An Overview of Passive Vision Techniques

Steven M. Seitz
The Robotics Institute
Carnegie Mellon University
seitz@cs.cmu.edu, http://www.cs.cmu.edu/∼seitz

This section surveys *passive* optical techniques for recovering scene shape and reflectance characteristics from images. The objective is to acquire images of a scene observed from different viewpoints and possibly different illuminations, and from these images to compute scene shape and reflectance at every surface point. Given estimates of shape and reflectance, we can fabricate duplicate 3D models and compute graphical models that can be rendered from new viewpoints and orientations (see SIGGRAPH 99 Course #39 for more on the latter topic).

We use the term passive sensing to refer to the sensing of visible radiation that is already present in the scene, in contrast to *active* techniques that modify scene radiance via projected light patterns. Although active sensing can facilitate the computation of scene structure, as discussed in sections 6-9 of these course notes, active approaches are not always feasible, especially for modeling distant or fast-moving objects. Note that term *active* has also been used in the computer vision community to refer to methods in which the camera is controlled purposively to facilitate reconstruction and other tasks–this is not the meaning used in this course.

Over the years, researchers in human perception and computer vision have identified numerous cues for sensing shape and reflectance properties and a wide range of computation strategies have emerged for exploiting these cues for 3D inference. Practical vision algorithms have been developed to developed to measure shape from:

- Texture
- Shading
- Focus
- Parallax
- Motion
- Reflection
- Shadows
- Symmetry
- Interreflection
- Polarization

While the above list is by no means complete, it demonstrates a wealth of information from which 3D shape and reflectance can be computed. Nevertheless, scene reconstruction is an inverse problem and generally does not admit a unique solution, i.e., it is *ill-posed* [1]. Consequently, additional assumptions and heuristics are generally needed to make the problem tractible. Some prominent examples include:

- Ideal reflectance – surfaces in the scene are often assumed to satisfy ideal reflectance models. For instance, stereo and shape-from-shading techniques generally assume a perfect Lambertian (isotropic) reflection model with no transparency. Consequently these techniques perform poorly in the presence of specularities and other deviations from the model.

- Smoothness – imposing smoothness, or *regularization* functionals is a very common method for making ill-posed inverse problems well-posed [1]. Choosing the reconstruction that is *smoothest* yields a better-conditioned problem but has it's own pitfalls, for example the tendency to smooth over sharp edges or miss thin structures in the scene.

- Ideal projection – simplified projection models, like orthographic projection and ideal pinhole projection are used to make the reconstruction equations more tractable. Consequently, techniques that use these approximations pay a penalty in terms of accuracy and are not well-suited for applications that demand high-accuracy surface measurements.

This course will consider three prominent approaches to passive 3D Photography: stereo, structure from motion, shape from shading, and photometric stereo.

When an point is imaged from two different viewpoints, its image projection undergoes a displacement from its position in the first image to that in the second image. The amount of displacement, alternatively called *parallax* or *optical flow* is inversely proportional to distance and may therefore be used to compute 3D geometry. Given a correspondence between points from two known viewpoints, it is possible to compute depth by triangulation – intersect the rays from each optical center through the point's projection on the image plane. The problem of establishing correspondence is a fundamental difficulty and is the subject of a large literature on stereo vision. One prominent approach is to correlate pixels of similar intensities in two images, using an assumption that each scene point reflects the same intensity of light in the two views. The included paper by Okutomi and Kanade extends this correlation approach two three or more images and demonstrates that using several cameras at different camera separations, or *baselines* yields a significant improvement in reconstruction accuracy.

The accuracy of stereo techniques depends critically on having precise knowledge of camera position, orientation, and internal parameters, i.e., focal length, aspect ratio, principle point, and distortion contributed by non-ideal lens optics. Therefore, calibrating these parameters is a key part of the success of any stereo system. The first paper included in this section, by Tsai [2], describes one of the mots widely used algorithms for calibrating cameras. Tsai's approach images a calibration object with known geometry and solving for camera parameters and a radial distortion coefficient given the image projections of several points on the calibration object.

Rather than image a scene with two or more cameras, an alternative approach is to acquire images from a single moving camera and to reconstruct a 3D model from the resulting video sequence. An additional challenge in this approach is that the camera path must be estimated since precise calibration of freely moving cameras is extremely difficult (although see [3]). This problem of recovering scene geometry from the motion of points in the image plane of a moving camera is called *structure from motion*. As one of the classical problems in computer vision, there are numerous structure from motion algorithms. Included in this volume is one such technique developed by Tomasi and Kanade [4]. Their elegant *factorization* method assembles point measurements in a measurement matrix that is factored into the product of a *motion* and *shape* matrix using singular value decomposition. This approach is based on an orthographic projection model.

Shape from shading is one of the simplest problems to state and one of the most complicated to solve: given a single intensity image of a smooth curved object, how can the shape of the object be recovered [5]? This problem and its solution was pioneered in the vision community by Berthold Horn in his 1970 doctoral dissertation [6, 7]. Given the intensity of a point in the image and a known directional light source, Lambert's law (I = kNL, where I is the intensity, k the reflectance, N the unit normal, and L the light source direction) yields a one-parameter family of solutions for the surface normal. Additional constraints are needed to make the problem well-posed–it is generally solved by assuming similarity of surface reflectance and orientation at nearby points.

The difficulties with shape from shading may be mitigated by acquiring two or more images of the object under different illuminations. This is precisely the approach of photometric stereo, the subject of the third paper included in these course notes [8]. Each image provides one constraint on the normal, and therefore two images is sufficient to recover the normal up to a small number of possible solutions, and three images yields a unique solution for each image pixel. Photometric stereo enables relaxing the strong smoothness conditions imposed by classical shape from shading approaches, and therefore yields more reliable shape estimates.

References

[1] T. Poggio, V. Torre, and C. Koch, "Computational vision and regularization theory," *Nature*, vol. 317, pp. 314–319, 1985.

[2] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf cameras and lenses," *IEEE Trans. Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[3] M. Pollefeys, R. Koch, and L. V. Gool, "Self-calibration and metric reconstruction in spite of varying unknown internal camera parameters," in *Proc. Sixth Int. Conf. on Computer Vision*, pp. 90–91, 1998.

[4] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *Int. J. of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[5] L. B. Wolff, S. A. Shafer, and G. E. Healey, eds., *Shape Recovery*. Physics-Based Vision: Principles and Practice, Boston, MA: Jones and Bartlett Publishers, 1992.

[6] B. K. P. Horn, *Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View*. PhD thesis, Massachusetts Institute Of Technology, Cambridge, MA, 1970.

[7] B. K. P. Horn and M. Brooks, *Shape from Shading*. Cambridge, MA: MIT Press, 1989.

[8] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Journal of Optical Engineering*, vol. 19, no. 1, pp. 138–144, 1980.