Omnivergent Stereo

Heung-Yeung Shum
Vision Technology Group
Microsoft Research
Redmond, WA 98052

Adam Kalai
Department of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Steven M. Seitz*
The Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

The notion of a virtual sensor for optimal 3D reconstruction is introduced. Instead of planar perspective images that collect many rays at a fixed viewpoint, *omnivergent cameras* collect a small number of rays at many different viewpoints. The resulting 2D manifold of rays are arranged into two multiple-perspective images for stereo reconstruction. We call such images *omnivergent images*, and the process of reconstructing the scene from such images *omnivergent stereo*. This procedure is shown to produce 3D scene models with minimal reconstruction error, due to the fact that for any point in the 3D scene, two rays with maximum vergence angle can be found in the omnivergent images. Furthermore, omnivergent images are shown to have horizontal epipolar lines, enabling the application of traditional stereo matching algorithms, without modification. Three types of omnivergent virtual sensors are presented: spherical omnivergent cameras, center-strip cameras and dual-strip cameras.

1 Introduction

Planar perspective images have long been the primary substrate of computer vision algorithms, due to the ease of acquiring such images from conventional cameras. The last few years, however, have seen a growing interest in fundamentally new types of image representations for a variety of applications in computer vision and computer graphics. Such new representations include *mosaics* [1, 2, 3, 4] that encode radiance information from all angles converging at a single optical center, as well as non-perspective representations like light fields [5], lumigraphs [6], multiperspective panoramas [7], and manifold mosaics [8]. These new image representations can be derived from one or more conventional images or acquired directly from new imaging devices, e.g., catadioptic cameras [9]. In the former case, we can think of derived images as representing the output of a *virtual sensor* with certain characteristics.

* The support of a grant from the Microsoft Corporation is gratefully acknowleged. Part of this work was conducted while Steven Seitz was employed by the Vision Technology Group at Microsoft Research.

The emergence of these new types of images suggests a powerful new paradigm for solving computer vision problems–rather than adapting the algorithm to suit the image, we can modify the image to best suit the algorithm. This approach has a powerful advantage: it becomes possible to *boost* the performance of a specific vision algorithm or even an entire class of algorithms by providing input data that is carefully crafted to improve the performance of the target algorithm(s). To demonstrate this approach, in this paper we formulate a new type of virtual sensor that is specifically designed to optimize the performance of stereo reconstruction algorithms.

Our approach is motivated by the fact that perspective image sequences are poorly suited as input for 3D reconstruction tasks for a variety of reasons. First, image sequences are highly redundant and represent scene appearance in a form that is difficult to process. Second, storing, transmitting, and processing long image sequences at high-resolution is cumbersome and computationally expensive on today's computer workstations. Third, traditional and multi-baseline stereo algorithms do not easily scale to handle hundreds or thousands of input images. In contrast, an ideal sensor would sample scene radiance in a manner that is designed to optimize the task of scene reconstruction. Furthermore, images from such a sensor would be amenable to efficient processing with scanline stereo algorithms.

To accomplish these objectives, we introduce a virtual sensor called an *omnivergent camera* that has the following properties:

- **Omnivergence**: every point in the scene is imaged from two cameras that are vergent on that point (see Fig. 1) *with maximum vergence angle*. This strategy is shown to simultaneously optimize the reconstruction accuracy of every point in the scene.

- **Uniform Accuracy**: scene radiance is sampled uniformly in all directions, yielding isotropic depth resolution. In contrast, stereo reconstructions obtained from planar or panoramic [10, 11] images are shown to generate strongly biased (non-isotropic) scene reconstructions.

- **Compactness**: radiance information from hundreds or thousands of input images is distilled into two or more composite images, thereby reducing an N-view reconstruction problem to a binocular stereo matching task, with minimal loss of accuracy.

- **Boosting Property**: due to their linear epipolar geometry, omnivergent images may be used directly as input to existing stereo algorithms, providing an algorithm-independent mechanism for optimizing reconstruction accuracy.

The idea behind the omnivergent camera is as follows, suppose you are given the task of providing two images as input to a binocular stereo matching algorithm, with the objective of obtaining the most accurate possible scene reconstruction. Suppose further that you can acquire images from as many viewpoints as you like, but may store at most a two-dimensional set of pixels (viewing rays) from which to construct the two images (note that the space of all viewing rays is five-dimensional [10]). Finally, you should construct images that satisfy the constraints of most stereo algorithms, i.e., epipolar lines should correspond to horizontal lines in the image.

Given this objective, which viewpoints and viewing rays would you select to form your images? In this paper, we demonstrate that the answer can be obtained on a point-by-point basis: for each 3D scene point *P*, we wish to choose the best pair of views from which to reconstruct *P*. Given certain assumptions, these views correspond to the two cameras that are vergent on *P* and for which the vergence angle is maximized. This strategy can be shown to yield a 2D manifold of rays that, when arranged into two images, produce the best possible stereo reconstruction. These two images can be thought of as an *omnivergent stereo pair*, output from a virtual sensor that is simultaneously vergent on every point in the scene. We call such images *omnivergent panoramas*, and the process of reconstructing the scene from such images *omnivergent stereo*.

This work was inspired by recent work on stereo from panoramas [10, 11]. The beauty of the panoramic stereo approach is that the same class of algorithms that have proven effective for binocular stereo reconstruction can also produce 360° reconstructions, simply by changing the input to the algorithm. However, stereo on panoramas has some critical shortcomings–first the epipolar geometry for cylindrical panoramas is nonlinear, significantly complicating the stereo search procedure [10]. Second, and more important, we shall see that stereo on panoramic images does not produce a fair reconstruction (accuracy is good in some regions but extremely poor in others). These two limitations are remedied by the omnivergent approach set forth in this paper.

Two of the mosaic representations presented in this paper, namely the *center-strip* and *dual-strip* varieties, were independently discovered by Peleg and Ben-Ezra [12] for the purpose of creating stereo panoramas for scene visualization. A similar stereo panorama representation based on rotating a stereo head was previously proposed by Huang and Hung [13]. In both of these approaches, an explicit 3D model is not computed, rather the images are visually fused with stereo glasses to achieve a stereo effect. In contrast, our objective is to obtain an explicit 3D scene reconstruction using stereo matching algorithms.

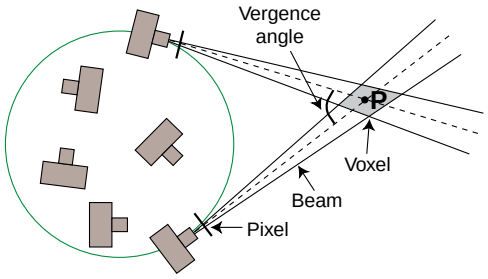


Figure 1: To reconstruct a point P with minimal uncertainty, the omnivergent sensor captures two beams of maximal vergence angle that contain P.

In the remainder of this paper, we describe the omnivergent stereo approach in detail. Section 2 introduces the omnivergent approach for the case of a 2D camera confined to move within a circular region of the plane. Sections 3 and 4 generalize the analysis to 3D where a camera moves along a sphere and a circle, respectively. Section 5 presents a practical implementation by taking two symmetrical off-centered slit images from a regular camera rotating around a circle. Several examples of omnivergent images are given, with results of processing these images with existing stereo algorithms. The paper concludes with a summary and future work.

2 Omnivergent Imaging

Suppose you can move a camera anywhere within some fixed region of space, with the objective of reconstructing the scene outside of that region as accurately as possible. Suppose further that you can acquire images from as many viewpoints as you like, but may store only a fixed number of pixels, due to limitations in resources. Which pixels would you save and from which cameras? In this section, we study the case of a camera moving within a circular region of the plane, depicted in Fig. 1. In order to accurately localize a point P, we argue that it is best to choose a pair of cameras that are vergent on P with maximum vergence angle. By choosing different cameras to reconstruct different scene points, it is possible to model virtual sensor that

is simultaneously vergent on every point on the scene.

In this section, we introduce the *omnivergent virtual sensor*, which maximizes the vergence angle for all points in the scene, and compare its performance to other plausible sensors for the specific case of a camera viewpoints within a circle. A theoretical justification for why maximizing vergence angle is desirable is given in the appendix. In our terminology, a pixel captures a *beam* of rays, as shown in Fig. 1, and the *direction* of the beam is defined to be its angle bisector. The intersection of two beams is a *voxel* and a *virtual sensor with resolution* N denotes a collection of N beams.

The central idea behind binocular stereo algorithms is that the location of a point in the world is identified by triangulation from corresponding pixels in two input images. Each pixel corresponds to a beam of viewing rays that project to a finite region of the image plane. Given two pixels in correspondence, one can deduce that the true location of the point is in a particular voxel, corresponding to the intersections of the two pixel beams (Fig. 1). This voxel represents the uncertainty of the reconstruction, since the point may lie anywhere within the voxel region. For purposes of visualization, it is convenient to assume that the point is at the *center* of the voxel, corresponding to the intersection of the two beam directions.

By plotting the voxel centers for a specified camera configuration, we can visualize the spatial sampling of a particular virtual sensor and the quality of the resulting reconstruction. Note that if we limit the number of beams (pixels) to N, there is a finite number of possible beam intersections and thus a finite number of voxels. To begin with, consider the traditional binocular stereo configuration, of two cameras oriented in a parallel or vergent configuration. A critical shortcoming of this configuration is that it enables reconstructing only the subset of the scene that is in front of the cameras. A recently proposed solution is to use two or more *panoramic* cameras [10, 11] instead. We can model such a pair of panoramic images as representing the output of a single sensor with two focal points.

Fig. 2(b) shows the voxel centers for the binocular panoramic sensor used in [10, 11], with 180 beams total. Clearly, this approach yields very biased reconstructions, as the voxels are distributed highly non-uniformly. In particular, the portion of the scene that lies near the line between the two camera centers is reconstructed with relatively poor accuracy, where the distance between voxel centers is much higher than in other parts of the scene.

To improve uniformity, a natural idea would be to use three cameras, as shown in Fig. 2(c-d). While the resulting distribution of voxels is somewhat more uniform than the binocular case, some portions of the plane are now covered by three beams and others by two, yielding an uneven distribution of voxels. Fig. 2(e-f) captures only the rays of the

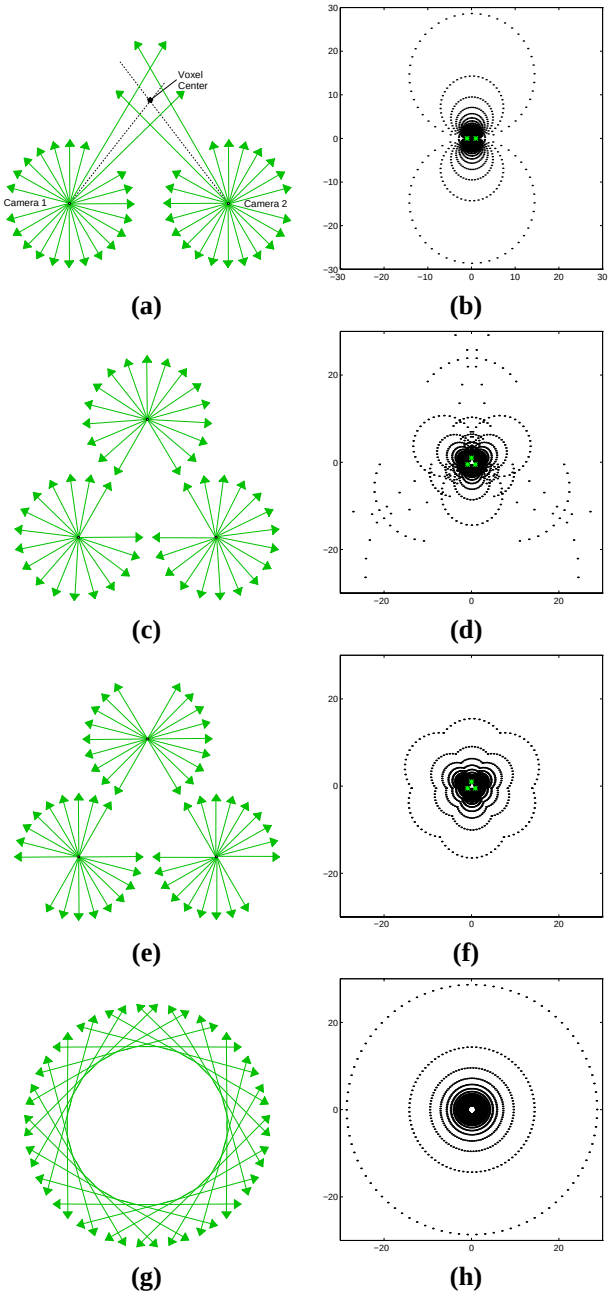
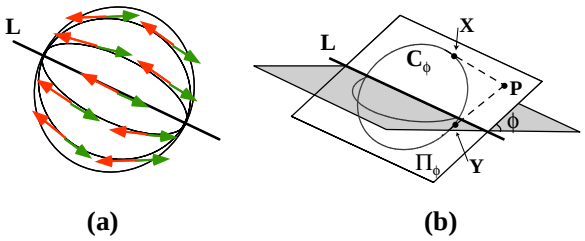


Figure 2: Each virtual sensor at left represents a configuration of pixel ray directions from cameras on a circle. The intersections of these rays define a lattice of voxels whose centers are shown at right. In each case, the sensor has 180 pixels total. (a) corresponds to a panoramic binocular camera rig and (g) to an omnivergent sensor. (g) Incoming light sensed along the rays tangent to a circle can be recorded in two separate 1D images, one for the clockwise and one for the counter-clockwise rays. Performing stereo matching on these two *omnivergent images* yields a reconstruction with optimal characteristics.



(a) **(b)**

Figure 3: Construction of spherical omnivergent images. For any scene point P, we wish to choose a pair of cameras in the sphere vergent on P with maximal vergence angle. To achieve this goal, it is sufficient choose a revolute axis L of the sphere and capture rays along the longitude lines (a). (b) For any scene point P, there exists a pair of longitude rays vergent on P for which the vergence angle is maximized over the entire sphere.

exterior angles of the triangle spanned by the three optical centers. This configuration covers every point in the scene by exactly two beams, and its voxel centers are more evenly distributed. If we extend this idea to a regular n-gon instead of a triangle, then for $n = N$ we get the omnivergent sensor. Note that for a fixed number of beams, the number of voxels can vary for different sensors, due to the fact that not every pair of beams will intersect. For the four sensors depicted in the figure, the total number of voxels are approximately 2000, 3000, 3200, and 4000, respectively.

2.1 Omnivergent Images

The omnivergent sensor of resolution N captures two sets of $N/2$ beams, evenly spaced around the circle, pointing in the direction of the tangents. Half of these we call *forward* and the other half *backward* tangents, according to whether the ray points in the clockwise or counter clockwise direction, respectively (Fig. 2(g)). Recording the forward tangents in one image and the backward tangents in another yields two one-dimensional *omnivergent images*. A correspondence between pixels in these two images can be obtained using standard stereo matching algorithms (restricted to a 1D search). For any two pixels that correspond, we know precisely which two beams are stored in these pixels, and straightforward trigonometry locates the voxel of the intersection between these two beams, yielding the reconstruction and its associated uncertainty.

The omnivergent sensor has three key advantages: first, it gives uniform angular accuracy over the entire scene. Second, it reduces the N-view stereo problem to a binocular stereo problem, amenable to conventional stereo algorithms. Third, it provides a mechanism for maximizing the vergence angle for every point in the scene.

3 Spherical Omnivergent Stereo

In this section we generalize the analysis of 2D camera motion to 3D camera motion within a spherical volume. The statement of the problem is similar to the planar case: suppose you could move a camera anywhere within a closed sphere and acquire as many images as you like. For any given point P in the scene, you would like to choose a pair of rays vergent on P such that the vergence angle is maximized. For any two points X and Y in the sphere, we define the *vergence angle* to be $\angle PXY$.

It is easily seen that the desired set of rays are tangent to the sphere. For any given point P, the set of all rays from P that graze the sphere defines a circle K_P where these rays intersect the sphere. The vergence angle is maximized by choosing any pair of rays on opposite poles of K_P:

> **Spherical Vergence Property**: Let S be a sphere with center C and let P be a point outside S. $\angle PXY$ is maximized for all points X and Y in S when X and Y are on K_P and X, Y, P, and C are co-planar.

If we allow P to vary over the entire scene, it is evident that capturing the incoming light along all tangent rays to the sphere is sufficient to image every point in the scene with two rays having maximum vergence angle. Note that the set of all tangent rays is three-dimensional—in contrast, the set of all rays from points within the sphere is four-dimensional. Therefore, we can obtain a 3D omnivergent image by storing tangent rays in a volumetric data structure. While some stereo algorithms have been developed to process 3D image volumes [14, 15], storing and processing 3D images poses a daunting task. Fortunately, it is possible to further reduce the dimensionality to a 2D set of rays, which we demonstrate next using a constructive argument.

Observe that a sphere can be generated by rotating a circle 180° about an axis L passing through the center of the circle. The resulting set of circles correspond to the longitude lines of a globe with polar axis L, as shown in Fig. 3(a). Each circle C_ϕ lies in a plane Π_ϕ, and the set of all such planes defines a one-parameter family of circles about L called a *pencil* [14]. Now consider a point P outside the sphere. P must lie in at least one plane in the pencil; let Π_ϕ be one such plane[1]. Consider the two rays from P that are tangent to C_ϕ. Let X and Y be the two tangency points on C_ϕ. Note that these two rays are also tangent to the sphere and therefore X and Y lie on K_P. Since, X, Y and P all lie on Π_ϕ, a plane that passes through the center of the sphere, it follows from the vergence property that these two rays yield the maximum possible vergence angle for all rays in the sphere.

[1] Points on L lie on all planes in the pencil.

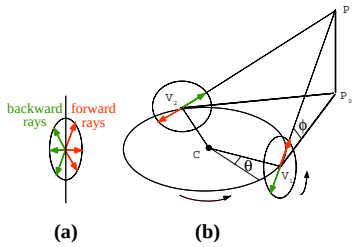


Figure 4: Construction of center-strip omnivergent images.

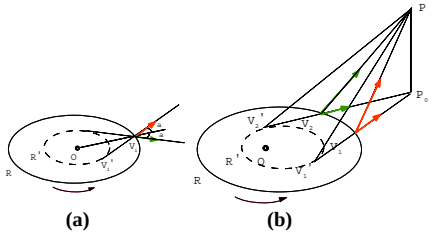


Figure 5: Construction of dual-strip images.

Consequently, it is sufficient to capture the 2D space of rays corresponding to the tangents to all circles C_ϕ. These rays represent the subset of tangents to the sphere that lie in the pencil defined by L. This construction ensures that every point in the scene is captured by a pair of rays on the sphere with maximum vergence angle. These rays may be acquired and stored in $\phi - \theta$ images by moving a camera in increments of an angle θ along C_ϕ and putting pixels corresponding to tangent rays in a row vector. Stacking row vectors for subsequent values of ϕ yields a *spherical omnivergent image* (SOI). Note that opposite sides of this image corresponding to $\theta = 0, 360$ and $\phi = 0, 360$ are identified, so the image has the topology of a torus.

3.1 Stereo on Spherical Omnivergent Images

We can capture all longitudinal tangent rays, shown in Fig. 3(a) by capturing only the forward tangents to each of the C_ϕ circles and sweeping 360° degrees in ϕ, yielding a single $\phi - \theta$ image in which each point is imaged twice[2]. Fig. 6(b) shows an example of an SOI image constructed in this way. Alternatively, we can construct an SOI stereo pair by producing two *hemispherical* omnivergent images; one for the *forward* tangents and one for the *backward* tangents. The forward (backward) image is formed by choosing the forward (backward) tangents to C_ϕ for a 180° rotation in ϕ. For stereo matching, the latter construction is useful because it produces two images directly (Fig. 7).

A key property of SOI's is that their epipolar geometry is identical to that of a pair of rectified plane-perspective images. In particular, corresponding points appear in the same horizontal scanline of both images of the spherical stereo pair. The epipolar planes for the SOI's correspond to the pencil Π_ϕ through L. To see why this is true, observe that all points in Π_ϕ are imaged in the row of the SOI pair corresponding to C_ϕ. The fact that SOI's have linear epipolar geometry is rather surprising, given the nonlinear geometry underlying the construction of these images. However, this property is extremely useful in that it enables spherical omnivergent images to be processed directly by existing stereo matching algorithms. To experimentally verify the horizontal epipolar lines, we show the results of applying a standard area-based correlation stereo algorithm to the stereo pair in Fig. 7(c).

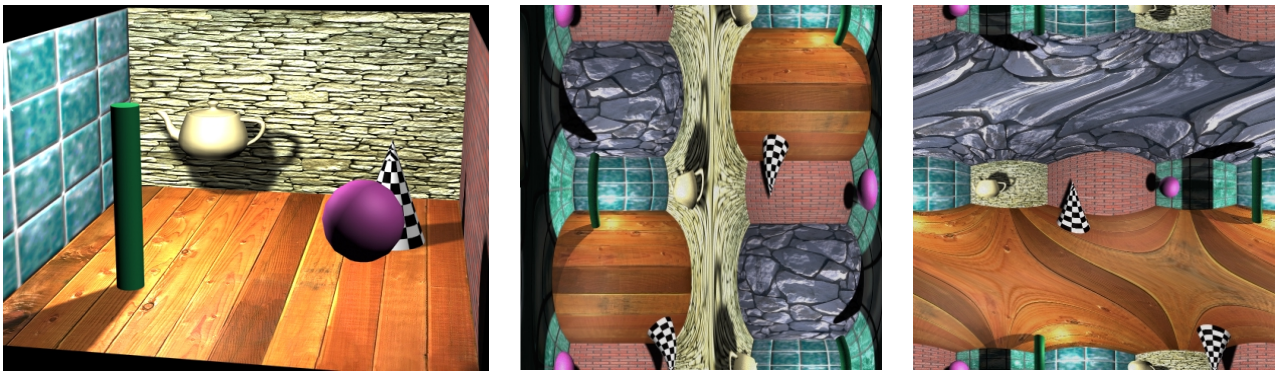
[2]Observe that the 360° rotation of C about L covers the sphere twice, and that the tangents are "flipped" in the second pass.

4 Center-Strip Stereo

Because it is difficult to move a camera over a spherical surface, we have designed another variant of an omnivergent camera, called a *center-strip omnivergent sensor*, that acquires only a single camera motion along a circular path. At each point on the circle, the camera captures incoming light along the 1D space of rays that are orthogonal to the circle radius, as shown in Fig. 4. When the optical axis of the camera is tangent to the circle and camera Y-axis points normal to the motion plane, these rays project to the center strip (center column) of the image. Center strips from consecutive positions along the circle are placed in consecutive columns of a new $\phi - \theta$ mosaic image called a *center-strip omnivergent image* (COI), an example of which is shown in Fig. 6(c). COI's are instances of *manifold mosaics* [8] and *multiperspective panoramas* [7] which were introduced by other authors, but to our knowledge have not been previously used for 3D reconstruction tasks.

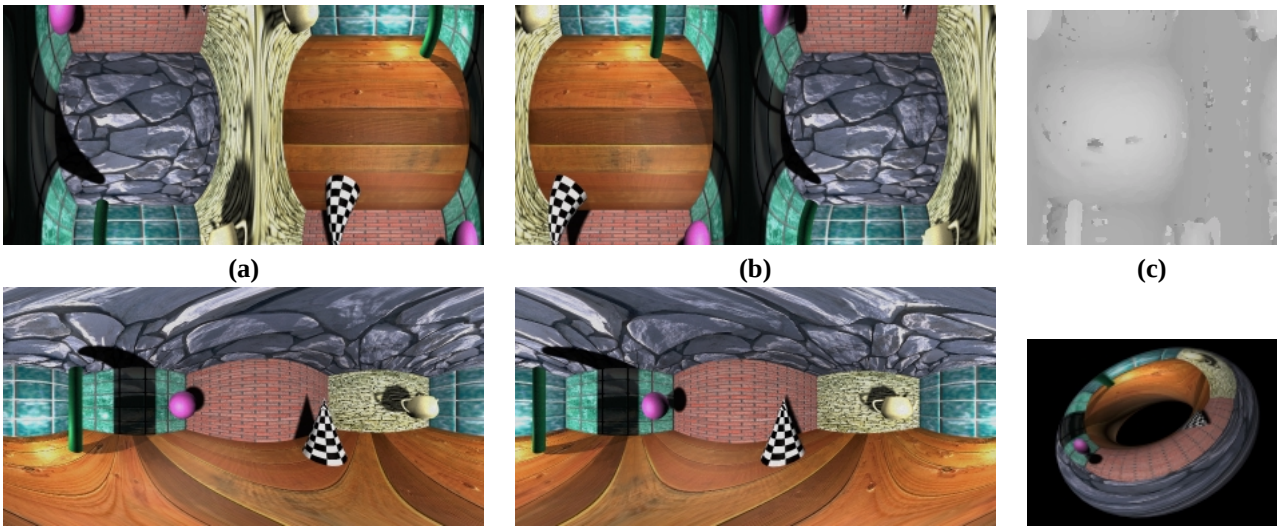
COI's have many similarities with SOI's. As in the spherical case, the COI has toroidal topology, as shown in Fig. 7(f). It can also be decomposed into a stereo pair by placing forward rays in one image and backward rays in another. The COI's also have horizontal epipolar lines; this can be shown by a simple geometric argument that is illustrated by Fig. 4(b). Consider projecting a scene point P orthographically to P_0 on the plane of the circle. P is imaged by two points V_1 and V_2 on the circle which lie on rays from P_0 that are tangent to the circle. Since $\|P_0V_1\| = \|P_0V_2\|$, it is easily seen that $\angle PV_1P_0 = \angle PV_2P_0$. Since rays in the same row of a COI correspond to the same value of ϕ (off-plane angle), it follows that scene points appear in the same row of both the forward and backward COI images.

Unlike the spherical case, however, the COI does not maximize the vergence angle for every 3D scene point. Rather, it maximizes $\angle P_0V_1V_2$ (*horizontal vergence angle*), which is a good approximation when the vertical field of view is limited.



(a) **(b)** **(c)**

Figure 6: Omnivergent images. (a) Perspective, (b) spherical omnivergent, and (c) cylindrical omnivergent images of the same synthetic room scene. Observe that each point appears twice in the omnivergent images.



(a) **(b)** **(c)**

(d) **(e)** **(f)**

Figure 7: Spherical (a-b) and cylindrical (d-e) omnivergent stereo pairs formed by decomposing the images in Fig. 6 into forward and backward tangent rays. (c) Computed disparity map of the left half of image (a). View of the cylindrical omnivergent image from Fig. 6 mapped onto a torus to illustrate its natural topology.

5 Dual-Strip Stereo

The center-strip omnivergent camera shown above requires capturing a full circle of rays at a point. Special optical devices (e.g., mirrors [16]) are needed to capture both forward and backward rays at a single point. In this section, we propose a practical alternative called a *dual-strip sensor*. Like the center-strip omnivergent sensor, a dual-strip sensor captures rays from viewpoints along a circle but is formed from two symmetrical off-center slit images. A top-down view of a dual-strip camera illustrates this property in Fig. 5(a). Both rays from the dual-strip camera on the plane are tangent to an inner circle $R' = R\sin\alpha$ where R is the radius of the outer circle that the camera moves along and α is the angle of dual-strip camera from the normal direction.

It can be easily shown (Fig. 5(b)) that the epipolar constraints from dual-strip images are again horizontal lines. Because $\|P_0V_1'\| = \|P_0V_2'\|$ (tangent to R') and $\|V_1V_2'\| = \|V_2V_2'\|$ (dual cameras are symmetric), $\angle PV_1P_0 = \angle PV_2P_0$. For dual-strip cameras, horizontal vergence angle relative to the inner circle R', not to the outer circle R, is maximized.

The significance of the dual-strip images is that they can be readily acquired from a regular perspective camera moved along a circle. Two columns of pixels, symmetric about the center, are selected from each image. One mosaic image is formed from the left column for consecutive positions on the circle, and another from the right. A pair of dual-strip images formed in this manner is shown

Figure 8: Dual-strip images formed by moving a single perspective camera 360° on a circle and building mosaics from the $60th$ column (top) and $260th$ column (bottom) of the input image sequence.

in Fig. 8. These images were created by stacking the $60th$ and $260th$ columns, respectively, from a sequence of 1350 frames (with resolution of 320×240) taken by a regular digital video camera. These two images have significant horizontal parallax but no vertical parallax due to the linear epipolar geometry.

6 Conclusions

Based on the belief that traditional perspective images are not best suited for solving many computer vision tasks, this paper introduced the concept of a virtual sensor. A key idea behind this approach is that it enables optimizing the *input* to computer vision algorithms in order to produce superior results. Toward this end, we introduced a new family of sensor called *omnivergent sensors* that are simultaneously vergent on every point in the scene, with maximum vergence angle. These sensors are ideally suited for the task of 3D reconstruction. Rather than physically construct these sensors, we described how to create omnivergent images by processing a sequence of perspective input images.

Omnivergent images represent scene appearance simultaneously from numerous perspective viewpoints, as opposed to traditional single perspective images, and are constructed so that the vergence angle for any 3D point is maximized. Consequently, omnivergent stereo produces accurate 3D scene models with minimal reconstruction error. Furthermore, omnivergent images have horizontal epipolar geometry, enabling the direct application of conventional binocular stereo algorithms for determining pixel correspondence.

We presented three different flavors of omnivergent sensors. The first is a spherical omnivergent sensor that satisfies the maximum vergence angle property. The second is a center-strip sensor that maximizes only the horizontal vergence angle, but is more straightforward to implement in practice. The center-strip omnivergent sensor is a good choice when the scene is relatively far away and the vertical field of view is limited. We also presented a dual-strip sensor that can be easily implemented by moving a regular camera on a circle.

There remain several important topics to be further studied. First, in this paper we considered only the problem of binocular stereo–better results may be obtained in practice by formulating an N-ocular stereo problem [17]. Because the omnivergent cameras have maximum vergence angle, stereo matching with omnivergent images is more sensitive to lighting changes than narrow baseline stereo images. Additional images may be generated by choosing more off-tangent rays and/or choosing different sets of input viewpoints [18]. Second, the generalization of the circle and sphere case to arbitrary camera paths and surfaces is an important topic of future work. Some of the analysis generalizes to any smooth surface–i.e., saving tangent rays is sufficient to maximize the vergence angle. The choice of camera path should ideally depend on *a priori* information about scene occupancy. Lastly, we are investigating how to build *real* devices that implement the omnivergent sensors presented in this paper in hardware, using mirrors to collect front and backwards rays with a single CCD.

References

[1] R. Kumar, P. Anandan, M. Irani, J. Bergen, and K. Hanna, "Representation of scenes from collections of images," in *Proc. IEEE Workshop on Representation of Visual Scenes*, pp. 10–17, 1995.

[2] S. Mann and R. Picard, "Virtual bellows: Constructing high-quality images from video," in *Proc. First Int. Conf. on Image Processing*, 1994.

[3] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, vol. 16, no. 2, pp. 22–30, 1996.

[4] S. E. Chen, "Quicktime VR — An image-based approach to virtual environment navigation," in *Proc. SIGGRAPH 95*, pp. 29–38, 1995.

[5] M. Levoy and P. Hanrahan, "Light field rendering," in *Proc. SIGGRAPH 96*, 1996.

[6] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proc. SIGGRAPH 96*, pp. 43–54, 1996.
[7] D. N. Wood, A. Finkelstein, J. F. Hughes, C. E. Thayer, and D. H. Salesin, "Multiperspective panoramas for cel animation," in *Proc. SIGGRAPH 97*, pp. 243–250, 1997.
[8] S. Peleg and J. Herman, "Panoramic mosaics by manifold projection," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 338–343, 1997.
[9] S. Baker and S. Nayar, "A theory of catadioptric image formation," in *Proc. Sixth Int. Conf. on Computer Vision*, pp. 35–42, 1998.
[10] L. McMillan and G. Bishop, "Plenoptic modeling," in *Proc. SIGGRAPH 95*, pp. 39–46, 1995.
[11] S. B. Kang and R. Szeliski, "3-D scene data recovery using omnidirectional multibaseline stereo," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 364–370, 1996.
[12] S. Peleg and M. Ben-Ezra, "Stereo panorama with a single camera," in *Proc. Computer Vision and Pattern Recognition Conf.*, pp. 395–401, 1999.
[13] H. C. Huang and Y. P. Hung, "Panoramic stereo imaging system with automatic disparity warping and seaming," *Graphical Models and Image Processing*, vol. 60, no. 3, pp. 196–208, 1998.
[14] R. C. Bolles, H. H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. of Computer Vision*, vol. 1, no. 1, pp. 7–55, 1987.
[15] A. Katayama, K. Tanaka, T. Oshino, and H. Tamura, "A viewpoint dependent stereoscopic display using interpolation of multi-viewpoint images," in *Proc. SPIE Vol. 2409A*, pp. 21–30, 1995.
[16] V. Nalwa, "A true omnidirectional viewer," tech. rep., Bell Labs, Holmodel, NJ, Feb 1996.
[17] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, no. 4, pp. 353–363, 1985.
[18] H.-Y. Shum and R. Szeliski, "Stereo reconstruction from multiperspective panoramas," in *Proc. Seventh Int. Conf. on Computer Vision*, (Kerkyra, Greece), September 1999.
[19] E. Krotkov, K. Henriksen, and R. Kories, "Stereo ranging with verging cameras," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, no. 12, pp. 1200–1205, 1990.
[20] S. Kang, J. Webb, L. Zitnick, and T. Kanade, "A multibaseline stereo system with active illumination and real-time image acquisition," in *Proc. Fifth Int. Conf. on Computer Vision*, pp. 88–93, 1995.

Appendix

Vergent stereo rigs have been shown to obtain better stereo reconstructions than parallel rigs for objects near the vergence point (e.g., [19, 20]). But why is maximizing vergence angle desirable? In this section we provide justification for maximizing the vergence angle that holds for an arbitrarily shaped connected region X of camera positions.

Suppose two beams intersect at points P and Q, defining two corners of a voxel, and that both P and Q lie outside the convex hull of X, as shown in Fig. 9. We claim that maximizing vergence angle approximately minimizes the area and perimeter among voxels with corners P and Q, and this approximation becomes perfect as the number of beams goes to infinity. While this does not prove that the omnivergent sensor is optimal for a finite set of beams, it demonstrates that it performs extremely well as the resolution increases.

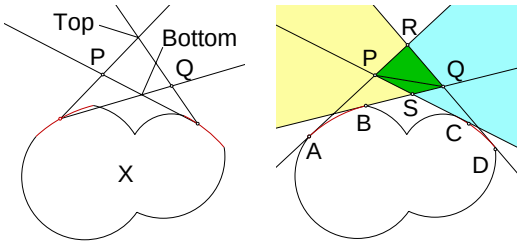


Figure 9: Maximum vergence gives the smallest voxel with corners P and Q.

We can choose many different pairs of beams from within X that result in a voxel with P and Q as corners, and each of these voxels actually has two different vergence angles: one at the top corner (R) and one at the bottom corner (S), as shown in the figure. Now if we decrease the distance between P and Q, i.e., by decreasing the width of the beams, the top and bottom vergence angles become closer, and, in the limit, are equal. Thus, if we chose to maximize the vergence angle between the two angle bisectors of the beams, we are approximately maximizing the top and bottom vergence angles, and this approximation gets better as the number of pixels increases.

Let the two tangents to X that pass through P intersect X at A and C. Similarly, let the two tangents to X that pass through Q intersect X at B and D. By construction, $PRQS$ is the intersection of all beams from points in X that contain both P and Q. Consequently, $PRQS$ must have smaller area and perimeter than any voxel containing P and Q. Note, however, that $PRQS$ is not the intersection of two pixel beams originating from within the circle. Also notice that $\angle PRQ$ is the maximum top vergence angle for any voxel, while $\angle PSQ$ is the maximum bottom vergence angle. So, we see that placing the beams' bases at A and D maximizes the top vergence angle, while placing them at B and D maximize the bottom vergence angle. These correspond to minimizing the areas of triangles PRQ and PSQ, respectively. As the number of pixels in increased, P and Q converge to the same point, as do A and B, so that in the limit, maximizing the vergence angle produces the smallest possible voxel.