Light Field Reconstruction Using a Planar Patch Model

Adam Bowen, Andrew Mullins, Roland Wilson, and Nasir Rajpoot

Signal & Image Processing Group,
Department of Computer Science,
University of Warwick, Coventry, CV4 7AL, England
{fade, andy, rgw, nasir}@dcs.warwick.ac.uk

Abstract. Light fields are known for their potential in generating 3D reconstructions of a scene from novel viewpoints without need for a model of the scene. Reconstruction of novel views, however, often leads to ghosting artefacts, which can be relieved by correcting for the depth of objects within the scene using disparity compensation. Unfortunately, reconstructions from this disparity information suffer from a lack of information on the orientation and smoothness of the underlying surfaces. In this paper, we present a novel representation of the surfaces present in the scene using a planar patch approach. We then introduce a reconstruction algorithm designed to exploit this patch information to produce visually superior reconstructions at higher resolutions. Experimental results demonstrate the effectiveness of this reconstruction technique using high quality patch data when compared to traditional reconstruction methods.

1 Introduction

A Light Field [1] captures a large array of images of a scene in a representation that allows fast reconstruction from a sufficiently arbitrary location and preserves view dependent effects. The scene is represented as a number of camera viewpoints of a common imaging plane. The pixel samples then correspond to the intersections of a ray with the image plane and the camera plane. Traditional light field reconstruction algorithms exploit this efficient data structure to rapidly sample light rays for every pixel being reconstructed. Unfortunately, it is often impractical or even impossible to capture the camera plane at sufficient resolution to represent all the desired viewpoints, resulting in noticeable artefacts in the reconstructions. Attempts have been made to alleviate this problem using variable focus and aperture [2], compensation with a low resolution model [3] and image warping [4]. Other techniques for image based rendering can also be applied to light field data, such as space carving [5] and photo-consistency approaches [6].

In fact, there is significantly more information in a light field than is exploited by a traditional reconstruction approach. Traditional reconstruction does not take advantage of the fact that all the camera views are of the same object to

H. Kalviainen et al. (Eds.): SCIA 2005, LNCS 3540, pp. 85–94, 2005.

infer properties of the object. By examining the light field data we can obtain information about the object of interest that will allow us to improve our reconstructions. Typically, this is the approach taken in image warping [4]. Warping extracts disparity information from the available images to then warp them to the novel viewpoint. However, this introduces problems during reconstruction, most significantly dealing with multiple conflicting samples of the same pixel and filling 'holes' in the reconstructed image. These problems arise because disparity information between images is not sufficient to model the shape and orientation of the surfaces present in the scene and so occlusion boundaries cannot be properly reconstructed. Other methods for computing reconstructions from light fields include photo-consistency [6] and space carving [5]. Using a photo-consistency approach [6] for reconstruction is very slow as not much preprocessing can be performed, whilst using a space carving [5] approach discards the view-dependent information.

We present a novel representation of the surfaces present in the scene using planar patches, and an algorithm for the reconstruction of these patches when the patch estimates may be unreliable.

2 Multiresolution Surface Estimation

When estimating the disparity between two images, such as with a stereo image pair, we can easily obtain an estimate of the depth of each pixel. Light field data sets (irrespective of the actual representation) have significantly more viewpoints than a single stereo pair. A surface patch provides information on not just the depth of a surface but also the normal to that surface. Figure 1 shows how these patches can be represented for a given image block using three parameters.

It is possible to describe the general projective properties of a camera using the position $\mathbf{o}_i$ of the camera i, and the direction $\mathbf{r}_i(x, y)$ of a ray passing through pixel (x, y) on the camera's image plane. We compute per pixel disparity values between horizontally and vertically adjacent cameras using the multiresolution approach described in [7]. Let $\delta x_{i,j}$ and $\delta y_{i,j}$ respectively denote the horizontal and vertical disparity between two cameras i and j. The disparity value tells us that these two pixels correspond to the same point in 3D space, hence we can obtain an equation of the form

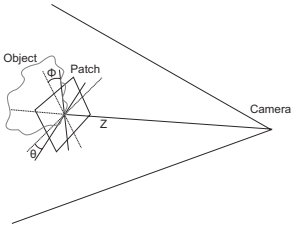


Fig. 1. The parameters for a surface patch, z, θ and ϕ

$$\mathbf{o}_i + z_i \mathbf{r}_i(x, y) = \mathbf{o}_j + z_j \mathbf{r}_j(x + \delta x_{i,j}, y + \delta y_{i,j}) \quad (1)$$

for each pair of disparity estimates and some scalars z_i and z_j. Because this is an over-constrained problem we apply a least squares solver to find a value for z_i (and not z_j because the point must lie along the ray from camera i but erroneous estimates may mean it does not lie along the ray from camera j). Given these depth values we can obtain a cloud of points in 3D space that map out the shape of the object by solving the set of equations given by the disparity maps for camera i and equation 1 and evaluating

$$\mathbf{o}_i + z_i \mathbf{r}_i(x, y) \quad (2)$$

for each pixel.

Once we have a cloud of points, it is possible to obtain patch parameters by first choosing the points that correspond to a pixel block in the source image and then fitting a plane through these points using principle component analysis. Larger blocks may not be fine enough to represent details, whilst smaller blocks are prone to error. To combat these problems we apply a multiresolution approach. We start at the coarsest resolution and attempt to fit a planar patch through the entire image. If the squared error between the point cloud and the patch is greater than a preset threshold value (for the properties of the teddy light field a value of 0.1 works well) then the patch is subdivided into four quadrants and we attempt to fit a new patch through the cloud of points found in each quadrant.

If the block used to generate a patch crosses an occlusion boundary the squared error will often be very high until the block size becomes very small. Once the block size approaches 2×2 it is often possible to fit a plane through any block in the image. However, a single plane does not model the two surfaces present at an occlusion boundary well. For this reason, we discard patches that cannot be represented using a 4×4 patch, and patches that become oblique to the camera (patches over 80 degrees) because they are very likely to be unreliable. We generated the patch data both for perfect disparity maps found from the scene geometry and estimated disparity maps in [7].

3 Reconstruction Algorithm

The estimation of planar patches, as described in section 2, takes place for every camera. The generated patches are locally consistent with the viewpoint from which they were estimated. If our patch data were perfect, this would be sufficient to construct a model of the object and recreate the novel view using traditional rendering techniques. However, the patch data is computed for each camera from disparity estimates and therefore is prone to error. Because these disparity estimates are only computed between pairs of cameras, we must also consider that patches for one camera may not be consistent with patches found for a camera some distance away. Our reconstruction algorithm takes account of these potential discrepancies by dividing the process into two stages. During

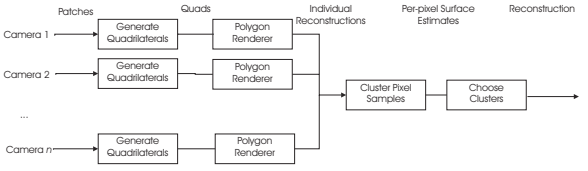


Fig. 2. Reconstruction Algorithm

the first stage a reconstruction is generated for every camera independently, using the patch data for that camera alone. The second stage then looks at the how consistent the data is across all the reconstructions to eliminate erroneous patches and select the best reconstruction. Figure 2 shows how the reconstruction algorithm proceeds.

3.1 Independent Reconstruction

Each patch is estimated using a block in the source camera's image. We generate an individual camera's estimate of the reconstruction by calculating a quadrilateral in 3D space that corresponds to the image block used to generate each patch, as illustrated by figure 1. Figure 3(a) shows the patches found from perfect disparity maps for one camera in the Teddy light field. The 'holes' seen in the image are regions that the camera cannot see, and so has no patch information for - most notably a 'shadow' of teddy is clearly visible on the background. Once the quadrilaterals have been computed, they are then textured and projected into the virtual viewpoint where a depth test is applied. Figure 3(b) shows the result of texturing and rendering the patches seen in figure 3(a) using standard OpenGL methods. We obtain an image similar to this for every available viewpoint. Only nearest neighbour interpolation is applied to the textures at this stage, to avoid blurring the textures during the second stage. This independent

(a) Shaded (b) Textured

Fig. 3. Surface patches estimated from scene geometry for a single camera

reconstruction stage can use graphics hardware to render the quadrilaterals as polygons and so is very fast.

3.2 Combining Reconstruction Images

Once each camera has generated an estimate of the reconstructed image, we attempt to identify the surfaces that are present at each reconstruction pixel using a clustering approach. For every pixel we wish to reconstruct, we have a colour sample and depth available from the estimate generated by each camera. Clustering these four dimensional vectors (red, green, blue and depth) gives us an estimate of the surfaces present in the reconstruction, and their corresponding depths.

To obtain these surface estimates, we apply a hierarchical clustering algorithm that finds the minimum number of clusters such that the total squared error within the cluster is below a threshold value. In our experiments we have found that, when the colour and depth values are between 0 and 1, a threshold values between 0.1 and 0.3 gave good clustering of the surfaces. The result is a variable number of clusters for each pixel that estimate the surfaces present along the ray. Small clusters may correspond to erroneous patches whilst larger clusters may correspond to genuine surfaces.

Given these clusters and their corresponding depths, we wish to select the cluster most likely to provide an accurate reconstruction. In other words, we wish to maximise the conditional probability

$$P(c_i|c_1, c_2...c_n) \tag{3}$$

for the selected cluster c_i and sample clusters $c_1, c_2...c_n$. Bayes' law gives us

$$P(c_i|c_1, c_2...c_n) = \frac{P(c_1, c_2...c_n|c_i).P(c_i)}{P(c_1, c_2...c_n)}. \tag{4}$$

Since $P(c_1, c_2...c_n)$ is constant across our maximisation, it can ignored. This simplifies the problem to maximising

$$P(c_1, c_2...c_n|c_i).P(c_i) \tag{5}$$

$P(c_i)$ is some measure of how reliable our cluster is. There are two factors to consider when calculating this measure. Firstly, we must consider the number of cameras that support the hypothesis that this cluster is a valid surface in our scene. Secondly, we must consider how much we trust the information provided by the supporting cameras. To achieve this we assign each camera j a weight w_j, the weight is computed as the dot (scaler) product of the direction of camera j and the direction of our reconstruction camera. If the direction of camera j is given by d_j and the direction of the reconstruction camera is d_{camera} then we find the weight as

$$w_j = \text{clamp}(0, (d_k.d_{\text{camera}})^\rho, 1) \tag{6}$$

where ρ is a tuning parameter used to control how closely aligned cameras must be before they are trusted and the clamp function clamps the value to the range

$[0, 1]$. Typically values of 5 to 8 cut out undesirable viewpoints. We say the probability of that cluster is

$$P(c_i) = \frac{\sum_{j \in c_i} w_j}{\sum_{k=1}^{C} w_k} \tag{7}$$

where $j \in c_i$ if camera j is in cluster c_i and C is the total number of cameras.

We now need to decide is how consistent the surfaces are with the selected surface - we say a surface is consistent with another surface if it occludes that surface, hence

$$P(c_1, c_2 ... c_n | c_i) = \frac{\sum_{j=1}^{n} \text{occludes}(c_i, c_j)}{n} \tag{8}$$

where

$$\text{occludes}(c_i, c_j) = \begin{cases} 1 & z_i \leq z_j, \\ 0 & \text{else.} \end{cases} \tag{9}$$

and z_i is the depth of the centroid of cluster c_i. Combining these two probabilities as in equation 5 gives us a measure of the quality of the surface represented by cluster c_i which we can then maximise for a value of c_i.

4 Results

We compared results of reconstruction using our patch model based rendering with three other techniques: traditional reconstruction, warping [4], and photo-consistency based reconstruction [6]. In order to assess the quality of different reconstructions, we computed the peak-signal-to-noise-ratio (PSNR) of the reconstructed images for all viewpoints as compared to the ground truth reconstruction. The reconstruction PSNR and time complexity for all the algorithms are summarised in Table 1, where N is the number of pixels, C is the number of cameras, and D is the number of depth samples (for the photo-consistency approach). In case of the photo-consistency reconstruction, we maximised the photo-consistency metric described in [7]. For warping and patch based reconstructions, we used disparity maps from scene geometry and estimations using [7]. Whilst the PSNRs are comparable, the patch based algorithm produces noticeably sharper and higher quality reconstructions.

Table 1. Summary of Results

Reconstruction Algorithm	PSNR	Time Complexity
Traditional Reconstruction	24.5dB	$O(N)$
Warping (perfect disparity maps)	31.5dB	$O(N)$
Warping (estimated disparity maps)	28.4dB	$O(N)$
Photo-consistency	27.0dB	$O(N.D.C^2)$
Patch Rendering (from geometry)	32.0dB	$O(N.C^2)$
Patch Rendering (from estimates)	26.0dB	$O(N.C^2)$

(a) Ground Truth

(b) Traditional Reconstruction

(c) Photo-consistency

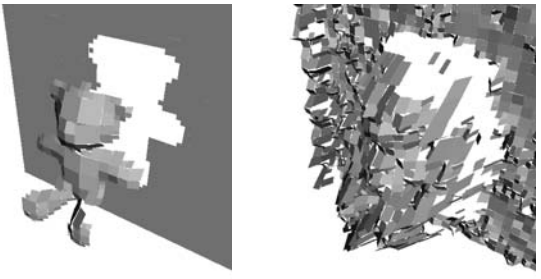
(d) Warping (Perfect Maps)

(e) Warping (Estimated Maps)

(f) Patch Rendering (Perfect Maps)

(g) Patch Rendering (Estimated Maps)

Fig. 4. Reconstruction Results for a Camera



(a) Perfect (b) Estimated

Fig. 5. Shaded images of some of the patches used for our reconstructions

Figure 4(b) shows the ghosting and blurring artefacts that typically result from a light field reconstruction when the camera plane is heavily under-sampled. Figure 4(c) shows the occlusion problems found with photo- consistency approaches. The photo-consistency technique performs well in unoccluded regions, but poorly in the occluded ones. Figures 4(d) and 4(e) alleviate the problems with the traditional reconstruction approach by realigning the images used in the reconstruction using disparity information. Reconstruction from perfect disparity maps suffers from hole filling problems due to occlusion between the legs and under the arm. This is because the warping approach only considers at most the 4 closest cameras for each pixel and in this case none of the cameras can see the desired region. It also suffers problems across the front of the legs. Because it has no model of how smooth or disjoint the surface is it cannot correctly interpolate nearby samples that belong to the same surface, the result is that parts of the background 'show through' the legs when no sample on the leg warps to the pixel. These problems are not visible when using the estimated maps because the

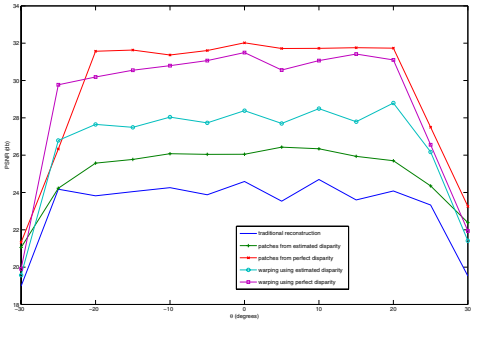


Fig. 6. Reconstruction quality (PSNR in dB) as we pan horizontally across the light field

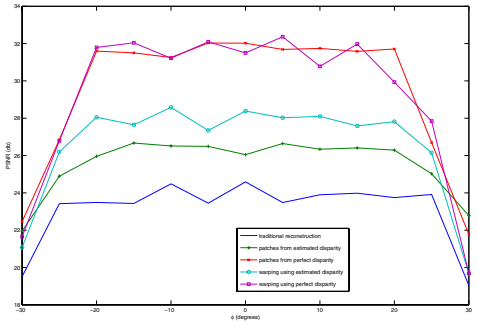


Fig. 7. Reconstruction quality (PSNR in dB) as we pan vertically across the light field

error in the maps prevents the samples from aligning. However, the lack of accuracy shows through when the samples from contributing cameras are blended together. Blending samples that do not come from the same surface results in a loss of detail in the image and often undesirable blurring or ghosting in the reconstruction. Figure 4(f) shows the reconstruction using perfect patches. This reconstruction is visually significantly superior to the other methods shown due to the accurate recovery of the edges. Because these reconstructions are generated at twice the resolution of the original light field, the technique is effectively achieving super-resolution on the reconstruction - making it more suitable for reconstructing scenes at different resolutions and from closer camera positions. The notable artefacts occur where part of the ear has been lost due to few cameras providing a reliable patch and a number of single pixel errors which could easily be restored using a prior based refinement of the reconstruction. Figure 4(g) shows the reconstruction from estimated patches. Whilst the technique performs well within teddy and on the background, it has significant problems with the edges. This is caused by the poor quality of the disparity values around the edges generating noisy patches from which the reconstruction algorithm cannot recover. Figure 5 compares some of the patch estimates with the perfect estimates, illustrating the problems our algorithm has reconstructing from the underlying data. Figure 6 shows how the reconstruction PSNR varies as we pan horizontally around the light field and figure 7 as we pan vertically around the light field. The peaks correspond to regions where the viewpoint aligns more closely with the source viewpoints. There are significant drops in PSNR towards the extreme angles because the arrangement is such that no camera can see some of the background needed to create the reconstruction.

5 Conclusions

We have presented a novel method of representing and reconstructing from light field data sets. The traditional and warping reconstruction approaches are computationally efficient, but do not exploit all the information that can be extracted

from the data set to produce the highest quality reconstructions. Instead they rely on a high volume of data to create accurate and high quality reconstructions - which is not ideal when it comes to the coding and transmission of light field data sets. Although our method is more computationally demanding, it is still relatively simple in terms of the approach and the scalability to higher resolutions. It provides more information on the structure of a scene whilst retaining the view-dependent properties of the surfaces in the scene. We can also generate visually superior reconstructions utilising the inherent super-resolution information available in light field data sets. While our algorithm is designed to be robust to erroneous data from a fraction of the input cameras, unfortunately it does not perform well when the patch data is extremely noisy. This leads us to believe that superior methods of estimating patch data are required, we are currently working on estimating patch properties directly from the light field data sets.

Acknowledgements

This research is funded by EPSRC project 'Virtual Eyes', grant number GR/S97934/01.

References

1. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of ACM Siggraph '96, New Orleans, LA, August 1996, ACM Press, New York (1996) 31–42
2. Isaksen, A., McMillan, L., Gortler, S.J.: Dynamically reparameterized light fields. In Akeley, K., ed.: Proceedings of ACM Siggraph 2000, New Orleans, Louisiana, July 2000, ACM Press, New York (2000) 297–306
3. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. In: Proceedings of ACM Siggraph '96, New Orleans, LA, August 1996, ACM Press, New York (1996) 43–54
4. Schirmacher, H.: Warping techniques for light fields. In: Proc. Grafiktag 2000, Berlin, Germany, September 2000. (2000)
5. Matusik, W.: Image-based visual hulls. Master of science in computer science and engineering, Massachusetts Institute of Technology (2001)
6. Fitzgibbon, A., Wexler, Y., Zisserman, A.: Image-based rendering using image-based priors. In: Ninth IEEE International Conference on Computer Vision, Nice, France, October 2003. Volume 2. (2003) 1176–1183
7. Bowen, A., Mullins, A., Rajpoot, N., Wilson, R.: Photo-consistency and multiresolution based methods for light field disparity estimation. In: Proc. VIE 2005, Glasgow, Scotland, April 2005. (2005)