

# The Walking Tree Method for 3D Vision

**Paul Cull**

Department of Computer Science  
Oregon State University  
Corvallis, OR 97331  
USA  
[pc@eeecs.oregonstate.edu](mailto:pc@eeecs.oregonstate.edu)

**Tai Hsu**

Department of Computer Engineering North-  
western Polytechnic University  
Fremont, CA 94539  
USA  
[hsut5@yahoo.com](mailto:hsut5@yahoo.com)

## Abstract

One of the major problems of visual perception is that the world we see is 3-dimensional, but we see it through a 2-dimensional retina. We all know part of the trick, binocular vision converts the 2-D image into a 3-D image. Here we describe a simple mechanism which makes this transformation computationally feasible.

## 1 Introduction

When we look at a forest, it resolves itself into a set of trees, some nearer and some further away. This seems so natural that it's only when we try to consider the mechanism that we realize how difficult a problem this is. Specifically, we start with light coming into our eyes, going through a lens into some light sensitive cells in the retina and then through some neural processing, and look and behold we see the trees. The specific problem is that the external (we assume 3D) world is mapped through a 2D retina and yet we "see" in 3 dimensions. We all know part of the trick, that is, binocular vision. Instead of one picture, we start with two 2D pictures and somehow "fuse" these two 2D representations into a 3D representation. How does a biological organism do this fusing? How could we instruct (or build) a robot to "see" in three dimensions? One way to think about the difference between two pictures of the same scene is to think in terms of objects. An object that is far away will be in roughly the same place in both pictures. But an object close to the viewer will be in very different places. Our idea is to align objects and use the distance between aligned pairs to estimate distance (depth) in the third dimension.

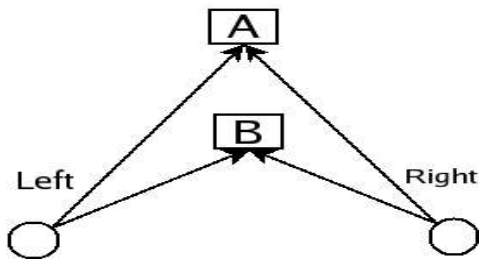
We convert 2D images into 3D images by adding an extra piece of information to every pixel of a 2D image: the distance from the observer to the pixel. 3D vision usually requires the comparison of two or more pictures. Sequences of pixels have to be aligned between pictures in order to use the distance "shift" of the same pixels to determine distances, e.g., the closer the object, the bigger its shift.

This image alignment problem is similar to the string alignment problem in genetics. For string alignment, dynamic programming algorithms can rapidly find an alignment, if one uses the edit-distance model [Gusfield, 1997; Setubal, et al, 1996] which assumes that changes between strings occur locally. But there is evidence showing that large scale non-local changes are possible [Devos, 1993], e.g., meiotic cell DNA replication is usually followed by homologous chromosome DNA recombination in which large pieces of DNA can be moved from one location to another (translocations), or replaced by their reversed complements (inversions). To handle such non-local changes, as well as the local changes, we created the Walking Tree method and demonstrated its use in phylogeny, gene discovery, and gene verification [Cavener et al, 2004, Cull et al, 1992, 1994, 1998, 1999a, 1999b, 2001, 2003a, 2003b, 2004]. This paper will show that the walking tree is also a reasonable tool for aligning 2D images to compute 3D images.

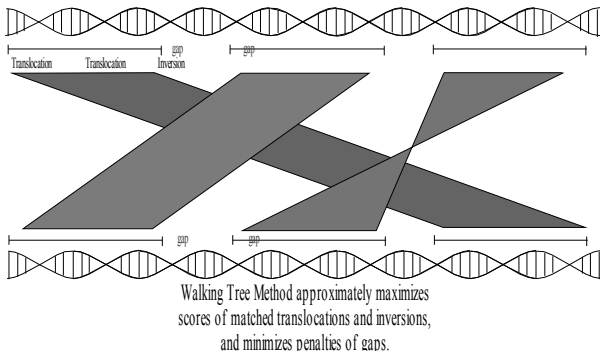
## 2 Walking Tree Methods

### 2.1 Why are Walking Tree Methods Needed?

Dynamic programming [Scharstein, 2002] can be utilized to implement stereo vision by comparing two pictures. If two pictures of a scene are taken by a pair of cameras, dynamic programming cannot align segments which are not in the same sequential order in both pictures (Figure 1). But, the Walking Tree Method can align segments even when the two segments' locations are swapped in the two pictures, and it can also handle insertions, deletions, and substitutions. Further, the Walking Tree heuristic tends to highlight features of objects, e.g. genes (Figure 2). Since inversions do not occur in pictures, this feature of the walking tree method can be turned off and the picture alignment can be computed more quickly than a genetic alignment.

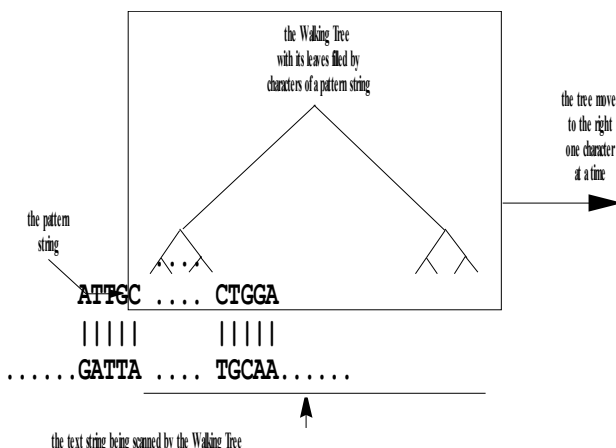


**Figure 1:** The segments ordering observed by the left camera is "AB", while "BA" is observed by the right camera. A standard dynamic programming method can't detect such a translocation, but the Walking Tree Method can.



**Figure 2:** This picture illustrates translocation, inversion, and gaps which can be located by the Walking Tree Method. DNA strings can be replaced by strings of pixels for image comparison.

## 2.2 The Basic Method



**Figure 3:** This picture shows the walking tree's structure, a binary tree. Leaves of the tree contain the characters of the pattern string  $P$ . After comparing each leaf with a corresponding character of the text string, the walking tree updates its nodes with new scores, then moves to the next position by moving each of its leaves one character to the right. Then it repeats the leaf comparison, and updates its node scores until it reaches the end of the text string.

The problem is to find an approximate alignment between two strings of pixels, one called pattern  $P$ , and the other called text  $T$ . Our metaphor is to consider the data structure as a walking tree [Python, 1989] with  $|P|$  leaves, one for each pixel in the pattern. When the walking tree is considering position  $l + 1$ , the internal nodes remember some of the information for the best alignment within the first  $l$  pixels for the text (see Figure 3). On the basis of this re-

membered information and the comparisons of the leaves with the text pixels under them, the leaves update their information and pass this information to their parents. The data will percolate up to the root where a new best score is calculated. The tree can then walk to the next position by moving each of its leaves one pixel to the right. The whole text has been processed when the leftmost leaf of the walking tree has processed the rightmost pixel of the text.

We use a function that gives a positive contribution based on the similarity between aligned pixels, and a negative contribution that is related to the number and length of gaps, translocations, and inversions. A gap in an alignment occurs when adjacent pixels in the pattern are aligned with non-adjacent pixels in the text. The length of the gap is the number of pixels between the non-adjacent characters in the text.

The computation at each leaf node makes use of two functions, MATCH and GAP. MATCH looks at the current text pixel and compares it with the pattern pixel represented by the leaf. In the simplest case we use

$$\begin{aligned} \text{MATCH}(P_i, T_j) &= c & \text{if } P_i = T_j \\ \text{MATCH}(P_i, T_j) &= 0 & \text{if } P_i \neq T_j \end{aligned}$$

For many of our examples we use  $c = 2$ . If we were matching similar pixels then MATCH could return a value depending on how similar the two compared pixels are.

If we only used the MATCH function, the leaves would simply remember if they had ever seen a matching pixel in the text. We use the GAP function to penalize the match for being far from the current position of the tree. So the leaf needs to remember both if it found a match and the position of the walking tree when it found a match. For example, a simple GAP function could be:

$$\text{GAP}(\text{currentpos}, \text{pos}) = \log |\text{currentpos} - \text{pos}|,$$

where  $\text{currentpos}$  is the current position of the walking tree, and  $\text{pos}$  is the position at which the walking tree found a match. Then the leaf could compute

$$\text{SCORE} = \max[ \text{MATCH}(P_i, T_j), \text{SCORE} - \text{GAP}(\text{currentpos}, \text{pos}) ]$$

and update  $\text{pos}$  to  $\text{currentpos}$  if MATCH is maximal. This means that a leaf will forget an actual match if it occurred far from the current position.

An internal node will only look at what it is remembering and at what its children have computed. Like a leaf node, the internal node computes

$$\text{SCORE} - \text{GAP}(\text{currentpos}, \text{pos})$$

which depends on what the node is remembering. From its two children, the node computes

$$\text{SCORE.left} + \text{SCORE.right} - \text{GAP}(\text{pos.left}, \text{pos.right})$$

This will penalize the sum of the children's scores because the position for the two scores may be different. But, the internal node also has to penalize this score because the left or right position may be far from the current position, so it also subtracts

$$\min[ \text{GAP}(\text{currentpos}, \text{pos.left}), \text{GAP}(\text{currentpos}, \text{pos.right}) ].$$

The internal node will keep the better of the score from its remembered data and the score computed from its children.

The walking tree will be computing a score for the current position of the tree. It is possible that it could forget a better score that was far from the present position. To avoid this problem, we add a special root node which simply keeps track of the best score seen so far.

In short, the walking tree finds an alignment  $f$  so that

$$f: [1, \dots, |P|] \rightarrow [1, \dots, |T|]$$

The alignment found approximately maximizes

$$\sum_{i=1}^{|P|} \text{MATCH}(P_i, T_{f(i)}) - \sum \text{GAP}(f(i), f(i+1)).$$

The actual functional being maximized is best described by the method itself and has no simple formula. Among other reasons, the above formula is only an approximation because the method tries to break strings into substrings whose lengths are powers of 2, even when using other lengths would increase the value of this formula.

## 2.2 Adjusting Gaps

The basic method places gaps close to their proper positions. If we use the method to align the string "ABCDEF" in the string "ABCXXDEF" the gap may be placed between 'B' and 'C', rather than between 'C' and 'D'. This is a result of the halving behavior of the basic method. By searching in the vicinity of the position that the basic method places a gap we can find any increase in score that can be obtained by sliding the gap to the left and right. The cost of finding better placements of the gaps is a factor of  $\log|P|$  increase in runtime since at each node we have to search a region of the text of length proportional to the size of the substring represented by the node.

## 3 Experiments

Two experiments are performed: one for the virtual world, and one for the real world. This allows us to explore the real world test using the knowledge based on the virtual world. Currently, the digital camera used in the experiment costs less than \$40 USD. No special treatment is used to preprocess the taken pictures.

### 3.1 Virtual World Test

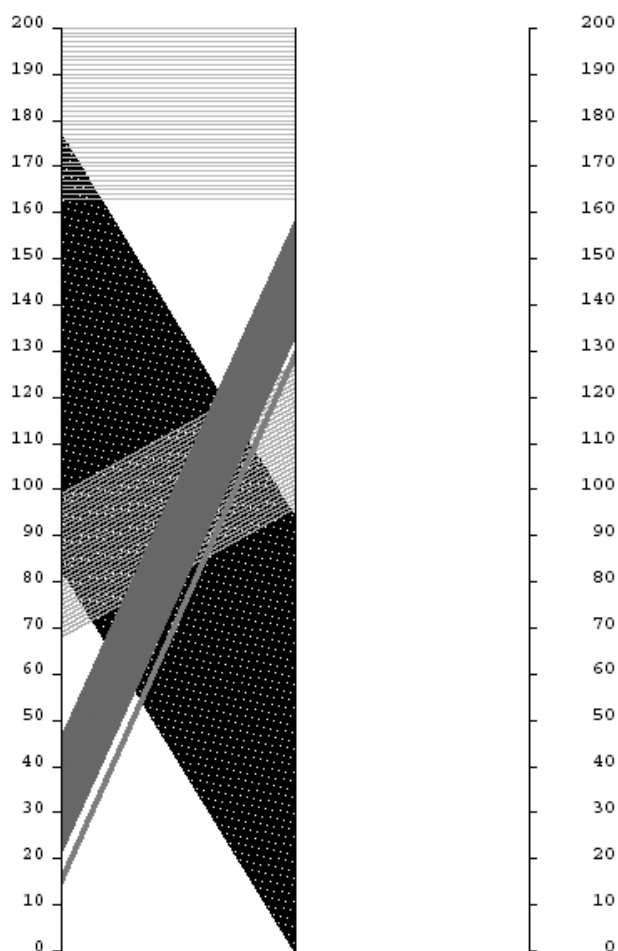


**Figure 4:** Analogous to Figure 1, the bigger block is A, and the smaller block is B. The left picture is what the left camera sees, and the right picture is what the right camera sees. Do

you know which object is closer? The big block is closer to the cameras, while the small block is further. Why? You may try to look at them using only one eye per picture. Namely, use a piece of paper to separate the pictures and eyes so that the left eye can see only the left picture, and the right eye can see only the right picture. When both eyes focus on the big block, your eyes are twisted inward, meaning the big block is closer. When the small block is focused, both eyes relaxed, meaning it's further.

We simulated two pictures (see Figure 4) in which two objects are placed in a way similar to Figure 1. Since both cameras' lines of sight are parallel, the objects taken by the left camera position a little to the right, while the objects taken by the right camera position a little to the left. Then, the Walking Tree Method processes the pixels of the two pictures, line by line, horizontally, to determine the depth of each pixel. The more "shift" a pixel does, the closer it is to the observer.

Why do we want to highlight this special simulated condition in which the two objects' locations are swapped in the two cameras? Because such a condition is the most difficult one for classic dynamic programming methods to solve.



**Figure 5:** This is the alignment computed by the Walking Tree Method using the vertical lines passing through both objects in pictures of Figure 4. The middle axis represents the pixel string of Figure 4's left picture, and the left axis represents the pixel string of Figure 4's right picture. Because it's impossible to have a positive "shift" from the left picture to

the right picture, this figure's two widest bands will be ignored, while only the bands of negative shift will be considered. And, the bigger shift of the red segment does tell us that it's closer to us, while the blue segment is further from us.

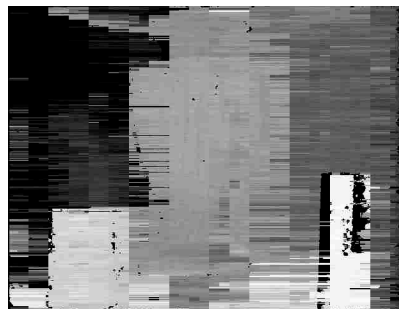
Figure 5 shows the alignment computed by the Walking Tree Method to demonstrate its ability to see swapped objects. We define a "positive" shift to be an object shifting its position, from the left location of a picture, to the right location of another picture. And, all other cases to be the "negative" shift. Because it's impossible to have a "positive" or "non-negative" shift from the left picture to the right picture, figure 5's two widest bands will be ignored, while only the bands of negative shift will be considered. So, the alignment tells us that the Walking Tree Method does see the two segments, even when their locations are swapped. And, the bigger shift of the big block does tell us that it's closer to us, while the small block's smaller shift tells us that it's further from us. The only problem left is that the area between both segments seems to be part of one of the two segments. This problem can be easily corrected by measuring the vertical shift using another pair of cameras.

### 3.2 Real World Test

We used an inexpensive digital camera to take two pictures as shown in Figure 6. In Figure 6, the left picture is taken when the camera is positioned at the left, and the right picture is taken when the camera is moved to the right. We ran the Walking Tree Method on two sequences of pixels, where one sequence is from the left picture, and one from the right picture. We constructed the 3D information of pixels based on the amount of their shifts (Figure 7). There are noises in the 3D picture, but we believe its quality can be improved by using more cameras taking pictures simultaneously (to avoid lighting changes) and using cameras that use X3 technologies [Merrill & Billings, 1999] (to avoid the mosaic effect of cameras that use single color photo sensors).



**Figure 6:** Two pictures are taken for the real world test. The left is taken by the left camera, and the right is taken by the right camera. For humans, it's easy to determine the relative distances of the three objects: the table lamp in the middle, and the black container on the left, and the shiny column on the right. We will let the Walking Tree Method to determine their relative distances by their "shifts" in the two pictures.



**Figure 7:** The distances of objects are represented by the brightness of a pixel. If the pixel is closer, it's brighter. This picture shows that the brightest part is at the lower right, i.e., the shiny column is the closest. The 2<sup>nd</sup> brightest object is at the lower left, i.e., the black container of the Figure 6. The 3<sup>rd</sup> brightest is the table lamp. And, the darkest is the wall. So, we know the relative distance from the closest to the furthest are the shiny column, the black container, the table lamp, and the wall.

## 4 Algorithmic Complexity

The Walking Tree Method is both time and space efficient. It can align two strings of length  $n$  in time proportional to  $n$  squared. The space used for computing this alignment is proportional to  $n \log n$ . Further, as shown in [Cull *et al.*, 2003a] a parallel version of the method will run in time proportional to  $n$ . In our experiments, a sequential version of the algorithm was run on a fairly slow PC and the alignments were still constructed in under a second. While our goal is to use this method for robotic vision, we believe that the Walking Tree is not unreasonable for a biological system. To verify this contention, we should show how to embody this algorithm as a neural network. Since the interconnection pattern is so simple, we believe that this will be an easy task.

## 5 Conclusion

From the result, we found that Walking Tree Method is also a powerful tool for stereo vision. The technique works by finding a "best" alignment between two strings of pixels. In addition to what can be done by a standard dynamic programming approach, the Walking Tree Method can find segments of swaps (translocations), which are required to accurately identify objects at various distances. Furthermore, since the Walking Tree Method computes the 3D information by comparing sequences of pixels, instead of objects, its capability is not limited to the amount of objects in a picture. What we did in the experiments is just a small step toward a much bigger goal of ours: to handle extremely complicated pictures for robots' 3D vision.

## 6. References

[Cavener *et al.*, 2004] J. Cavener, P. Cull, J. Holloway, and T. Hsu, "Walking Tree Heuristics for Comparative Genomic Alignments", *Mathematical Biosciences*, 188, pp207-219, 2004.

- [Cull et al, 1992] P. Cull and J. Holloway, "Divide and Conquer Approximate String Matching: When Dynamic Programming is not Powerful Enough", Technical Report 92-20-06, Computer Science Department, Oregon State University, 1992.
- [Cull et al, 1994] P. Cull and J. Holloway, "Aligning genomes with inversion and swaps", Second International Conference on Intelligent Systems for Molecular Biology. Proceedings of ISBM '94, 195-202. AAAI Press, Menlo Park CA 1994.
- [Cull *et al.*, 1998] P. Cull, J. Holloway, J. Cavener, "Walking Tree Heuristics for Biological String Matching and Gene Location", Cybernetics and Systems '98, (R. Trapp, editor), Austrian Society for Cybernetic Studies, pp374-379, 1998.
- [Cull *et al.*, 1999a] P. Cull, J. Holloway, J. Cavener, "Walking Tree Heuristics for Biological String Alignment, Gene Location, and Phylogenies", CASYS '98, Computing Anticipatory Systems (D. M. Dubois, editor), American Institute of Physics, Woodbury, New York, pp201-215, 1999.
- [Cull *et al.*, 1999b] Paul Cull, Tai Hsu, "Improved Parallel and Sequential Walking Tree Algorithms for Biological String Alignments", Supercomputing Conference, 1999.
- [Cull *et al.*, 2001] Paul Cull, Tai Hsu, "Gene Verification and Discovery by Walking Tree Method", Pacific Symposium on Biocomputing, pp287-298, 2001.
- [Cull *et al.*, 2003a] Paul Cull, Tai Hsu, "Fast Walking Tree Method via Recurrent Reduction for Biological String Alignment", the 7th World Multi-Conference on Systemics, Cybernetics and Informatics, p.228-233, vol. XIV, 2003.
- [Cull *et al.*, 2003b] Paul Cull, Tai Hsu, "Recent Advances in the Walking Tree Method for Biological Sequence Alignment", EUROCAST 2003, Springer, Lecture Notes in Computer Science 2809, pp349-359, 2003.
- [Cull *et al.*, 2004] Paul Cull, Tai Hsu, "Parallel Walking Method for Sequence Recombination", Parallel and Distributed Computing Systems, pp7-12, 2004.
- [Devos *et al.*, 1993] K. M. Devos, M. D. Atkinson, C. N. Chinoy, H. A. Francis, R. L. Harcourt, R. M. D. Koebner, C. J. Liu, P. Masojc, D. X. Xie, M. D. Gale, "Chromosomal rearrangements in the rye genome relative to that of wheat", Theoretical and Applied Genetics 85:673-680, 1993
- [Gusfield, 1997] D. Gusfield, "Algorithms on strings, trees, and sequences: computer science and computational biology." Cambridge University Press, New York, NY, USA, 1997.
- [Merrill & Billings, 1999] "Color separation in an active pixel cell imaging array using a triple-well structure" United States Patent No. 5,965,875; Oct 12, 1999.
- [Scharstein, 2002] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", International Journal of Computer Vision, 47(1/2/3):7-42, April-June 2002
- [Setubal *et al.*, 1996] J. Setubal and J. Meidanis, "Introduction to Computational Molecular Biology." PWS Publishing, Boston, MA, 1996
- [Python, 1989] M. Python, "Just the Words.", vol. 2, p338. Metheun, London, 1989.

