Surface visibility probabilities in 3D cluttered scenes

Michael S. Langer

School of Computer Science, McGill University
Montreal, H3A2A7, Canada
email:langer@cs.mcgill.ca
http://www.cim.mcgill.ca/~langer

Abstract. Many methods for 3D reconstruction in computer vision rely on probability models, for example, Bayesian reasoning. Here we introduce a probability model of surface visibilities in densely cluttered 3D scenes. The scenes consist of a large number of small surfaces distributed randomly in a 3D view volume. An example is the leaves or branches on a tree. We derive probabilities for surface visibility, instantaneous image velocity under egomotion, and binocular half–occlusions in these scenes. The probabilities depend on parameters such as scene depth, object size, 3D density, observer speed, and binocular baseline. We verify the correctness of our models using computer graphics simulations, and briefly discuss applications of the model to stereo and motion.

1 Introduction

The term *cluttered scene* typically refers to a scene that contains many visible objects [2, 18, 6] distributed randomly over space. In this paper, we consider a particular type of cluttered scene, consisting of a large number of small surfaces distributed over a 3D volume. An example is the leaves or branches of a tree.

Reconstructing 3D geometry of a cluttered scene is a very challenging task because so many depth discontinuities and layers are present. Although some computer vision methods allow for many depth discontinuities, such methods typically assume only a small number of layers are present, typically two [28].

The goal of this paper is to better understand the *probabilistic* constraints of surface visibilities in such scenes. We study how visibility probabilities of surfaces depend on various geometric parameters, namely the area, depth and density of surfaces. We also examine how visibility probabilities depend on observer viewpoint. Such probability models are fundamental in many methods for computing optical flow [24, 29, 19] and stereo[1, 3]. The hope is that the probability models could be used to improve these methods, for example, in a Bayesian framework by providing a *prior* on surface visibilities in 3D cluttered scenes.

2 Related Work

The probability models that we present are related to several visibility models that have appeared in the literature. We begin with a brief review.

The first models are those that describe atmospheric effects such as fog, clouds, and haze [15] and underwater [22]. In these models, the size of each occluder is so small and the number of occluders is so large that each occluder projects to an image area much smaller than a pixel. This is the domain of partial occlusion *i.e.* transparency [26]. This scenario may be thought of as a limiting case of what is considered in this paper. Another interesting case for which partial occlusion effects dominate the analysis is that of rain [5].

A second related model is the *dead leaves* model [11, 23] which consists of a sequence of 2D objects such as squares or disks that are dropped into a 2D plane, creating occlusion relationships. Here one is typically concerned with the distribution of object boundaries and sizes [21, 20, 10, 17]. The key difference between the dead leaves model and the one we present is that the dead leaves model assumes 2D shapes that lie in the same plane *i.e.* dead leaves on the ground, and thus assumes orthographic projection rather than perspective projection. In the dead leaves model, there is no notion that shapes lying at greater depths appear smaller in the image, as is the case considered in this paper of "living leaves" in 3D space and viewed under perspective projection. While the dead leaves model is sufficient for understanding certain aspects of visibility in 3D cluttered scenes, for example, probability of surface visibility along a single line of sight as a function of scene depth (see Sec. 3), it cannot describe other aspects such as how visibility changes when the observer moves (Sec. 4 and 5).

A third set of related models consider both occlusion and perspective, and combine geometric and statistical arguments (see two monographs [7, 30]). Such models have been used in computer graphics in the context of hidden surface removal [14] as well as for rendering foliage under ambient occlusion lighting [8]. Similar geometric and statistical reasoning has been used in computer vision for planning the placement of cameras to maximize scene visibility in the presence of random object occlusions[12].

3   Static monocular observer

The model developed here is based on several assumptions. First, the scene consists of surfaces that are randomly distributed over a 3D view volume. Each surface is represented by a particular point (e.g. its center of mass). The distribution of these points assumed to be Poisson [4] with density η, which is the average number of surface centers per unit volume. We assume that the density η is constant over a 3D volume. We also ignore interactions between surface locations such as clustering effects that may be found in real cluttered scenes and the fact that real surfaces such as leaves cannot intersect.

According to the Poisson model, the probability that any volume V contains k surface centers is

$$p(k\ surface\ centers\ in\ V) = e^{-\eta V}\frac{(\eta V)^k}{k!}$$

and so the probability that there are no surface centers in the volume V is

$$p(no\ surface\ centers\ in\ V) = e^{-\eta V}. \tag{1}$$

This last expression will be used heavily in this paper. Let's consider two instantiations of this model.

3.1 Planar patches

Suppose the surfaces are planar patches of area A such as in the first row of Figure 1. What can we say about visibility in this case? The expected total area of the surfaces per unit volume is ηA. If the surface normals of all patches were parallel to the line of sight, then the expected number of surfaces intersected by a unit line that is parallel to these surface normals would be ηA. (To see this, imagine cutting a surface of area ηA into pieces of unit area and stacking them perpendicular to the line.) The more general and interesting case is that the surface normals are oriented in a uniformly random direction. In this case, it is easy to show that the average number of surface intersections per unit length line is reduced by a factor $\frac{1}{2}$, so the expected number of surfaces intersected by a line of unit length is:

$$\lambda = \frac{\eta A}{2}.$$

Using standard arguments about Poisson distributions[4], one can show that the probability $p(z)dz$ that the first visible surface is in depth interval $[z, z + dz]$ is:

$$p(z)dz = e^{-\lambda z}\ \lambda dz.$$

The probability density $p(z) = \lambda e^{-\lambda z}$ is defined over $z \in [0, \infty]$. Often, though, we are interested in scenes whose objects are distributed over some finite volume beyond some distance z_0 from the viewer. In this case, the probability density is

$$p(z) = \lambda e^{-\lambda(z - z_0)}\ \ . \tag{2}$$

In computer graphics terminology, z_0 is the distance to the near clipping plane.

3.2 Spheres

The mathematical model presented in Sec. 5 for binocular visibility will be easier to derive in the case that the objects are spheres of radius R, so we consider this case next. Consider an imaginary line segment with one endpoint at the viewer and the other endpoint at distance z. In order for the 3D scene point at this distance z to be visible from the viewer, no sphere center can fall in a cylinder[1] whose radius is the same R as above, and whose axis is the line segment of length z. Such a cylinder has volume

$$V = \pi R^2 z.$$

---

[1] To be more precise, we would need to consider a capped cylinder [30], namely a cylinder capped at both ends by two half spheres. However, the extra volume of the capped cylinder has only a small effect on the model, and so we ignore this detail.

As with the planar patch scene, if all surfaces lie beyond a distance z_0, then the cylinder that cannot contain any sphere centers has volume:

$$V = \pi R^2 (z - z_0). \tag{3}$$

In order for the first surface along the line of sight to be in the depth interval $[z, z + dz]$, two conditions must hold: first, the cylinder must contain no sphere centers, and second, a sphere center must fall in a small volume slice $\pi R^2 dz$ capping the cylinder. From the above expression for V and from Eq. (1), the probability that the first surface visible along the line of sight is in depth interval $[z, z + dz]$ is

$$p(z)dz = \eta \pi R^2 e^{-\eta \pi R^2 (z - z_0)} \ dz \tag{4}$$

This model is the same as Eq. (2) where now for the case of spheres we have

$$\lambda = \eta \pi R^2.$$

We next present computer graphics simulations to illustrate this model.

3.3   Experiments

Scenes were rendered with the computer graphics package RADIANCE[2] [9]. Each rendered scene consisted of a set of identical surfaces randomly distributed over the rectanguloid

$$(x, y, z) \in [-4, 4] \times [-4, 4] \times [2, 8]$$

i.e. near and far clipping planes at $z = 2, 8$, respectively. The width of the field of view was 30 degrees. Each rendered image was 512×512 pixels.

Results are presented for scenes consisting of either squares (planar surfaces) or spheres. For both types of scene, three scene densities were used, namely $\eta = 14, 56, 225$ surfaces per unit volume. The object sizes for each of these densities were chosen to be large, medium and small, such that the λ value was the same over all scenes and so the theoretical $p(z)$ curves are the same as well. Specifically, for the sphere scenes, the radii were arbitrarily chosen to be $R = 0.1, \ .05, \ .025$. The corresponding areas for the square scenes were $2\pi R^2$ which kept the λ values the same for all scenes.

Figure 1 compares the model of $p(z)$ in Eqs. 2 and 4 with the average histogram of *visible depths* from the RADIANCE Z buffer over ten scenes. For each histogram, 15 bins were used to cover the depth range $z \in [2, 8]$. Standard errors for each histogram bin are also shown.

For the scenes with larger values of R, the standard errors within each histogram are greater than for the scenes with small R values. To understand this effect, consider for example the histogram bin for the nearest depth. The number of pixels in that histogram bin depends on the number of sphere centers that

---

[2] Unlike OpenGL, RADIANCE computes cast shadows which we will use later in the paper to examine binocular half-occlusions.
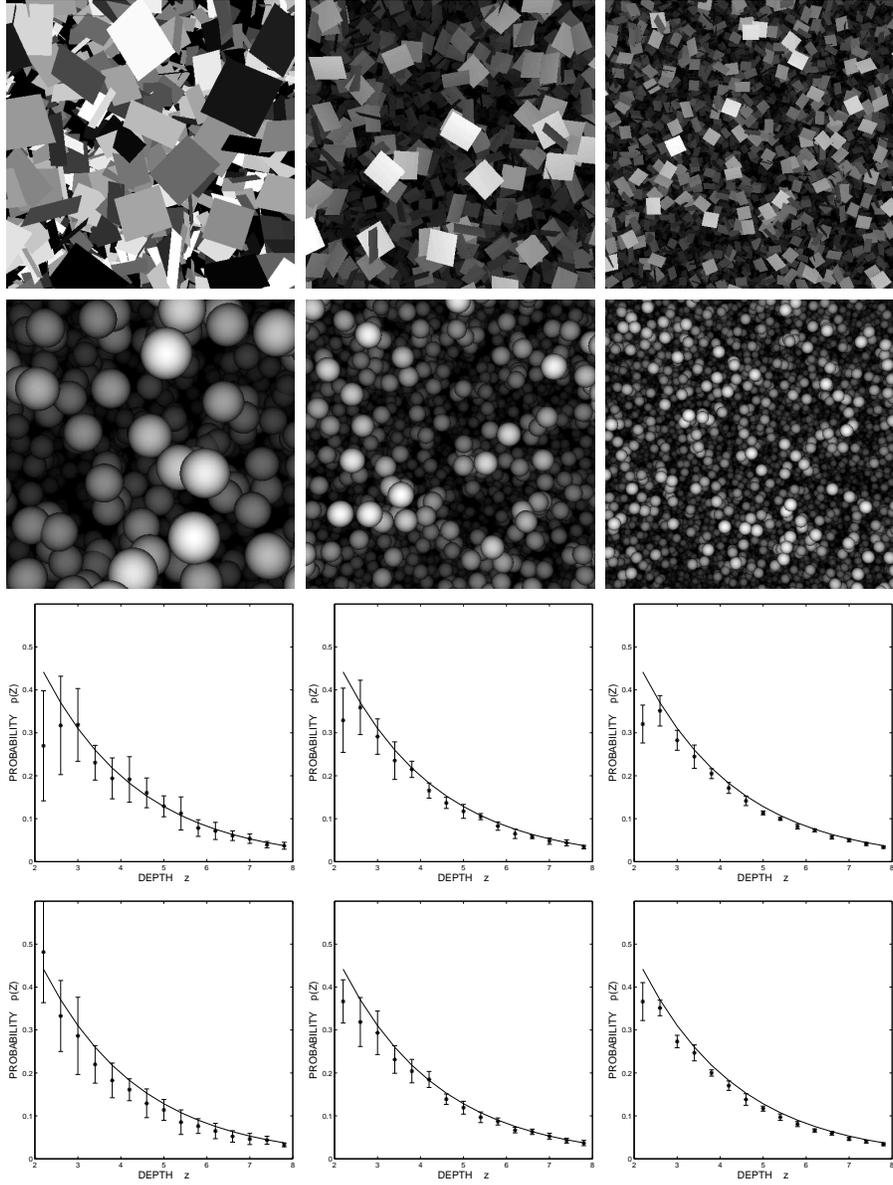
Fig. 1. Examples of rendered scenes consisting of squares or spheres distributed in a view volume. The density and areas of the three scenes have the same λ and so the expected visibility probabilities are the same. Histograms of depths from ten scenes and standard errors in each histogram bin are shown and compared with the theoretical model. The fits are excellent in each case. (See text.)

are in the closest depth layer. When the surfaces are large and the density is small, relatively few spheres will contribute to that bin, but each sphere that does contribute will contribute a large number of pixels, hence larger variance.

A second observation is that the theoretical curves slightly overestimate the probabilities, especially at nearest depth bin. The reason for the bias is that the RADIANCE Z buffer gives z values in Euclidean distance from the observer – as do Eqs. (2) and (4). However, as the surfaces are distributed in a rectanguloid whose boundaries are aligned with the viewer coordinate system, there are slightly fewer surfaces at the smallest z values, since for example distance $z = z_0$ falls in the view volume only at one point – where the optical axis meets the near plane. Because the field of view is small, the bias is also small and so we ignore it in the subsequent discussion.

4   Moving monocular observer

What happens when the observer moves, in particular, when the observer moves laterally? What can be said about the probability of different image speeds occuring? The key idea is that image speed depends on scene depth and, since we have a probability density on depth, we can convert it to a probability density on image speed.

Assume the projection plane is at depth $z = 1$ unit and the observer's horizontal translation speed is T units per second. The horizontal image speed v_x for a visible surface at depth z is

$$v_x = -\frac{T}{z} \tag{5}$$

and the units are radians per second. Following van Hateren [27], the probability $p(z)\,dz$ that the visible surface lies in the depth interval $[z, z+dz]$ can be related to the probability $p(v_x)\,dv_x$ that a visible surface has image speed in $[v_x, v_x+dv_x]$ as follows. Eq. (5) implies

$$dz = -\frac{T}{v_x^2}dv_x \; . \tag{6}$$

To interpret the minus sign, note that an increase in depth ($dz > 0$) implies a decrease in speed ($dv_x < 0$). Combining Eq. (4) and (6) with

$$p(z)dz = p(v_x)dv_x$$

and defining $v_0 = \frac{T}{z_0}$ to be the image speed for a point on the near clipping plane, we get

$$p(v_x) = \frac{\eta \pi R^2 T}{v_x^2} e^{-\eta(\pi R^2 T(\frac{1}{v_x} - \frac{1}{v_0}))}. \tag{7}$$

Figure 2 shows histograms of the image speeds, plotted from slow to fast (corresponding to depths far to near). Note that the probability density $p(v_x)$ is non-monotonic. While at first blush this may seem to contradict the depth

histograms from Fig. 1, in fact it does not. We have binned the velocities in uniform Δv bins, but these do not correspond to the uniform Δz bins.

Again note that the standard errors in the velocity histograms are greater for the plot on the left, which corresponds to large objects (R large). The reason is similar to what was argued for Fig. 1.

One final note is that if we define $v_x' \equiv \frac{v_x}{T}$ then

$$p(v_x)dv_x = \eta\pi R^2(\frac{T}{v_x})^2 e^{-\eta(\pi R^2(\frac{T}{v_x} - \frac{T}{v_0}))}d\frac{v_x}{T} = p(v_x')dv_x'. \tag{8}$$

Thus the observer speed T merely rescales the abscissa of the histogram. This is intuitively what we might expect, for example, doubling the observer speed doubles the image speed at each pixel.

$$R = 0.1 \qquad\qquad R = 0.05 \qquad\qquad R = 0.025$$
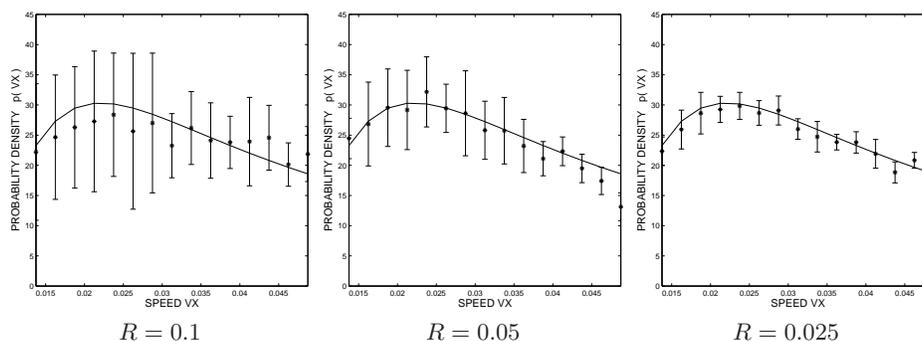
Fig. 2. Image speed histograms for the sphere scenes from Figure 1. Standard errors decrease as R decreases i.e. from left plot to right. Data are similar for the square scenes (not shown). The observer is moving at $T = 0.1$ space units per second and so the velocities are in units of radians per second.

5   Binocular Half Occlusions

The velocities discussed in the previous section were instantaneous. When an observer moves over a finite distance in 3D cluttered scenes, however, surfaces can appear from and disappear behind each other at occlusion boundaries [13]. This problem also arises in binocular stereo, of course, and arguably is more severe in stereo since the stereo baseline T is typically larger than the distance moved by a moving observer between frames of a video sequence. In stereo this appearance and disappearance phenomenon is know as *binocular half-occlusion* [1, 3]. We now introduce a probability model for half occlusions in 3D cluttered scenes.

Consider the two cylinders C_l and C_r whose axes are the line segments joining some point at depth z to the two eyes (see Appendix). Following similar reasoning to case of spheres in Sec. 3.2, the point at depth z is visible to *both* eyes if and only if *neither* cylinder contains a sphere center. Moreover, the conditional probability that a point at depth z is visible to the right eye, given it is visible to the left eye is:

$$p(z \ visible \ to \ right \mid visible \ to \ left) = \frac{p(z \ visible \ to \ both \ eyes)}{p(z \ visible \ to \ left)}.$$

Recalling Eq. (1) and noting that $vol(C_r \cup C_l) = vol(C_l) + vol(C_r \backslash C_l)$, we get:

$$p(z \ visible \ to \ right \mid visible \ to \ left) = \frac{e^{-\eta \ vol(C_l \cup C_r)}}{e^{-\eta \ vol(C_l)}} = e^{-\eta \ vol(C_r \backslash C_l)} \qquad (9)$$

We thus need an expression for $vol(C_r \backslash C_l)$. An exact expression is quite complicated since it involves the intersections of two cylinders at a general angle. But using arguments of a similar flavor as those used in [30, 14] we can derive the approximation (see Appendix):

$$vol(C_r \backslash C_l) \approx \begin{cases} \pi(\frac{2Rz}{T} - z_0)R^2 + \frac{2z}{T}R^3, & \text{if } \frac{z-z_0}{z} > \frac{2R}{T} \\[2ex] \frac{RT(z-z_0)^2}{z} & , & \text{otherwise.} \end{cases}$$

This model is examined using computer graphics simulations and the same scenes as before. To compute which points are visible to both eyes, the following trick is used. Treating the viewing position as the left eye, two versions of each scene are rendered: one in which the scene is illuminated by a point source from the viewing position (left eye) and a second in which the scene is illuminated by a point light source at the right eye's position. Then pixels that have non-zero intensity in both rendered images are visible to both eyes. Conditional probabilities are estimated from the relative number of pixels at which the surface is visible to both eyes versus the number visible just to the left eye.

Figure 3 compares the model of conditional probabilities with data obtained from RADIANCE simulations. The distance between the eyes was $T = 0.1$ which corresponds to the radius of the spheres in the large sphere scenes. The conditional probabilities are very well fit by the model. While this is not so surprising for the sphere case – it merely verifies that our approximation for $vol(C_r \backslash C_l)$ is good – it is somewhat surprising that the model should produce a good fit for the squares case as well, as the model was derived under the assumption that the scene consists of spheres.

Another interesting observation is that, for both λ and T fixed, the conditional probabilities depend strongly on R, that is, on the size of the surfaces. For the scenes with high-density and small-R objects, the conditional probability falls off faster with z. That is, all other things (z, λ, T) being equal, binocular half occlusions are more common when the scene consists of smaller objects.
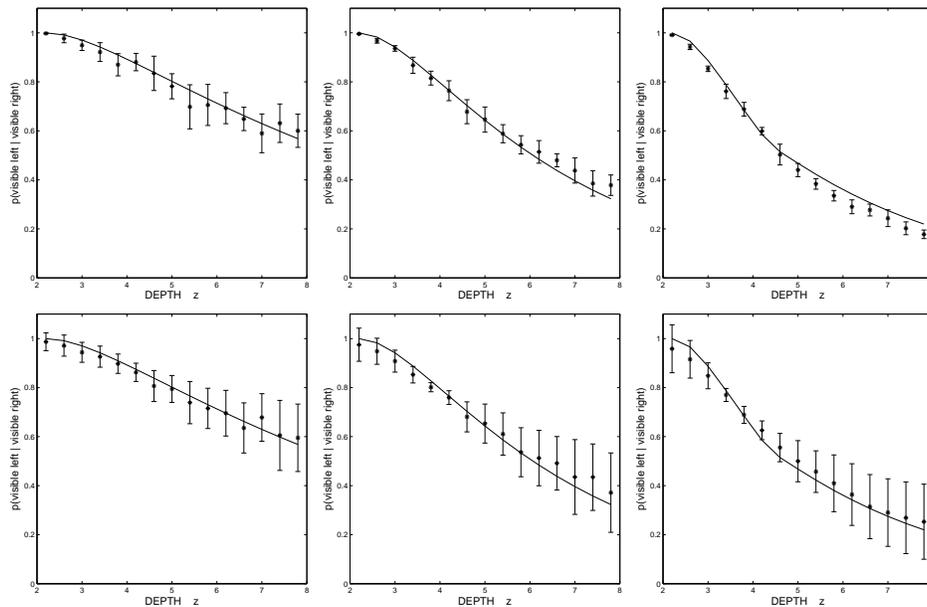
Fig. 3. Conditional probabilities that a surface at depth z is visible to right eye, given it is visible to left eye, for large (left), medium (middle), small (right) scenes.

6 Limitations

Here we discuss a few limitations of the model, which will be investigated in future work when the model is compared to real imagery. One limitation is the assumption of uniform density and size of objects. For example, the distribution of leaves and branches on typical trees and shrubs is not uniform, *e.g.* leaves tend to clump together near branches and the branch distribution itself is not uniform. This issue has been studied, for example, by plant scientists who are interested how light penetrates a tree canopy [25] and how the penetration depends on geometric and statistical properties of the plant such as leaf area, orientation, and branching factors [16].

A second limitation is that the model ignores partial occlusions which can arise for example from the finite size of pixels. To account for partial occlusion, we would need to modify the model slightly by considering an R-dilated cone rather than an R-dilated ray, such that the solid angle of the cone is that of a pixel. We claim that the difference between the two is insignificant for the case we are considering in which the objects subtend a solid angle much greater than that of a single pixel. However, in other cases in which the objects subtend a much smaller visual angle (or in which the depth of field is small), we would need to consider partial occlusions. Both of the above limitations are interesting topics to explore in future research.

Acknowledgements

Appendix

Figures 4 and 5 illustrate two configurations that must be considered for us to calculate $vol(C_l \backslash C_r)$. In both cases, I only need to consider the parts of the cylinder that are beyond z_0 since by definition no sphere centers are present for $z < z_0$. The two cases are distinguished by whether the cross sections of the cylinders (disks) overlap at depth z_0.

If the two cylinders overlap at depth z_0, then we approximate $C_l \backslash C_r$ to be the new volume swept out by a cylinder of radius R if a viewer were to move continuously from the left to right eye's position.[3] By inspection this swept volume is $\frac{RT(z-z_0)2}{z}$, where the factor $\frac{(z-z_0)T}{z}$ is the projection of the inter-ocular distance T to the z_0 plane.
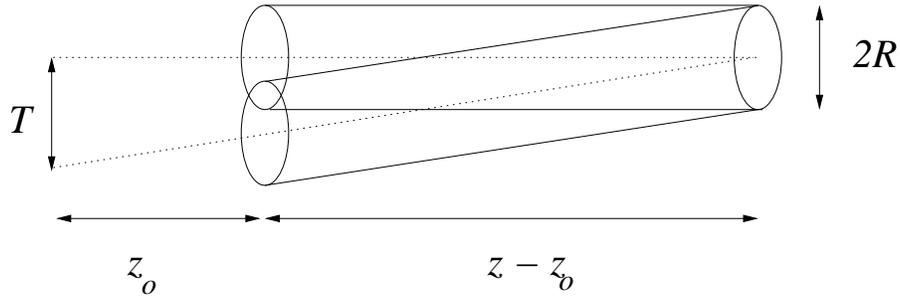
Fig. 4. A point at depth z is visible to two eyes separated by a distance T if no sphere centers lie in the cylinders whose axes join the point to the two eyes. Here we consider the case that the two cylinders overlap at the clipping plane $z = z_0$.

If the two cylinders do not overlap at depth z_0, then there must be a minimal depth z_1 where the two cylinders overlap such that $z_0 < z_1 < z$. This is the situation shown below. In this case we approximate $C_l \cup C_r$ by partitioning $C_l \backslash C_r$ into two sets, namely the points in front of and beyond this depth z_1, respectively. For the points in front of z_1, the volume within $C_l \backslash C_r$ is $\pi R^2(z_1 - z_0)$. Using a similar triangle argument, we note $\frac{z-z_1}{2R} = \frac{Z}{T}$, and with a little manipulation

---

[3] This swept volume is a slightly overestimate of $C_l \cup C_r$. It assumes the point at depth z, which lies on the axis of the cylinders, is seen by *all* viewpoints in between the left and right eye's position.

we can rewrite this volume as $\frac{2R^3 Z}{T}$. For the points beyond z_1, we approximate the volume within $C_l \backslash C_r$ using a similar idea as the case of Fig. 4. Now the two cylinders are exactly tangent to each other at depth z_1 so the distance between their axes at this depth is $2R$. we approximate the volume of $C_l \backslash C_r$ that lies beyond z_1 as the volume swept out by the cylinder, namely $\frac{2R}{2}(z - z_0)$. This is a slight overestimate of the volume i.e. an approximation.
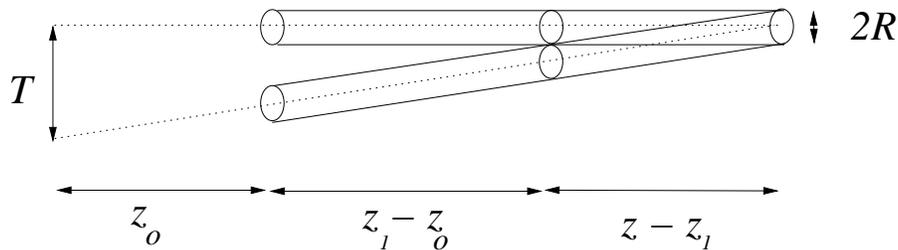
Fig. 5. Similar to Fig. 4 but now we consider the case that the two cylinders do not overlap at the clipping plane $z = z_0$, and that the overlap occurs starting at depth $z = z_1 > z_0$.

References

1. P. Belhumeur. A Bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19(3):237–260, 1996.
2. M. J. Bravo and H. Farid. A scale invariant measure of clutter. *Journal of Vision*, 8(1):1–9, 2008.
3. G. Egnal and R.P. Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1127–1133, August 2002.
4. W. Feller. *Introduction to Probability Theory and Its Applications, Vol. 1*. Wiley Series in Probability and Mathematical Statistics, 1968.
5. K. Garg and S.K. Nayar. Vision and rain. *International Journal of Computer Vision*, 75(1):3–27, October 2007.
6. Ulf Grenander and Anuj Srivastava. Probability models for clutter in natural images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(4):424–429, 2001.
7. P. Hall. *Introduction to the Theory of Coverage Processes*. John Wiley & Sons, Inc., 1988.
8. K. Hegeman, S.Premože, M. Ashikhmin, and G. Drettakis. Approximate ambient occlusion for trees. In *I3D '06: Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pages 87–92, New York, NY, 2006. ACM.
9. G. Ward Larson and R. Shakespeare. *Rendering with Radiance: The Art and Science of Lighting Visualization*. Morgan Kaufmann, 1998.

10. A. B. Lee, D. Mumford, and J. Huang. Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1/2):35–59, 2001.

11. G. Matheron. *Random Sets and Integral Geometry.* John Wiley and Sons, 1975.

12. A. Mittal and L.S. Davis. A general method for sensor planning in multi-sensor systems: Extension to random occlusion. *International Journal of Computer Vision*, 76(1):31–52, January 2008.

13. K.M. Mutch and W.B. Thompson. Analysis of accretion and deletion at boundaries in dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7(2):133–138, March 1985.

14. B. Nadler, G. Fibich, S. Lev-Yehudi, and D. Cohen-Or. A qualitative and quantitative visibility analysis in urban scenes. *Computers and Graphics*, 23(5):655–666, 1999.

15. S.G. Narasimhan and S.K. Nayar. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3):233–254, July 2002.

16. Prusinkiewicz P. Modeling of spatial structure and development of plants: a review. *Scientia Horticulturae*, 74:113–149, 1998.

17. N. M. Grzywacz R. M. Balboa, C. W. Tyler. Occlusions contribute to scaling in natural images. *Vision Research*, 41(7):955–964, 2001.

18. R. Rosenholtz, Y. Li, and L. Nakano. Measuring visual clutter. *Journal of Vision*, 7(2):1–22, 2007.

19. S. Roth and M. J. Black. On the spatial statistics of optical flow. *International Journal of Computer Vision*, 74(1):33–50, 2007.

20. D. L. Ruderman. Origins of scaling in natural images. *Vision Research*, 37(23):3385–3398, November 1997.

21. D. L. Ruderman and W. Bialek. Statistics of natural images: scaling in the woods. *Physical Review Letters*, 73:814–817, 1994.

22. Y.Y. Schechner and N. Karpel. Clear underwater vision. *IEEE Conf. on Computer Vision and Pattern Recognition*, 1:I–536–I–543 Vol.1, 2004.

23. J. P. Serra. *Image Analysis and Mathematical Morphology.* Academic Press, London, 1982.

24. E P Simoncelli, E H Adelson, and D J Heeger. Probability distributions of optical flow. In *Proc Conf on Computer Vision and Pattern Recognition*, pages 310–315, Mauii, Hawaii, 1991. IEEE Computer Society.

25. H. Sinoquet, G. Sonohat, J. Phattaralerphong, and C. Godin. Foliage randomness and light interception in 3d digitized trees: an analysis of 3d discretization of the canopy. *Plant Cell and Environment*, 29:1158–1170, 2005.

26. R. Szeliski and P. Golland. Stereo matching with transparency and matting. *International Journal of Computer Vision*, 32(1):45–61, 1999.

27. J. H. van Hateren. Theoretical predictions of spatiotemporal receptive fields of fly LMCs, and experimental validation. *Journal of Comparative Physiology A*, 171:157–170, 1992.

28. Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *IEEE Conf. on Computer Vision and Pattern Recognition*, pages 520–526, 1997.

29. Y. Weiss and D. J. Fleet. *Probabilistic Models of the Brain: Perception and Neural Function*, chapter Velocity likelihoods in biological and machine vision, pages 77–96. MIT Press, 2002.

30. S. Zacks. *Stochastic Visibility in Random Fields, Lecture Notes in Statistics 95.* Springer Verlag, 1994.