

Classification and Characterization of Image Acquisition for 3D Scene Visualization and Reconstruction Applications

S ou- ang ei u-Fei uang and ein ard lette

Abstract

This paper discusses the techniques of image acquisition for 3D scene visualization and reconstruction applications 3DSV . The existing image acquisition approaches in 3DSV applications are briefly reviewed. There are still lacks of the studies about what principles are essential in the design and how we can characterize the limitations of an image acquisition model in a formal way. This paper addresses some of the main characteristics of existing image acquisition approaches summarized through a classification scheme and illustrated with many examples. The results of the classification lead to general characterizations in establishing the notions basic components for design analysis and assessment of image acquisition models. The notions introduced include: focal set receptor set reflector set etc. The definitions of the notions are given and supported with various examples e.g. concentric panoramas cataoptrical panoramas . The observations important issues and future directions from this study are also provided.

Classification and Characterization of Image Acquisition for 3D Scene Visualization and Reconstruction Applications

Shou-Kang Wei, Fay Huang, and Reinhard Klette

CITR, Computer Science Department, The University of Auckland
Tamaki Campus, Auckland, New Zealand

Abstract

This paper discusses the techniques of image acquisition for 3D scene visualization and reconstruction applications (3DSVR). The existing image acquisition approaches in 3DSVR applications are briefly reviewed. There are still lacks of the studies about what principles are essential in the design and how we can characterize the limitations of an image acquisition model in a formal way. This paper addresses some of the main characteristics of existing image acquisition approaches, summarized through a classification scheme and illustrated with many examples. The results of the classification lead to general characterizations in establishing the notions (basic components) for design, analysis and assessment of image acquisition models. The notions introduced include: focal set, receptor set, reflector set etc. The definitions of the notions are given and supported with various examples (e.g. concentric panoramas, cataoptrical panoramas). The observations, important issues, and future directions from this study are also provided.

1 Introduction

Image acquisition is a process for obtaining data from real 3D scenes. The role of the image acquisition process has critical impacts on subsequent processes in 3D scene visualization and reconstruction (3DSVR) applications¹. An *image acquisition model* defines image-acquiring components and their usages in the image acquisition process for a particular application. The

¹For instance, the applications using panoramic images include stereoscopic visualization [HH98, PBE99, WHK99a], stereo reconstruction [IYT92, Mur95, KS97, WHK99b, SS99], image-based rendering [Che95, MB95, KD97, RB98], localization, route planning or obstacle detection in robot-navigation [IYT92, ZT92].

specifications of image acquisition models typically differ between different applications except of some basic characterizations which will be discussed later. A scenario for developing an image acquisition model should cover the following steps: (1) list the requirements and specifications of the application under investigation; (2) sketch the problem(s) and available solution(s) or approaches; (3) design an image acquisition model (involving pose planning, sensor design, illumination conditioning etc) which may lead to solutions and satisfaction of practical constraints; (4) implement and test the image acquisition model.

Conceptually, the closer the relation between the data acquired and the outcome expected in a 3DSVR application, the simpler the processes involved and the better the performance. QuicktimeVR [Che95] serves as a good example. However in reality there are some physical constraints (such as temporal and spatial factors) and practical issues (e.g. cost, availability) that complicate the design and the realization of an image acquisition model for a 3DSVR application. Therefore it is important to study the constraints and issues as well as how they influence the design of image acquisition models.

Since different image acquisition models result into different subsequent processes providing different characteristics in respect to both geometrical and photometrical analysis, it is very risky developing 3DSVR applications without serious considerations of the suitability of the image data acquired for use in the intended application. Failures to assess the data may not only cause an unnecessary complexity to the subsequent processes but even lead to an inability to fulfill the requirements of the application.

Some researchers see a need for designing new image acquisition system(s) especially for 3DSVR applications. The point of view is that the traditional image acquisition models may/should not be able to serve all kinds of tasks in 3DSVR applications. Researchers in the image-based rendering community have also noticed this need/inadequacy. They reconfigured some components from the traditional image acquisition models (e.g. a pinhole projection model with a pre-defined camera motion) and received some interesting results (i.e. novel views generation without 3D reconstructions) [LH96, GGSC96, WFH⁺97, RB98, SH99]. However there are still lacks of the studies about what principles are essential in the design and how we can characterize the limitations of an image acquisition model in a formal way.

For being able to design, analyze and assess an image acquisition model, we need to establish the building-blocks (basic components) which construct the architecture of image acquisition models. In the next section, the classification of recent image acquisition approaches for 3DSVR applications is presented. The results of the classification characterize the existing image acquisition approaches and lead to general characterizations in Section 3 establishing basic/general components/notions for design, analysis and assessment of image acquisition models. The observations, important issues, and future directions from this study are addressed in the conclusion.

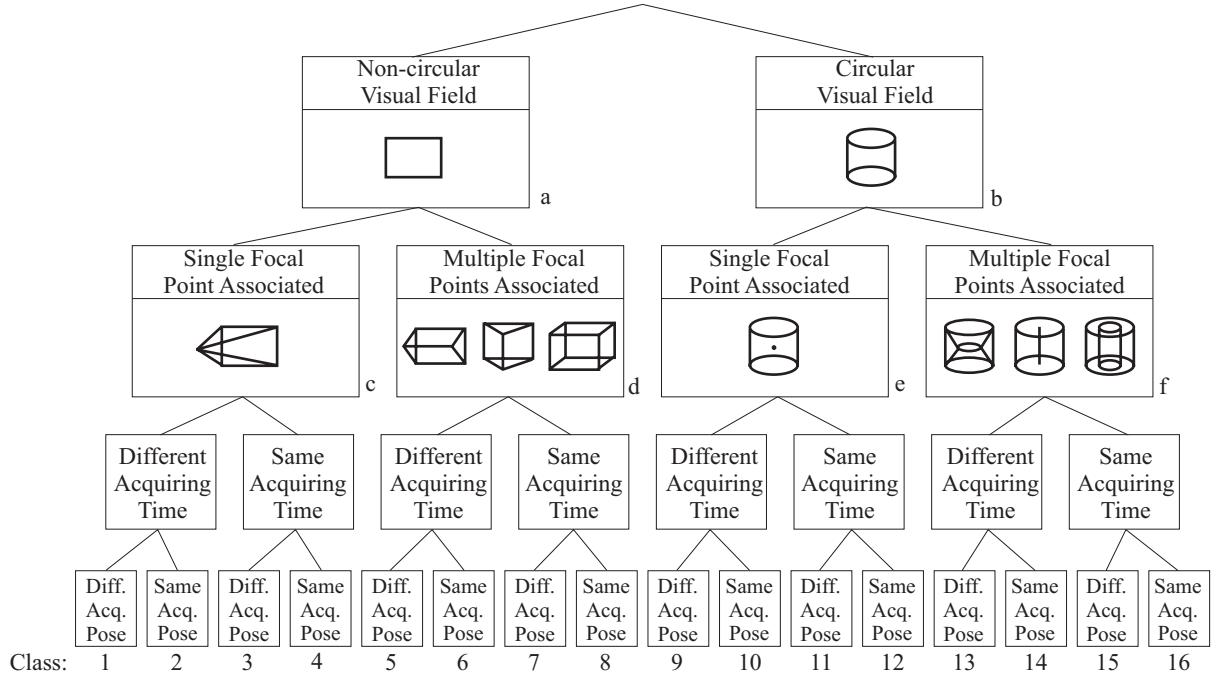


Figure 1: A classification tree of image acquisition approaches for 3D scene visualization and reconstruction applications. See text for details.

2 Classification

For simplicity, this paper mainly focuses on passive imaging² and leaves the factors introduced by active imaging to be incorporated later. To avoid the problems where the resulting classification becomes untraceable and lost in excessive detail, four binary classifiers are used. They are defined as follows:

Visual field: Circular/Non-circular

Focal point(s) associated with each image: Single/Multiple

Acquiring time: Different/Same

Acquiring pose: Different/Same

In Fig.1 the examples of the intermediate classes from the first two classifiers are depicted. A planar image and a cylindrical image (Fig.1 a and b) are examples for the non-circular and the circular visual field classes. A pinhole projection model (Fig.1 c) is the example for the class of a planar image associated with a single focal point. For the class of a planar image associated with multiple focal points the three examples (Fig.1 d) are: each image column associates with a focal point (left); each image row associates with a focal point (middle); and each image pixel associates

²The terms, *active imaging* and *passive imaging*, are frequently used to distinguish whether or not equipment (such as lighting device or laser) is acting on the physical scene while carrying out image acquisition [KSK98].

with a focal point (i.e. orthographic projection) (right). A central-projection panorama (Fig.1 e) is the example for the class of a cylindrical image associated with a single focal point. For the class of a cylindrical image associated with multiple focal points the three examples (Fig.1 f) are: each image column associates with a focal point (left); each image row associates with a focal point (middle); and each image pixel associates with a focal point (right).

Note that the temporal classifier may be used in a flexible way, that is, the acquiring time can be conceptually rather than physically the same. For example, a binocular stereo pair, used in stereo matching, acquired in a static scene under an almost constant illumination condition can be regarded conceptually as having the same acquiring time. For the acquiring pose classifier, two other characteristics, translation and rotation, may be used to further specify the possible classes.

Now let us look at some examples of the existing image acquisition approaches in the different classes. A typical example for class 1 is a video sequence. A video surveillance system is an example for class 2 because it fixes the image acquisition system at a particular place. In class 3, the binocular stereo pair acquisition is a typical example. More recently approaches such as light field [LH96] and Lumigraph [GGSC96] also fall into class 3, where both approaches arrange the poses of a pinhole camera to a planar grid layout and assume a constant acquiring time for their applications. Re-sampling the acquired image data, which are parameterized into a 4D function, generates a novel view for use in their visualization applications. Other examples in class 3 are [LF94, AS97, SD97, MMB97, SGHS98, KS99]. In class 4, single still image capturing, currently the most common image acquisition model, serves as an example.

In class 5, the image acquisition model used in three light-sources photometric stereo method [KSK98] for 3D reconstruction is one of the examples, where the orthographic projection (non-circular visual field and multiple focal centers); with only one of lights on for each image capturing (different acquiring time); and multiple views (different acquiring poses) for a full 3D reconstruction are used. Another example in class 5 is reported in [RB98]. Considering 2.5D reconstruction of a 3D scene from a single viewing pose, a well-known passive approach, depth from defocusing, is usually adopted. An example of such an approach which falls into class 6 is the multi-focus camera with a coded aperture proposed in [HM98].

For class 7, an example is a three-line scanner system (pushbroom camera [GH97] on an airplane) used in the terrain reconstruction or heat-spot sensing [SE96, RS97]. This setup is characterized by (1) non-circular visual field; (2) each image column associates with a focal point, (3) acquiring time is conceptually the same as performing the stereo matching; and (4) the poses of each line-scanner are inherently different.

Now we look at some examples of the approaches with circular visual field. In class 9, there are quite a few systems already developed in research institutes or commercialized in industry [ZWEa]. One of the applications for this so-called *dynamic-panorama-video* system is to allow the user to visualize a real scene by virtually walking along a path (where data is acquired) and

looking around 360° from any point in the path³. For instance, to visualize the interior of a building (i.e. walking through corridor, lobby, rooms, etc), the path is planed and a robot (on which a dynamic-panorama-video system is installed) implements the image acquisition. An interactive explorer⁴ is developed, providing an interface for the user to explore the interior of a building. Some image-based rendering techniques can be used to interpolate the missing data (i.e. gaps/holes in synthesized image) or extrapolate the acquired data to some extent such that the viewing space can be expanded to a certain degree.

The applications (e.g. environmental study; surveillance, etc) requiring a panoramic-video system to be fixed at one position for a long period of time are members of class 10 (i.e. *static-panorama-video* system). For example, a system was deployed on the bank of a lake (West-lake, China) for monitoring environmental change and acquiring 360° panorama continuously for the whole year in 1998. Some examples of the surveillance application can be found in [ZWeb]. A well-known example for a single-focal-point panorama is QuickTimeVR [Che95] from Apple Inc., which falls into class 12. Using multiple single-focal-point panoramas to reconstruct a 3D scene, class 11, S.B. Kang and R. Szeliski reported their results in [KS97]. Other similar examples are [MB95, HH98, MW98]. The families of cataoptrical panorama used for 3D scene reconstruction [GNT98, OHS99, Svo99, BN99] mostly belong to class 11, except for the configuration: pinhole projection model with a spherical mirror, which is of class 15.

H. Ishiguro et al. first proposed an image acquisition model that is able to produce multiple panoramas by a single swiveling of a pinhole-projection camera, where each panorama is associated with multiple focal points. It is of class 15. The model was created for the 3D reconstruction of an indoor environment. Their approach reported in 1992 in [IYT92] already details essential features of the image acquisition model. The modifications and extensions of their model have been discussed by other works such as [PBE99, SH99, SKS99, SS99, HP00]. Other examples of this class are [WHK99a, PPBE00].

3 Characterization

In this section we discuss some general notions which characterize the basic components that are essential to most of the image acquisition models for 3DSVR applications. The notions include *focal set*, *receptor set*, *projection-ray set*, *reflector set*, *reflected-ray set*, temporal and spatial factors.

A focal set is a non-empty set of focal points in 3D space. A receptor set is a non-empty set of receptors (photon-sensing elements) in 3D space. In some cases, it is convenient to express

³The density of view points on a path depends on the frame rate of the video camera and the speed of the movement during the image acquisition.

⁴A software that synthesizes a view (image) via resampling the acquired data according to current user's viewing condition.

the points in a point-set by a geometric primitive such as a straight line, curve, plane, quadratic surface etc where all of the points lies on. For instance, the pinhole projection model contains one focal point and a set of coplanar receptors. The orthographic projection model contains a set of coplanar focal points and a set of coplanar receptors. The single-focal-point panorama (e.g. QuickTimeVR) contains one focal point and a set of receptors laid on a cylinder/sphere while the multiple-focal-points panorama contains a set of focal points on various geometrical forms (such as vertical straight line, circle, disk, cylinder etc, see Fig.1 f) and a set of receptors on a cylinder/sphere.

The association between focal points and receptors determines a particular subset of light rays being captured in 3D space and the geometrical relations between light rays. A complete bipartite set of focal and receptor sets is defined as $\{(p, q) : p \in \text{focal set and } q \in \text{receptor set}\}$. A projection-ray set is a non-empty subset of the complete bipartite set, which satisfies the following conditions:

1. $(p, q) \in \text{projection ray set}$ *iff* p and q satisfy pre-defined association rule(s);
2. $\forall p \in \text{focal set}$, there is at least a $q \in \text{receptor set}$ such that $(p, q) \in \text{projection-ray set}$;
3. $\forall q \in \text{receptor set}$, there is at least a $p \in \text{focal set}$ such that $(p, q) \in \text{projection-ray set}$.

For instance, the projection-ray set of a concentric panorama [HP00] can be characterized formally as follows. The focal points are an ordered finite sequence, p_1, p_2, \dots, p_n , which form a circle in 3D space. The set of receptors form 2D grid points on a cylinder which is co-axis to the circle of the focal points. The number of column of the grid must be equal to n . The association rules for p and q that determine whether (p, q) belongs to projection-ray set are as follows. (1) $\forall q \in \text{receptor set}$ which belongs to the same column must be paired up to an unique $p_i \in \text{focal set}$. (2) There is an ordered one-to-one mapping between each $p_i \in \text{focal set}$ and each column of the grid. In other words, the column of the grid in either counterclockwise or clockwise are indexed as c_1, c_2, \dots, c_n such that $\forall q \in c_i$ is mapped to $p_i, i \in [1..n]$.

A reflector set, e.g. mirror(s), is used to characterize how light rays can be captured indirectly by the system with respect to the actual scene. For instance, a hyperbolic mirror is used in conjunction with the pinhole projection model for acquiring a wide visual field of a scene (e.g. 360° panorama). Similarly, with the orthographic projection model, the parabolic mirror is adopted. Such type of image acquisition model allows all the reflected projection rays intersected in the focus of the hyperboloid [Svo99, BN99], which poses a simple computational model for use in 3DSVR applications.

A reflector set is a set of reflectors' surface equations, usually a set of first or second order continuous and differentiable surfaces in 3D space. Let $\mathcal{P}(\text{reflector set})$ denote the power set of the reflector set. Define a geometrical transformation T as follows:

$$T : ((p, q), S) \mapsto (p', q'),$$

where $(p, q) \in$ projection-ray set, $S \in \mathcal{P}(\text{reflector set})$, and $(p', q') \in$ reflected-ray set. The transformation T is a function which transforms a projection ray with respect to an element of $\mathcal{P}(\text{reflector set})$, and outputs the resulting reflected ray. A reflected-ray set consists of elements with the same properties as in the projection-ray set, and obeys the condition 2 and 3 of the projection-ray set defined before. When the transformations of a projection-ray set take place, a unique element of $\mathcal{P}(\text{reflector set})$ is used. Formally,

$$\text{reflected-ray set} = \{T((p, q), S) : (p, q) \in \text{projection-ray set}\},$$

where S is one particular element of the power set of a reflector set. In particular, when $\emptyset \in \mathcal{P}(\text{reflector set})$ is chosen, the resulting reflected-ray set is identical to the original projection-ray set. If the number of elements of the chosen S is more than one, the transformation can behave as ray-tracing.

A single projection-ray set (or reflected-ray set, we omit repeating this in the following text) is not enough for most 3DSVR applications. Two factors are used to characterize the relationships between multiple projection-ray sets. Temporal factor characterizes the acquiring-time difference between projection-ray sets. Spatial factor describes geometrical relations (e.g. orientation and translation) among the subsets of light rays captured in different poses. The multiple projection-ray sets can then be described by $\{J_{t,\rho}\}$ to distinguish the projection-ray sets taken at time t and pose ρ . Multiple images, a collection of the projection-ray sets acquired in different time or pose $\{J_{t,\rho}\}$, are a subset of a light field. All possible light rays in a specified 3D space and time intervals form a light field. Regardless time factor, to acquire a complete light field of a medium-to-large scale space is already known to be very difficult, or say, almost impossible to achieve based on the technology available to date. Usually, few sampled projection-ray sets are acquired for approximating a complete light field. Due to the nature of scene complexity, the selections of optimal projection-ray samples become important factor to determine the quality of the approximation of a complete light field of a 3D scene.

4 Conclusions

This paper discusses image acquisition approaches for 3D scene visualization and reconstruction applications. The importance of the role of image acquisition and the impacts to the subsequent processes in developing a 3DSVR application are addressed. It may be risky and inappropriate in both the research and the developments of 3DSVR applications if we do not consider a possibility that other/better image acquisition models might exist.

We designed and applied a classification scheme for the existing image acquisition approaches for 3DSVR applications. Some existing image acquisition approaches in 3DSVR applications are briefly reviewed. The results of the classification lead to general characterizations in establishing notions (basic components) for design, analysis and assessment of image acquisition models.

In future we will look further to the relationship between applications and image acquisitions. Given some conditions with respect to a particular 3DSVR application: (1) what is the capability and limitation of an image acquisition model; and (2) what criteria should be used to evaluate the developed image acquisition model in respect to its application?

An extension of this study could develop into a model that is able to automatically generate (optimal) solution(s) of image acquisition model satisfying the image acquisition requirements from a 3DSVR application. The success of the model has direct practical benefits for 3DSVR applications. With respect to a theoretical aspect⁵ it helps us to understand what image acquisition can support to a 3DSVR application (capability analysis) as well as how far the support may go (limitation analysis).

References

- [AS97] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *CVPR97*, pages 1034–1040, June 1997.
- [BN99] S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *IJCV*, 35(2):1–22, November 1999.
- [Che95] S. E. Chen. QuickTimeVR - an image-based approach to virtual environment navigation. In *Proc. SIGGRAPH'95*, pages 29–38, 1995.
- [GGSC96] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *Proc. SIGGRAPH'96*, pages 43–54, New Orleans, Louisiana, August 1996.
- [GH97] R. Gupta and R. I. Hartley. Linear pushbroom cameras. *PAMI*, 19(9):963–975, September 1997.
- [GNT98] J. Gluckman, S. K. Nayar, and K. J. Thoresz. Real-time omnidirectional and panoramic stereo. In *DARPA98*, pages 299–303, 1998.
- [HH98] H. C. Huang and Y. P. Hung. Panoramic stereo imaging system with automatic disparity warping and seaming. *GMIP*, 60(3):196–208, May 1998.
- [HM98] S. Hiura and T. Matsuyama. Multi-focus camera with coded aperture: Real-time depth measurement and its applications. In *Proc. of Workshop in Cooperative Distributed Vision*, pages 101–118, Nov. 1998.
- [HP00] F. Huang and T. Pajdla. Epipolar geometry in concentric panoramas. Research Report CTU–CMP–2000–07, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, March 2000.

⁵The study of the image acquisition should be carried out beyond the current technology for the theoretical merit.

- [IYT92] H. Ishiguro, M. Yamamoto, and S. Tsuji. Omni-directional stereo. *PAMI*, 14(2):257–262, February 1992.
- [KD97] S. B. Kang and P. K. Desikan. Virtual navigation of complex scenes using clusters of cylindrical panoramic images. Technical Report CRL 97/5, Digital Equipment Corporation, Cambridge Research Lab, September 1997.
- [KS97] S. B. Kang and R. Szeliski. 3-d scene data recovery using omnidirectional multibaseline stereo. *IJCV*, 25(2):167–183, November 1997.
- [KS99] K. N. Kutulakos and S. Seitz. A theory of shape by space carving. In *ICCV99*, pages 307–314, 1999.
- [KSK98] R. Klette, K. Schlüns, and A. Koschan. *Computer Vision - Three-Dimensional Data from Images*. Springer, Singapore, 1998.
- [LF94] S. Laveau and O.D. Faugeras. 3-d scene representation as a collection of images. In *ICPR94*, pages A:689–691, 1994.
- [LH96] M. Levoy and P. Hanrahan. Light field rendering. In *Proc. SIGGRAPH'96*, pages 31–42, New Orleans, Louisiana, August 1996.
- [MB95] L. McMillan and G. Bishop. Plenoptic modeling: An image-based rendering system. In *Proc. SIGGRAPH'95*, pages 39–46, 1995.
- [MMB97] W. R. Mark, L. McMillan, and G. Bishop. Post-rendering 3d warping. In *Proceedings of 1997 Symposium on Interactive 3D Graphics*, pages 7–16, Providence, RI, April 1997.
- [Mur95] D. W. Murray. Recovering range using virtual multicamera stereo. *CVIU*, 61:285–291, 1995.
- [MW98] T. Matsuyama and T. Wada. Cooperative distributed vision - dynamic integration of visual perception, action, and communication. In *Proc. of Workshop in Cooperative Distributed Vision*, pages 1–40, Nov. 1998.
- [OHS99] M. Ollis, H. Herman, and S. Singh. Analysis and design of panoramic stereo vision using equi-angular pixel cameras. Technical Report 04, The Robotics Institute, Carnegie Mellon University, Pittsburgh, USA, 1999.
- [PBE99] S. Peleg and M. Ben-Ezra. Stereo panorama with a single camera. In *CVPR99*, pages I:395–401, 1999.
- [PPBE00] S. Peleg, Y. Pritch, and M. Ben-Ezra. Cameras for stereo panoramic imaging. Technical report, Computer Vision Lab, Institute of Computer Science of The Hebrew University, 2000.

- [RB98] P. Rademacher and G. Bishop. Multiple-center-of-projection images. In *Proceedings of SIGGRAPH 99*, volume 32, pages 199–206, August 1998.
- [RS97] R. Reulke and M. Scheel. Ccd-line digital imager for photogrammetry in architecture. In *Int.Archives of Photogrammetry and Remote Sensing*, volume XXXII (5C1B), pages 195–201, 1997.
- [SD97] S. Seitz and C. Dyer. Photorealistic scene reconstruction by space coloring. In *CVPR97*, page A9: Shape, 1997.
- [SE96] R. Sandau and A. Eckardt. The stereo camera family waoss/waac for space-borne/airborne applications. In *Int.Archives of Photogrammetry and Remote Sensing*, volume XXXI(B1), pages 170–175, 1996.
- [SGHS98] J. W. Shade, S. J. Gortler, L.-W. He, and R. Szeliski. Layered depth images. In *Proc. SIGGRAPH'98*, volume 32, pages 231–242, August 1998.
- [SH99] H.-Y. Shum and L.-W. He. Rendering with concentric mosaics. In *Proc. SIGGRAPH'98*, pages 299–306, Los Angeles, California., August 1999.
- [SKS99] H. Shum, A. Kalai, and S. Seitz. Omnivergent stereo. In *ICCV99*, pages 22–29, 1999.
- [SS99] H. Shum and R. Szeliski. Stereo reconstruction from multiperspective panoramas. In *ICCV99*, pages 14–21, 1999.
- [Svo99] T. Svoboda. *Central Panoramic Cameras Design, Geometry, Egomotion*. PhD Thesis, Center for Machine Perception, Czech Technical University, Prague, Czech Republic, 1999.
- [WFH⁺97] Daniel N. Wood, Adam Finkelstein, John F. Hughes, Craig E. Thayer, and David H. Salesin. Multiperspective panoramas for cel animation. In Turner Whitted, editor, *SIGGRAPH 97 Conference Proceedings*, Annual Conference Series, pages 243–250. ACM SIGGRAPH, Addison Wesley, August 1997.
- [WHK99a] S. K. Wei, F. Huang, and R. Klette. Three-dimensional scene navigation through anaglyphic panorama visualization. In *CAIP99*, pages 542–549, 1999.
- [WHK99b] S. K. Wei, F. Huang, and R. Klette. Three dimensional view synthesis from multiple images. Technical Report 42, CITR, Auckland University, New Zealand, March 1999.
- [ZT92] J.Y. Zheng and S. Tsuji. Panoramic representation for route recognition by a mobile robot. *IJCV*, 9(1):55–76, October 1992.
- [ZWEa] <http://www.cis.upenn.edu/~kostas/omni.html>.
- [ZWEb] <http://www.eecs.lehigh.edu/~tboult/vsam/>.