Chapter 7

Model-Based Stereo

7.1 Motivation

The modeling system described in Chapter 5 allows the user to create a basic model of a scene, but in general the scene will have additional geometric detail (such as friezes and brickwork) not captured in the model. This chapter presents a new method of recovering such additional geometric detail automatically through stereo correspondence, called *model-based* stereo. Model-based stereo differs from traditional stereo in that it measures how the actual scene deviates from the approximate model, rather than trying to measure the structure of the scene without any prior information. The model serves to place the images into a common frame of reference that makes the stereo correspondence possible even for images taken from relatively far apart. The stereo correspondence information can then be used to render novel views of the scene using image-based rendering techniques.

7.2 Differences from traditional stereo

As in traditional stereo, given two images (which we call the *key* and *offset*), model-based stereo computes the associated depth map for the key image by determining corresponding points in the key and offset images. Like many stereo algorithms, our method is *correlation-based*, in that it attempts to determine the corresponding point in the offset image by comparing small pixel neighborhoods around the points. Because of this, correlation-based stereo algorithms generally require the neighborhood of each point in the key image to resemble the neighborhood of its corresponding point in the offset image.

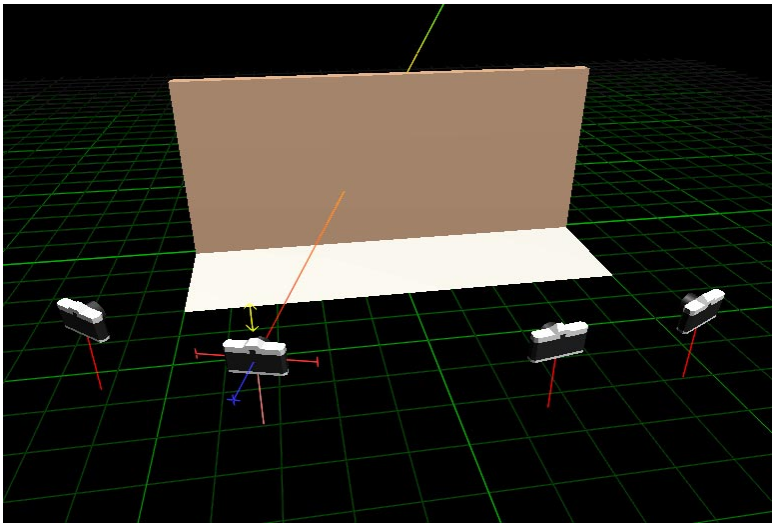


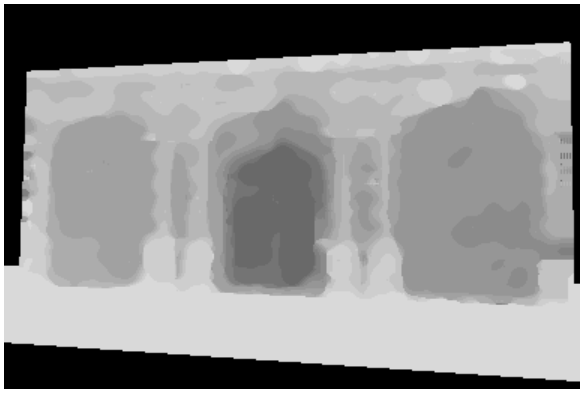
Figure 7.1: The façade of the Peterhouse chapel was photographed from four camera locations and modeled as two flat surfaces: one for the façade itself, and one for a piece of the ground in front of the chapel.

The problem is that when the key and offset images are taken from relatively far apart, as is the case for our modeling method, corresponding pixel neighborhoods can be foreshortened very differently. In Figs. 7.2(a) and (c), pixel neighborhoods toward the right of the key image are foreshortened horizontally by nearly a factor of four in the offset image.

(a) *Key Image*

(b) *Warped Offset Image*

(c) *Offset Image*



(d) *Computed Disparity Map*

Figure 7.2: **(a)** and **(c)** Two images of the entrance to Peterhouse chapel in Cambridge, England. The Façade program was used to model the façade and ground as a flat surfaces (see Fig. 7.1) and to recover the relative camera positions. **(b)** The warped offset image, produced by projecting the offset image onto the approximate model and viewing it from the position of the key camera. This projection eliminates most of the disparity and foreshortening with respect to the key image, greatly simplifying stereo correspondence. **(d)** An unretouched disparity map produced by our model-based stereo algorithm.

The key observation in model-based stereo is that even though two images of the same scene may appear very different, they appear similar after being projected onto an approximate model of the scene. In particular, projecting the offset image onto the model and viewing it from the position of the key image produces what we call the *warped offset* image, which appears very similar to the key image. The geometrically detailed scene in Fig. 7.2 was modeled as two flat surfaces with the photogrammetric modeling program, which also computed the original camera positions (see Fig. 7.1). As expected, the warped offset image (Fig. 7.2(b)) now exhibits the same pattern of foreshortening as the key image.

In model-based stereo, pixel neighborhoods are compared between the key and warped offset images rather than the key and offset images. When a correspondence is found, it is simple to convert its disparity to the corresponding disparity between the key and offset images, from which the point's depth is easily calculated. Fig. 7.2(d) shows a disparity map computed for the key image in (a).

The reduction of differences in foreshortening is just one of several ways that the warped offset image simplifies stereo correspondence. Some other desirable properties of the warped offset image are:

- Any point in the scene which lies on the approximate model will have zero disparity between the key image and the warped offset image.
- Disparities between the key and warped offset images are easily converted to a depth map for the key image.
- Depth estimates are far less sensitive to noise in image measurements since images pairs with large baselines can be used.

- Places where the model occludes itself relative to the key image can be detected and indicated in the warped offset image.

- A linear epipolar geometry (Sec. 7.3) exists between the key and warped offset images, despite the warping. In fact, the epipolar lines of the warped offset image coincide with the epipolar lines of the key image.

7.3 Epipolar geometry in model-based stereo

In traditional stereo, the *epipolar constraint* (see [13]) is often used to constrain the search for corresponding points in the offset image to a linear search along an epipolar line. This reduction of the search space from two dimensions to one not only speeds up the algorithm, but it also greatly reduces the number of opportunities to select a false matches. This section shows that taking advantage of the epipolar constraint is no more difficult in the model-based stereo case, despite the fact that the offset image is a non-uniformly warped version of the original offset image.

Fig. 7.3 shows the epipolar geometry for model-based stereo. If we consider a point P in the scene, there is a unique *epipolar plane* which passes through P and the centers of the key and offset cameras. This epipolar plane intersects the key and offset image planes in *epipolar lines* e_k and e_o. If we consider the projection p_k of P onto the key image plane, the epipolar constraint states that the corresponding point in the offset image must lie somewhere along the offset image's epipolar line.

In model-based stereo, neighborhoods in the key image are compared to the warped offset image rather than the offset image. Thus, to make use of the epipolar constraint, it is necessary to determine where the pixels on the offset image's epipolar line project to in the warped offset image.

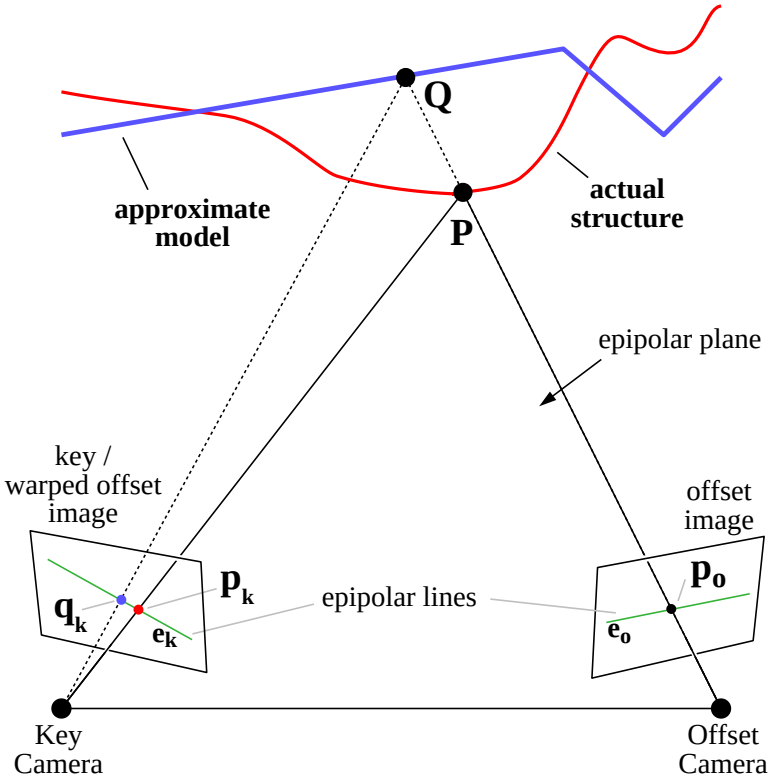


Figure 7.3: The epipolar geometry for model-based stereo. This figure illustrates the formation of the warped offset image, and shows that points which lie on the model exhibit no disparity between the key and warped offset images. Furthermore, it shows that the epipolar line in the warped offset image of a particular point in the key image is simply that point's epipolar line in the key image. The text of this chapter provides a more detailed explanation of these properties.

The warped offset image is formed by projecting the offset image onto the model, and then reprojecting the model onto the image plane of the key camera. Thus, the projection p_o of P in the offset image projects onto the model at Q, and then reprojects to q_k in the warped offset image. Since each of these projections occurs within the epipolar plane, any possible correspondence for p_k in the key image must lie on the *key* image's epipolar line in the warped offset image. In the case where the actual structure and the model coincide at P, p_o is projected to P and then reprojected to p_k, yielding a correspondence with zero disparity.

The fact that the epipolar geometry remains linear after the warping step also facilitates the use of the ordering constraint [3, 13] through a dynamic programming technique.

7.4 The matching algorithm

Once the warped offset image is formed, stereo matching proceeds in a straightforward manner between the key and warped offset images. The one complication is that the two images are not rectified in the sense of the epipolar lines being horizontal; instead, the epipolar lines which need to be searched along converge at a finite epipole. Since this epipole can be either within or outside of the borders of the key image, special care must be taken to ensure that the epipolar lines are visited in a reasonable fashion. The approach taken in this work is to traverse the pixels of the border of the key image in a clockwise manner, examining the corresponding epipolar line between the current border pixel and the epipole at each step.

The matching window we used was a 7×7 pixel neighborhood, and the matching function we used was the normalized correlation between the forty-nine pixel intensity values in the two regions. Normalized correlation makes a good stereo matching criterion because it is not sensitive

to overall changes in brightness and contrast between the two images.

The ordering contraint [3] was exploited using a dynamic programming technique. The ordering constraint is useful in avoiding stereo matching errors since it enforces that each piece of the model be a single, connected surface.

Lastly, the stereo disparity maps were post-processed using a nonlinear smoothing technique related to those used in [38, 6, 33]. Our process used a relaxation technique to smooth the disparity map in regions where there was little texture, while avoiding excessive smoothing near intensity edges. Importantly, this postprocessing step helps enforce smoothness in the stereo matching information among different epipolar lines.

7.5 Results

Fig 7.2 shows the results of running the model-based stereo algorithm on two images of the Peterhouse chapel façade. To achieve even better results, stereo was run on each of the four images from the camera positions shown in Fig. 7.1.

Once a depth map was computed for each image, the image can be rerendered from novel viewpoints using the image-based rendering methods described in [57, 44, 27, 37]. In this case, several images and their corresponding depth maps are available. This makes it possible to use the view-dependent texture-mapping method of Chapter 6 to composite multiple renderings of each of the images to produce far better results. Novel views of the chapel façade in Fig. 7.2 generated using both model-based stereo and view-dependent texture-mapping of four images are shown in Fig. 7.4.

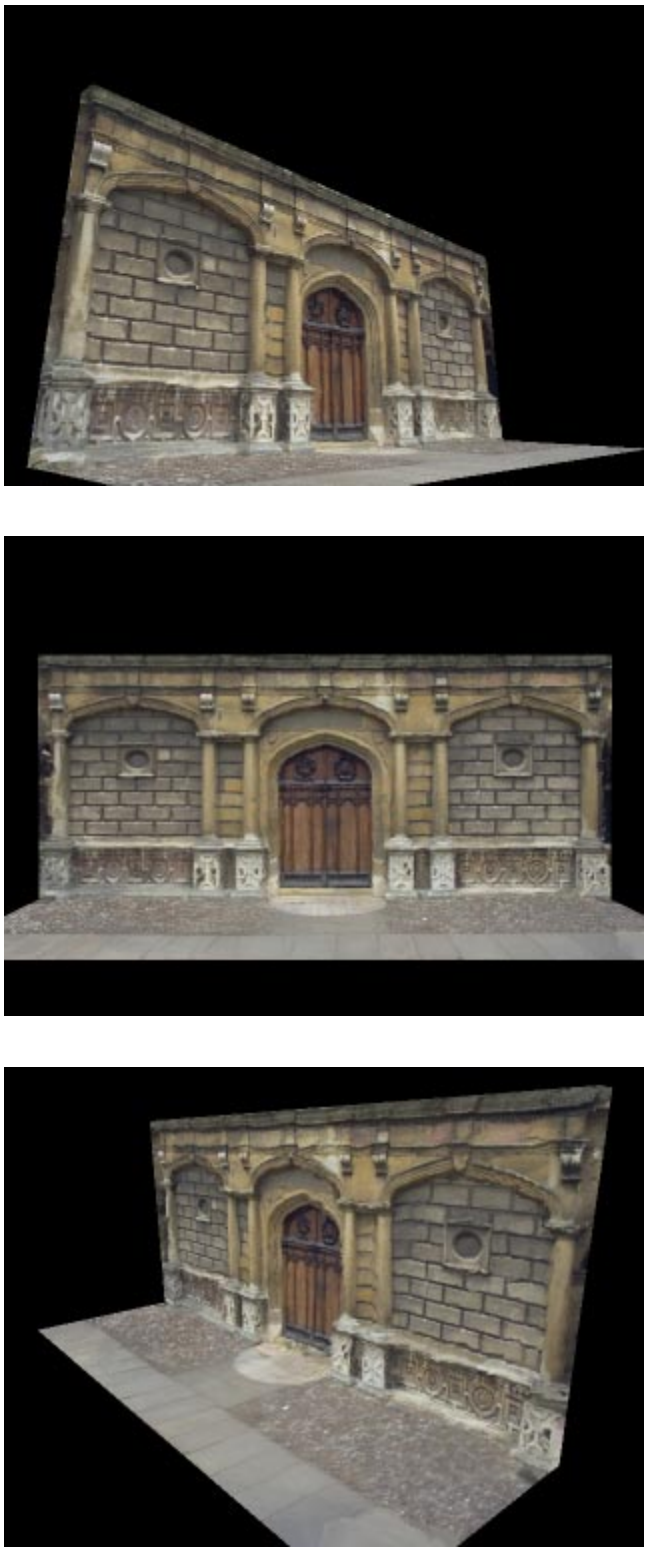


Figure 7.4: Novel views of the scene generated from four original photographs. These are frames from an animated movie in which the façade rotates continuously. The depth is computed from model-based stereo and the frames are made by compositing image-based renderings with view-dependent texture-mapping.