in the forest as a undirected graph. An arbitrary queue will render a great amount of the neighbor checking steps unsuccessful, since we cannot predict which vertices have been derived yet and which not.

However, we can make use of the inherent geometry of the problem by employing a graph traversing algorithm such as *Breadth First Search*. By performing an initial *BFS* traversal into the graph and storing this precalculated order we can ensure that if at every iteration, the nodes are interpreted in the same order, then we will be maximizing vertices whose neighbors all have the same iteration value.

Consequently, the algorithm will start deriving at a given starting point of the graph and traversing it following *BFS*. When the other end is reached, a number of processors may not have work to do in said end, so they can go back to the starting point of the graph and

with a really high probability, they will be able to derive the next iteration of the vertices there, maximizing the processing power.

With all this optimizations the program works almost as good as before (for some inputs the neighbor checking slows the algorithm) when $p \mid N$ and much better in cases where $N \bmod p$ is really small which implies that for the previous implementation, a big fraction of the processors will be idle for an entire derivation step at the end of each iteration. Since the complexity of the derivation steps grows exponentially, this enhances the simulation runtime significantly.

4.3 Connected components optimization

A further optimization can be made. The main disadvantage of the solution proposed in section 4.2 comes from the fact that it introduces a whole neighbor checking phase that can consume a non-negligible amount of resources if the graph presents great connectivity. Also, the simulation uses as inputs both dense and sparse groups of trees in order to satisfactorily model different types of environments. This implies that in some cases we will have a single highly connected graph and in other cases we may have several smaller subgraphs or even individual vertices.

Moreover, as we briefly commented in the previous section the neighbor checking may be unnecessary in some cases. Specifically, when all processors are executing trees in the same iteration, neighbor checking needs not be done. By using a series of flag variables we can check if all the processors are in the same iteration and skip the neighbor checking.

In addition, to improve the algorithm we can also use the fact that we are performing a initial BFS to compute a order of visit to the vertices and also identify the different connected components of the graph. This is useful, since connected

components in the graph render completely independent subgroups of L-Systems that will not share any information whatsoever during all the iterations. Using this information we can split the main queue into smaller queues which helps reducing the overhead of scheduling the work.

4.4 Splitting L-Systems

To try to extend the parallelization further an attempt was made to parallelize the derivation of individual L-Systems. This rendered a problem when either N was really small or k was exceedingly big. To successfully divide a L-System into smaller work items, we can use the fact that L-Systems are parallel rewriting grammars. Since all the L-Systems covered in this implementation are context free, we can just evenly divide the number of symbols in the system in m different pieces and add them into the dynamic queue. It is important that they are not dispersed in the queue since as we analyzed before, the neighbor checking pattern works better when given a BFS order for the nodes. The L-Systems were slightly modified to cope with the multiple updates that now a derivation step mean. The iteration value will not change until the last derivation part is finished.

Following this scheme and building on top of the previous optimizations, we successfully reached a correct solution that split the work for L-Systems when idle processors are present and directly depend on that system to continue the work with another the derivation of another systems. However, the overhead involved in both checking if one L-System is bottlenecking the derivation and maintaining a queue with both entire L-System states and partial ones rendered the solution useless since it increases the serial time by at least one order of magnitude for almost all inputs. Consequently, this implementation was not used for the final results.

5 Results

To thoroughly test the parallelization, the parallelization techniques described in section 4, were implemented incrementally and tested to test them independently. As data inputs a random combination of monopodial, system and ternary trees were used in each of the systems, all sharing the same defining parameters, since forests are usually made of trees with similar characteristics. Randomness was also used to place the points in a plane. A square surface was used, whose size was randomly chosen at the beginning. The value for the size oscillated within the range $\sqrt{\sqrt{N}} \leq s \leq \sqrt{2N}$, since bigger sizes led to very sparse systems and smaller systems produced highly concentrated systems, and neither of them resembled a realistic forest model. Therefore, by randomizing the area we are getting a random density for each testcase.

As we also explained in previous section randomness is a crucial part in the derivation of stochastic L-Systems. At the interpretation of every nontrivial symbol, randomness is used to decide between using the default production rule or the plain identity. This coupled with the interdependence strategy models the forest behavior as desired.

this

6 Discussion

this

7 Conclusion

This report explores the viability of using parametric stochastic interdependent Lindenmayer sys-

tems to model the growth behavior of a forest, and a nontrivial way of parallelizing the problem. The final parallelization scheme involved a number of components which include a dynamically scheduled queue, an initial traversal search to determine the order of the nodes, a neighbor checking algorithm that deserialized the iteration steps allowing multiple iterations happen at once and finally a decomposition of the graph into smaller subgraphs to further optimize the results.

An important conclusion that we can gather from the analysis given is the fact that the nature of the L-Systems structure was not directly used in the parallelization (except for Section 4.4, which was unsuccessful). This means that the obtained parallelization algorithm can adequately model similar problems of the same characteristics. Namely, given a simulation problem with local dependence, and problems that grow with the same complexity after each iteration; the present algorithm should render similar results with similar inputs, even if the internal elements are not precisely L-Systems.

Future work: The main caveat of the produced outputs was that all trees had the same “age”, which translates in similar complexities (within one order of magnitude). This scenario is not unrealistic, (forest after fires develop in this fashion), but is far from general. We would like to analyze how the system would grow if new trees had the possibility to develop at any given iteration given certain circumstances like the presence of neighboring trees. This greatly breaks the current parallelization so it would be an interesting problem to analyze.

References

- [AD86] H. Abelson and A.A. DiSessa. *Turtle Geometry: The Computer as a Medium for*

- Exploring Mathematics*. Artificial Intelligence Series. AAAI Press, 1986.
- [AK84] M. Aono and T.L. Kunii. Botanical tree image generation. *Computer Graphics and Applications, IEEE*, 4(5):10–34, May 1984.
- [KS02] Winfried Kurth and Branislav Sloboda. Growth grammars simulating trees - an extension of l-systems incorporating local variables and sensitivity. *Silva Fennica*, 31(3):285 – 295, 2002.
- [LH04] Patrick Lacz and John C. Hart. Procedural geometry synthesis on the gpu, 2004.
- [LWW09] Markus Lipp, Peter Wonka, and Michael Wimmer. Parallel generation of l-systems. In Holger Theisel Marcus Magnor, Bodo Rosenhahn, editor, *Vision, Modeling, and Visualization Workshop (VMV) 2009*, pages 205–214, November 2009.
- [LWW10] Markus Lipp, Peter Wonka, and Michael Wimmer. Parallel generation of multiple l-systems. *Computers & Graphics*, 34(5):585–593, October 2010.
- [PJM94] Przemyslaw Prusinkiewicz, Mark James, and Radomír Měch. Synthetic topiary. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '94*, pages 351–358, New York, NY, USA, 1994. ACM.
- [PL96] Przemyslaw Prusinkiewicz and Aristid Lindenmayer. *The Algorithmic Beauty of Plants*. Springer-Verlag New York, Inc., New York, NY, USA, 1996.
- [PM01] Yoav I. H. Parish and Pascal Müller. Procedural modeling of cities. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '01*, pages 301–308, New York, NY, USA, 2001. ACM.
- [PMKL01] Przemyslaw Prusinkiewicz, Lars Mündermann, Radoslaw Karwowski, and Brendan Lane. The use of positional information in the modeling of plants. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '01*, pages 289–300, New York, NY, USA, 2001. ACM.
- [RLP07] Adam Runions, Brendan Lane, and Przemyslaw Prusinkiewicz. Modeling trees with a space colonization algorithm. In *Eurographics Workshop on Natural Phenomena*, 2007.
- [YHL⁺07] Tingjun Yang, Zhengge Huang, Xingsheng Lin, Jianjun Chen, and Jun Ni. A parallel algorithm for binary-tree-based string rewriting in l-systems. In *Computer and Computational Sciences, 2007. IMSCCS 2007. Second International Multi-Symposiums on*, pages 245–252, Aug 2007.
- [ZST⁺06] Long Zhang, Chengfang Song, Qifeng Tan, Wei Chen 0001, and Qunsheng Peng. Quasi-physical simulation of large-scale dynamic forest scenes. In Tomoyuki Nishita, Qunsheng Peng, and Hans-Peter Seidel, editors, *Computer Graphics International*, volume 4035 of *Lecture Notes in Computer Science*, pages 735–742. Springer, 2006.