

Visibility of Multiple Cameras in a Scene with Unknown Geometry

Liuxin ZHANG and Yunde JIA

*Beijing Laboratory of Intelligent Information Technology,
School of Computer Science, Beijing Institute of Technology, Beijing 100081, P.R. CHINA
Email: zhangliuxin@bit.edu.cn jiayunde@bit.edu.cn*

Abstract—In this paper, we investigate the problem of determining the visible regions of multiple cameras in a 3D scene without a priori knowledge of the scene geometry. Our approach is based on a variational energy functional where both the unresolved visibility information of multiple cameras and the unknown scene geometry are included. We cast visibility estimation and scene geometry reconstruction as an optimization of the variational energy functional amenable for minimization with the Euler-Lagrange driven evolution. Starting from any initial value, the accurate visibility of multiple cameras as well as the true scene geometry can be obtained at the end of the evolution. Experimental results show the validity of our approach.

Keywords-visibility estimation; image-based modeling; level set method; implicit surface; multiple cameras;

I. INTRODUCTION

Visibility estimation of multiple cameras is to determine the visible and invisible regions of some given cameras in a 3D scene due to the presence of obstacles. It has played a central role in many applications such as modeling, rendering, visualization, surveillance, navigation, etc.

Computational geometry and combinatorics are the primary tools to solve the visibility problem. The combinatorial approach [1] defines visibility on polygons and more general planar environments with special structure. The simplified representation of the environment is a major limitation of this algorithm, and its extension to three dimensions may be extremely complicated. Visibility algorithms with the ray tracing technique [2] in computational geometry can be applied to general types of environment and easily extended to three dimensions, but tracing a ray and checking whether it intersects the obstacle surfaces explicitly are more computationally expensive. Tsai [3] introduced an implicit ray tracing technique which can avoid the time-consuming calculation and determine the visibility of a single viewpoint efficiently, but it needs to know the geometry of the obstacles to construct a level set representation of the scene. A more thorough discussion about the state-of-the-art solutions for the visibility problem can be found in [4], and nearly all of them require a priori knowledge of the scene geometry. However, this information may not be available in some real life applications, such as navigation in an unknown scene.

In this paper, we consider the visibility problem of multiple cameras in a 3D scene with unknown geometry. Similar to [3], the problem is formulated in an implicit framework

where the scene geometry and the visibility information are represented by level set functions. But the differences are that we extend the implicit ray tracing technique to 3D space incorporating multiple views, and combine visibility estimation with scene geometry reconstruction through an image-based modeling technique. These make our approach possible to handle the visibility problem even when the true scene geometry is unknown.

The rest of the paper is organized as follows: Section 2 gives the level set representation of visibility. Section 3 explains how this representation of visibility is incorporated into an image-based modeling technique. The experimental results and conclusions are discussed in Section 4 and Section 5 respectively.

II. LEVEL SET REPRESENTATION OF VISIBILITY

We assume that the obstacles in the scene are some closed and opaque objects. When the geometry of these obstacles is known, a level set representation of the scene can be constructed. In this representation, the surfaces of the obstacles are considered as the zero level set of a real valued function ϕ . Furthermore, the points where $\phi < 0$ represent the interior of the obstacles, and the points where $\phi > 0$ are the exterior of the obstacles. This real valued function ϕ is called the level set function.

Suppose there are N viewpoints in the scene denoted by $\mathbf{v}_i (i = 1, 2, 3, \dots, N)$, and any other point in the scene is denoted by \mathbf{x} . According to [3], the visibility status of the point \mathbf{x} with respect to the viewpoint \mathbf{v}_i can be represented by a level set function ψ :

$$\psi(\mathbf{v}_i, \mathbf{x}) = \min_{\xi \in \overline{\mathbf{x}\mathbf{v}_i}} (\phi(\xi)), \quad (1)$$

where ξ is a point on the ray $\overline{\mathbf{x}\mathbf{v}_i}$, and ϕ is the level set representation of the scene. If $\psi > 0$, then \mathbf{x} is visible to \mathbf{v}_i ; otherwise, \mathbf{x} is invisible to \mathbf{v}_i .

Consider the numerical implementation of (1). A 3D Cartesian grid, described as $\{(x_i, y_j, z_k) | 0 \leq i \leq nx, 0 \leq j \leq ny, 0 \leq k \leq nz\}$ in a rectangular domain Ω_d , is used to discretize the whole scene Ω , and all of the functions and quantities are defined on this grid. With this notation, a voxel in Ω_d is defined as a cube with vertices $\{(x_{i+p}, y_{j+q}, z_{k+r}) | p, q, r \in \{0, 1\}\}$. At each grid node $\mathbf{x} = (x_i, y_j, z_k)$, we first find the voxel containing \mathbf{x} . From

this voxel, we find the face F that intersects the ray $\overline{\mathbf{x}\mathbf{v}_i}$. If $\overline{\mathbf{x}\mathbf{v}_i}$ intersects F at a point, we denote this intersection by \mathbf{x}' . If $\overline{\mathbf{x}\mathbf{v}_i}$ lies in F , this problem is reduced to a 2D problem, and \mathbf{x}' is the intersection of $\overline{\mathbf{x}\mathbf{v}_i}$ and an edge of F . In fact, \mathbf{x}' can be regarded as the closest point right before \mathbf{x} in $\overline{\mathbf{x}\mathbf{v}_i}$ within the accuracy of the Cartesian grid.

Note that the visibility status of points satisfies a causality condition: if a point is occluded, then all other points behind it in the same ray are also occluded. Under this condition, we can use a recursion to update the value of ψ in (1):

$$\psi(\mathbf{v}_i, \mathbf{x}) = \min(\psi(\mathbf{v}_i, \mathbf{x}'), \phi(\mathbf{x})). \quad (2)$$

In most cases, \mathbf{x}' does not lie on the grid node, and we need to interpolate the value of $\psi(\mathbf{v}_i, \mathbf{x}')$ from the grid nodes closest to \mathbf{x}' in F . In order to make the interpolation feasible, the values of ψ on all the neighboring grid nodes around \mathbf{x}' must have been computed before estimating $\psi(\mathbf{v}_i, \mathbf{x})$. This can be achieved by organizing all the grid nodes in Ω_d in a special sequence. Tsai [3] reported an implementation of this sequence in a two dimensional form, while we construct it in a more reasonable way in three dimensions.

We first find the voxel C_0 in Ω_d containing the viewpoint \mathbf{v}_i , i.e., $\mathbf{v}_i \in C_0 := [x_{i_0}, x_{i_0+1}] \times [y_{j_0}, y_{j_0+1}] \times [z_{k_0}, z_{k_0+1}]$. The special sequence of the grid nodes is denoted by Q . We can therefore start from C_0 and store the grid nodes into Q following the ray directions outwards. In order to represent our idea simply, we use three parameters s_1, s_2 , and s_3 to describe multiple directions from C_0 for i, j and k index respectively. The relationship between s_1 and i is given by

$$i = \begin{cases} i_0 - 1, i_0 - 2, i_0 - 3, \dots, 0 & \text{when } s_1 = -1 \\ i_0, i_0 + 1 & \text{when } s_1 = 0 \\ i_0 + 2, i_0 + 3, i_0 + 4, \dots, nx & \text{when } s_1 = +1 \end{cases}$$

and similarly, we have the relationship between s_2 and j , as well as s_3 and k . In the first step, put the grid nodes (x_i, y_j, z_k) where (i, j, k) satisfy $(s_1, s_2, s_3) = (0, 0, 0)$ into Q . Actually, (x_i, y_j, z_k) are the vertices of C_0 in this case (see Fig. 1(a), red round points). In the second step, put the grid nodes (x_i, y_j, z_k) where (i, j, k) satisfy $(s_1, s_2, s_3) = (0, 0, -1), (0, 0, 1), (0, -1, 0), (0, 1, 0), (-1, 0, 0),$ or $(1, 0, 0)$ into Q (see Fig. 1(b), blue square points). In the third step, put the grid nodes (x_i, y_j, z_k) where (i, j, k) satisfy $(s_1, s_2, s_3) = (0, -1, -1), (0, -1, 1), (0, 1, -1), (0, 1, 1), (-1, 0, -1), (-1, 0, 1), (1, 0, -1), (1, 0, 1), (-1, -1, 0), (-1, 1, 0), (1, -1, 0),$ or $(1, 1, 0)$ into Q (see Fig. 1(c), green triangle points). In the last step, put the remaining grid nodes (x_i, y_j, z_k) where (i, j, k) satisfy $(s_1, s_2, s_3) = (-1, -1, -1), (-1, -1, 1), (-1, 1, -1), (-1, 1, 1), (1, -1, -1), (1, -1, 1), (1, 1, -1),$ or $(1, 1, 1)$ into Q (see Fig. 1(c), purple star points).

After all the grid nodes are inserted into Q , the visible regions of \mathbf{v}_i in Ω_d can be represented by a level set function ψ . A simple algorithm for this reads:

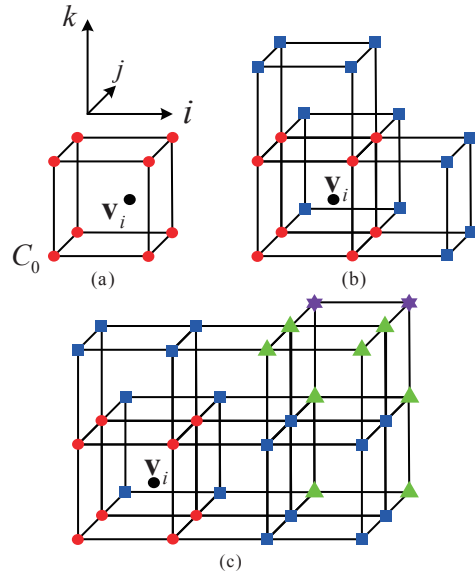


Figure 1. A special sequence of the grid nodes

```

for each grid node  $\mathbf{x}$  in the queue  $Q$ 
  if  $\mathbf{x}$  is the vertices of  $C_0$ 
    Set  $\psi(\mathbf{v}_i, \mathbf{x}) = \phi(\mathbf{x})$ ;
  else
    Find  $\mathbf{x}'$  right before  $\mathbf{x}$  in the ray direction  $\overline{\mathbf{x}\mathbf{v}_i}$ ;
    Update  $\psi(\mathbf{v}_i, \mathbf{x}) = \min(\psi(\mathbf{v}_i, \mathbf{x}'), \phi(\mathbf{x}))$ ;
  end
end

```

III. VISIBILITY OF UNKNOWN SCENE

In some applications, when the true scene geometry is unknown, the level set representation of the scene may not be available. Equation (2) cannot be employed directly to estimate the visibility of each camera in the scene. In this case, some techniques for recovering the true scene geometry should be taken. We mainly focus on the image-based modeling technique that reconstructs the unknown scene geometry from a number of scene images.

We cast image-based modeling as an energy minimization problem by assigning an energy cost based on the unresolved visibility of the cameras. In general, the surface of the unknown scene is supposed to be nearly Lambertian, i.e., only elements belonging to the scene surface should have a consistent appearance in the input images (*be photo-consistent*). Under this assumption, the energy cost of a surface element is defined as

$$\mathbf{A}(\mathbf{x}, \mathbf{n}) = \frac{1}{T} \sum_{i,j=1, i \neq j}^N H(\psi(\mathbf{v}_i, \mathbf{x})) H(\psi(\mathbf{v}_j, \mathbf{x})) \mathbf{A}_{ij}(\mathbf{x}, \mathbf{n}), \quad (3)$$

where (\mathbf{x}, \mathbf{n}) is an infinitesimal element located at point \mathbf{x} and having unit outward normal \mathbf{n} , T is the number of items in the summation, N is the number of all views, $H(\cdot)$ is the

Heaviside function, and \mathbf{A}_{ij} between two visible views i and j is based on the normalized cross correlation (NCC):

$$\mathbf{A}_{ij}(\mathbf{x}, \mathbf{n}) = 1 - \text{NCC}(\mathbf{x}, \mathbf{n}, i, j). \quad (4)$$

The term $\text{NCC}(\mathbf{x}, \mathbf{n}, i, j)$ is the normalized cross correlation of the projections of the element (\mathbf{x}, \mathbf{n}) in views i and j . We direct readers to [5] for more details on how to compute this term. $\mathbf{A}_{ij}(\mathbf{x}, \mathbf{n})$ ranges between 0 (best correlation) and +2 (worst). Thus, the smaller is the energy cost $\mathbf{A}(\mathbf{x}, \mathbf{n})$, the more likely does the element (\mathbf{x}, \mathbf{n}) belong to the surface of the scene. By setting the scene surface as the zero level set of ϕ and integrating the elements' costs $\mathbf{A}(\mathbf{x}, \mathbf{n})$ over the scene surface, we get the overall surface energy cost:

$$E_P(\phi) = \int_{\Omega} \mathbf{A}(\mathbf{x}, \mathbf{n}) \delta(\phi) |\nabla \phi| d\Omega, \quad (5)$$

where $\delta(\cdot)$ is the univariate Dirac function, and the whole scene is located within some bounding volume Ω . The integral in (5) is actually the weighted surface area of the zero level set of ϕ with the weight coefficient $\mathbf{A}(\mathbf{x}, \mathbf{n})$. So, a level set function ϕ minimizing $E_P(\phi)$ can be regarded as the level set representation of the true scene geometry.

Now, we have expressed the surface photo-consistency with an energy term $E_P(\phi)$. Minimizing $E_P(\phi)$ solely is, however, uninteresting as it has an obvious global minimum $\{\mathbf{x} | \phi(\mathbf{x}) = 0, \mathbf{x} \in \Omega\} = \emptyset$ that equals zero. In fact, it has been demonstrated in [6] that in the absence of noise, different scenes can be consistent with the same set of color images. Therefore, surface reconstruction based solely on photo-consistency is an ill-posed problem. To regularize it, we propose to augment the energy functional with a regularization term:

$$E_R(\phi) = \lambda_1 E_{R_1}(\phi) + \lambda_2 E_{R_2}(\phi), \quad (6)$$

where λ_1 and λ_2 are constants, and the terms $E_{R_1}(\phi)$ and $E_{R_2}(\phi)$ are defined as

$$E_{R_1}(\phi) = \int_{\Omega} (\min_{i=1}^N I_i(\mathbf{x}))^2 \cdot H(\phi) d\Omega \quad (7)$$

and

$$E_{R_2}(\phi) = \int_{\Omega} (255 - \min_{i=1}^N I_i(\mathbf{x}))^2 \cdot H(-\phi) d\Omega \quad (8)$$

respectively, where $I_i(\mathbf{x})$ is the silhouette value of the projection of the point $\mathbf{x} \in \Omega$ onto the viewpoint \mathbf{v}_i :

$$I_i(\mathbf{x}) = \begin{cases} 255 & \mathbf{x} \text{ projects to the foreground in } \mathbf{v}_i \\ 0 & \mathbf{x} \text{ projects to the background in } \mathbf{v}_i \end{cases}. \quad (9)$$

Minimizing the energy term $E_R(\phi)$ in (6) will force the zero level set of ϕ to converge to the visual hull [7], which can supply some good prior knowledge about the scene geometry and reduce the ambiguity of the final result.

Because the level set function may develop very sharp

and/or flat shape during the evolution, it is crucial to keep the evolving level set function close to a signed distance function for stable purpose. Here, we use a penalizing term

$$P(\phi) = \int_{\Omega} \frac{1}{2} (|\nabla \phi| - 1)^2 d\Omega \quad (10)$$

to characterize how close a function ϕ is to a signed distance function in Ω . It will force the level set function ϕ to be close to a signed distance function in the evolution process. We direct readers to [8] for more details about this energy term.

Combining all the energy terms (5), (6), and (10) leads to the full energy functional for the reconstruction:

$$E(\phi) = \alpha P(\phi) + \beta E_P(\phi) + E_R(\phi). \quad (11)$$

The second and third energy terms drive the zero level set of ϕ toward the true scene geometry, while the first energy term penalizes the deviation of ϕ from a signed distance function during its evolution. All the terms in the final energy functional are weighted by constants α , β and 1.

The function ϕ that minimizes $E(\phi)$ satisfies the Euler-Lagrange equation $\partial E / \partial \phi = 0$. Employing an artificial time variable $t > 0$, the steepest descent process for minimizing $E(\phi)$ is the following gradient flow:

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= - \frac{\partial E}{\partial \phi} \\ &= \alpha (\Delta \phi - \kappa) + \beta \delta(\phi) (\nabla \mathbf{A} \cdot \mathbf{n} + \mathbf{A} \cdot \kappa) \\ &\quad - \lambda_1 \delta(\phi) (\min_{i=1}^N I_i)^2 + \lambda_2 \delta(\phi) (255 - \min_{i=1}^N I_i)^2, \end{aligned} \quad (12)$$

where Δ is the Laplacian operator, κ and \mathbf{n} are the mean curvature and unit normal of the scene surface respectively, which can be calculated by

$$\kappa = \text{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) \quad \text{and} \quad \mathbf{n} = \frac{\nabla \phi}{|\nabla \phi|}. \quad (13)$$

In practice, the Dirac function $\delta(\phi)$ in (12) is replaced by a regularized form $\delta_{\varepsilon}(\phi)$ defined as

$$\delta_{\varepsilon}(\phi) = \frac{1}{\pi} \frac{\varepsilon}{\varepsilon^2 + \phi^2} \quad (14)$$

with a small $\varepsilon > 0$. As discussed in [9], this regularized form of $\delta(\phi)$ can give (12) the tendency to compute a global minimum of $E(\phi)$. In this way, the evolution result of our method is not sensitive to the initial value of ϕ . The final evolution equation of ϕ is hence given by

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \alpha (\Delta \phi - \kappa) + \beta \delta_{\varepsilon}(\phi) (\nabla \mathbf{A} \cdot \mathbf{n} + \mathbf{A} \cdot \kappa) \\ &\quad - \lambda_1 \delta_{\varepsilon}(\phi) (\min_{i=1}^N I_i)^2 + \lambda_2 \delta_{\varepsilon}(\phi) (255 - \min_{i=1}^N I_i)^2. \end{aligned} \quad (15)$$

Because of the penalizing term in (10), we no longer need the upwind scheme to numerically implement (15). Instead, all the spatial partial derivatives $\partial \phi / \partial x$, $\partial \phi / \partial y$ and $\partial \phi / \partial z$ in (15) are approximated by the central difference, and the

temporal partial derivative $\partial\phi/\partial t$ is approximated by the forward difference. Thus, the approximation of (15) can be simply written as an iterative process:

$$\phi_{i,j,k}^{n+1} = \phi_{i,j,k}^n + \Delta t \cdot f(\phi_{i,j,k}^n), \quad (16)$$

where Δt is the time step, $\phi_{i,j,k}^n$ is the value of ϕ on the grid node (x_i, y_j, z_k) at the time n , and $f(\phi_{i,j,k}^n)$ is the approximation of the right hand side in (15). Once the initial level set function ϕ^0 has been given, we can constantly update ψ and ϕ through (2) and (16) until the iteration is over. The final ψ and ϕ is considered as the true visibility of each camera and the true scene geometry.

IV. EXPERIMENTAL RESULTS

We show results of our method on the Middlebury datasets [10, 11]. The datasets consist of two scenes, a dinosaur scene and a temple scene, and three different sets of input images for each scene, with viewpoints forming a sparse ring, a full ring, and a full hemisphere around the scene. The advantage of using these datasets is that the ground truth laser-scanned scene geometry is known and the quality of the reconstructed results can be evaluated. We perform our experiments on the dinoSparseRing and the templeSparseRing inputs (see Fig. 2, first and third column). The initialization is taken as a sphere with arbitrary locations and radius, and the unknown scene geometries of these two datasets can be reconstructed through an evolution process (see Fig. 2, second and fourth column). Because it is usually difficult to get the ground truth visibility of each viewpoint, we cannot evaluate the accuracy of our visibility estimation results directly. Here, we evaluate the quality of the reconstructed scene geometry instead. The evaluation results are shown in Table I, where the accuracy metric is the distance d (in millimeters) that brings 90% of the reconstructed surface within d from some points on the ground truth surface, and the completeness score measures the percentage of points in the ground truth surface that are within 1.25mm of the reconstructed surface. It is clear that our method can yield good results.

Table I
QUANTITATIVE EVALUATION OF THE RECONSTRUCTION RESULTS.

DinoSparseRing		TempleSparseRing	
Acc. (mm)	Comp. (%)	Acc. (mm)	Comp. (%)
0.89	93.4	1.02	92.7

V. CONCLUSION

In this paper, we have presented an approach to estimating the visibility of multiple cameras in a scene with unknown geometry. It relies on an image-based modeling technique. The accurate visibility information of multiple cameras as well as the true scene geometry can be obtained during an evolution process. The numerical implementation of the evolution process is easy and efficient by using very simple finite difference scheme. The main advantage of our method



Figure 2. Reconstruction results. First column: 2 of the 16 images in dinoSparseRing. Second column: Reconstruction results of dinosaur in front and back view. Third column: 2 of the 16 images in templeSparseRing. Fourth column: Reconstruction results of temple in front and back view.

is its capability for the visibility problem even when the prior knowledge of the scene is unavailable.

ACKNOWLEDGEMENTS

This work was partially supported by the Natural Science Foundation of China (90920009) and the Chinese High-Tech Program (2009AA01Z323).

REFERENCES

- [1] W. P. Chin and S. Ntafos, "Shortest watchman routes in simple polygons," *Discrete and Computational Geometry*, vol. 6, pp. 9–31, 1991.
- [2] S. C. Franklin and F. David, "Interactive ray tracing with the visibility complex," *Computers and Graphics*, vol. 23, no. 5, pp. 703–717, 1999.
- [3] R. Tsai, L. T. Cheng, P. Burchard, S. Osher, and G. Sapiro, "Dynamic visibility in an implicit framework," UCLA CAM Report 02-06, Tech. Rep., 2002.
- [4] F. Durand, "3d visibility: analysis study and applications," Ph.D. dissertation, MIT, 1999.
- [5] T. Alejandro, S. B. Kang, and S. Seitz, "Multi-view multi-exposure stereo," in *The 3rd International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006, pp. 861–868.
- [6] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," *International Journal of Computer Vision*, vol. 38, no. 3, pp. 199–218, 2000.
- [7] A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 2, pp. 150–162, 1994.
- [8] C. M. Li, C. Y. Xu, C. F. Gui, and M. D. Fox, "Level set evolution without re-initialization: a new variational formulation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [9] T. Chan and L. Vese, "Active contours without edges," *IEEE Transactions on Image Processing*, 2001.
- [10] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, pp. 519–528.
- [11] The middlebury multi-view stereo webpage. [Online]. Available: <http://vision.middlebury.edu/mview/>