

View Synthesis with Occlusion Reasoning Using Quasi-Sparse Feature Correspondences

David Jelinek and Camillo J. Taylor

GRASP Laboratory, CIS Department
University of Pennsylvania
3401 Walnut Street, Rm 335C
Philadelphia, PA, 19104-6229
{davidj2,cjttaylor}@grasp.cis.upenn.edu
Phone: (215) 898 0376
Fax: (215) 573 2048

Abstract. The goal of most image based rendering systems can be stated as follows: given a set of pictures taken from various vantage points, synthesize the image that would be obtained from a novel viewpoint. In this paper we present a novel approach to view synthesis which hinges on the observation that human viewers tend to be quite sensitive to the motion of features in the image corresponding to intensity discontinuities or edges. Our system focuses its efforts on recovering the 3D position of these features so that their motions can be synthesized correctly. In the current implementation these feature points are recovered from image sequences by employing the epipolar plane image (EPI) analysis techniques proposed by Bolles, Baker, and Marimont. The output of this procedure resembles the output of an edge extraction system where the edgels are augmented with accurate depth information. This method has the advantage of producing accurate depth estimates for most of the salient features in the scene including those corresponding to occluding contours. We will demonstrate that it is possible to produce compelling novel views based on this information.

The paper will also describe a principled approach to reasoning about the 3D structure of the scene based on the quasi-sparse features returned by the EPI analysis. This analysis allows us to correctly reproduce occlusion and disocclusion effects in the synthetic views without requiring dense correspondences. Importantly, the technique could also be used to analyze and refine the 3-D results returned by range finders, stereo systems or structure from motion algorithms. Results obtained by applying the proposed techniques to actual image data sets are presented.

Keywords: Structure From Motion, Surface Geometry, Image Based Rendering

1 Introduction

The goal of most image based rendering systems can be stated as follows: given a set of pictures taken from various vantage points, synthesize the image that would be obtained from a novel viewpoint. In this paper we present a novel approach to view synthesis which hinges on the observation that human viewers tend to be quite sensitive to the motion of features in the image corresponding to intensity discontinuities or edges. These discontinuities may arise from a number of sources including albedo changes, cast shadows and depth discontinuities in the scene. Our system focuses its efforts on recovering the 3D position of these features so that their motions can be synthesized correctly. In the current implementation these feature points are recovered from image sequences by employing the epipolar plane image (EPI) analysis techniques proposed by Bolles, Baker, and Marimont [1]. The output of this procedure resembles the output of an edge extraction system where the edgels are augmented with accurate depth information. In the sequel we will demonstrate that it is possible to produce compelling novel views based on this information.

Section 3 will describe a principled approach to reasoning about the 3D structure of the scene based on the quasi-sparse feature set produced by the EPI analysis. This analysis proceeds by considering the freespace volumes defined by the depth maps associated with each viewpoint. This analysis provides us with an implicit description of the 3D structure of the scene and allows us to correctly reproduce occlusions and disocclusions in the synthetic views. Importantly, this analysis could also be applied to the 3D results returned by range finders, stereo systems and structure from motion algorithms.

For the purposes of this discussion, previous approaches to the image based rendering problem can be divided into three categories. The first set of approaches are based on the plenoptic sampling approach described by Levoy and Hanrahan [9] and Gortler et al [5]. In these schemes, novel views are reproduced by sampling the appropriate rays from the input images. Shum and He [19] describe an interesting and effective approach for extending these techniques to immersive environments using a sampling system based on concentric mosaics. The method proposed in this paper differs from the techniques in this category by attacking the view generation problem as a morphing task rather than a plenoptic sampling problem.

The second category of approaches consists of techniques which proceed by constructing a detailed geometric model of the scene in the form of per pixel depth information for every image in the data set. Laveau and Faugeras [8], Pollefeys et al [13], Kanade et al [7] and Werner et al [23,6] all propose stereo based techniques for recovering the required depth or disparity maps from the input image data. Other researchers assume that the depth maps can be obtained from auxiliary range sensors [15,12]. Once these depth maps have been obtained, it is a relatively straightforward matter to produce a novel view of the scene by computing where each of the pixels in the original views will appear in the novel image. The Layered Depth Image representation proposed by Shade et al. [18] provides a particularly efficient method for rendering data sets of this

form. These authors also describes a method for determining the relative depth of points in the scene by estimating the motion of various layers in the input imagery.

The problem with recovering dense depth maps from image sequences is that there are two important situations where it is exceedingly difficult for traditional correlation-based stereo algorithms to produce accurate depth estimates. The first situation corresponds to texture free regions in the scene, such as blank walls, which do not produce a sufficiently distinctive correlation signature. Occluding edges in the scene can also cause significant difficulties since the regions in the image surrounding such an edge will contain half occluded regions which cannot be adequately matched between frames by the correlation metric. The proposed method overcomes these problems by employing an epipolar plane image analysis to recover the positions of the feature points rather than a correlation based approach and by using an interpolation scheme which produces acceptable results in texture free regions.

The third category of approaches consists of techniques that draw inspiration from the image morphing schemes described by Chen and Williams [2]. Seitz and Dyer [16] propose a technique for producing physically correct images by interpolating between a given pair of views. Lhuiller and Quan [10,11] describe a view morphing technique which also seeks to produce interpolated views which correctly reproduce the motion of salient points in the scene. They describe a scheme for triangulating the input image in such a way as to respect intensity discontinuities. These papers demonstrate that it is possible to produce compelling interpolated images from a relatively sparse set of correspondences. Both of these approaches deliberately avoid the problem of estimating the actual 3D locations of the feature points that are used as correspondences. This means that the techniques can be applied to uncalibrated imagery but it also limits the systems to producing views that lie along the straight line connecting the two original images. The technique proposed in this paper eliminates this restriction by estimating the actual 3D locations of the observed feature points. This allows the system to predict where the features will appear in any viewpoint.

Genc and Ponce [4] propose an interesting approach to view interpolation by characterizing mathematically the possible positions of a selected set of feature points in the imagery. The proposed technique improves on this work by providing an automated technique for reproducing the motion of a much larger set of features in the scene, all of the edgels, and a more direct method for specifying the desired position of the novel viewpoint.

2 Obtaining Quasi-Sparse Feature Correspondences

As mentioned in the previous section, the Epipolar Plane Image (EPI) analysis technique proposed by Baker, Bolles, and Marimont [1] is used to recover the 3D positions of salient features in the scene. Other methods, such as feature based stereo or structure from motion could also be used for this purpose without affecting the correctness of subsequent arguments. EPI analysis is simply an

expedient, reliable, automatic technique for recovering the correspondences we are interested in.

In EPI analysis we consider the spatiotemporal intensity volume formed by collecting images at regularly spaced intervals as the camera moves along an axis parallel to one of the image axes. Figure 1a shows a slice of a typical spatiotemporal volume. This slice indicates how the intensity pattern along one row of the image changes as the camera moves along an axis parallel to that row. Note the characteristic banded structure of this image which can be explained by noting that feature points in the scene will correspond to straight lines in the epipolar plane imagery.

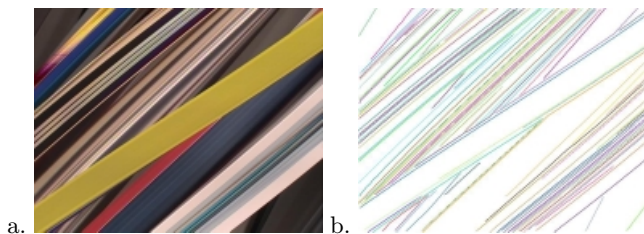


Fig. 1. a. A typical epipolar plane image obtained from the imagery. b. The result of applying the edge enhancement procedure to the epipolar plane image.

Each of these epipolar plane images is subjected to an analysis which seeks to extract these straight line features. This analysis is based on the techniques described by Baker et al [1] and Yamamoto [24]. A series of filters designed to enhance edges at various orientations is run over each epipolar plane image and the resulting edge elements are linked together to form straight line segments. Line segments that are deemed long enough and straight enough are interpreted as 3D feature points. Typical results obtained by invoking this procedure are shown in Figure 1b. The 3D location of the corresponding feature point in the scene is computed from the position and slope of the extracted line segments. The end result of the EPI analysis is a set of 3D points corresponding precisely to intensity discontinuities in the image.

In situations where the intensity discontinuity corresponds to the occluding edge of a curved object the trajectory in the spatiotemporal volume will correspond to the motion of the osculating ray. In most imaging situations this curve will still be a line to a good approximation since there will not be much variation in the position of the tangent point over the camera's trajectory.

Note that unlike correlation based approaches, this method for recovering the depth of feature points in the imagery does *not* assume that the intensity values surrounding corresponding points in the images will be strongly correlated. It simply exploits the fact that feature points in the scene correspond to straight lines in the EPI. The resulting line fitting problem is heavily overdetermined which serves to improve the accuracy and robustness of the method. This means

that the technique produces accurate estimates for the depth of occluding edges and other features that are problematic for correlation based methods. Occluding edges are particularly salient features in the scene and it is important to correctly predict where these features will appear in the novel view. Figure 2 shows an example of how drastically image regions corresponding to occluding edges can vary as the camera moves from one location to another.

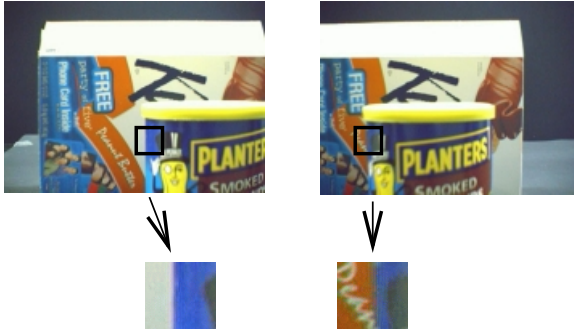


Fig. 2. The two neighborhoods shown differ greatly in appearance, even though they are centered around the same feature. Nevertheless, this feature corresponds to a single edge in the EPI and can, therefore, be recovered accurately.

One limitation of EPI analysis is that it will have difficulty recovering the depth of edges that are parallel to the direction of motion of the camera due to the aperture problem. In other words, a horizontal camera motion can be used to accurately recover the location of vertical edges but not horizontal edges and vice versa.

For this reason the gantry robot shown in Figure 3a is used to move the camera along two orthogonal axes. Images are acquired by scanning the camera along a set of horizontal and vertical paths as shown in Figure 3b. EPI analysis is applied to each scan in turn. For the images at the intersections of these scan paths we can combine the results of the horizontal and vertical analysis to produce depth estimates for all of the edgels in the image as shown in Figure 4. We refer to this structure as a Quasi-Sparse depth map since we have estimates for depth at a quasi-sparse collection of locations in the image, the intensity boundaries.

2.1 Interpolating Depth Values

Once a quasi-sparse depth map has been obtained for a particular image, one can produce a conservative approximation for the depth map associated with that viewpoint by obtaining the Delaunay triangulation of the projections of

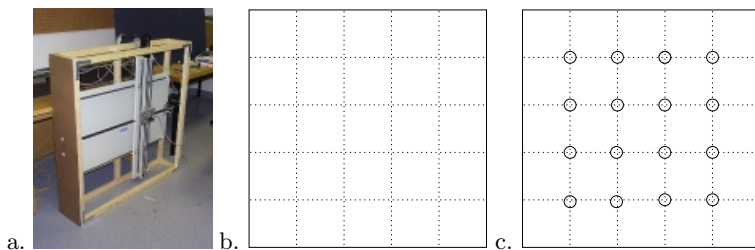


Fig. 3. a. The XY-positioning table used in the experiments b. An illustration of the camera scan paths. c. The circles show the camera locations corresponding to the intersections of the scan paths. For these camera positions the system can combine the results of the horizontal and vertical analyses to produce a quasi-sparse depth map.

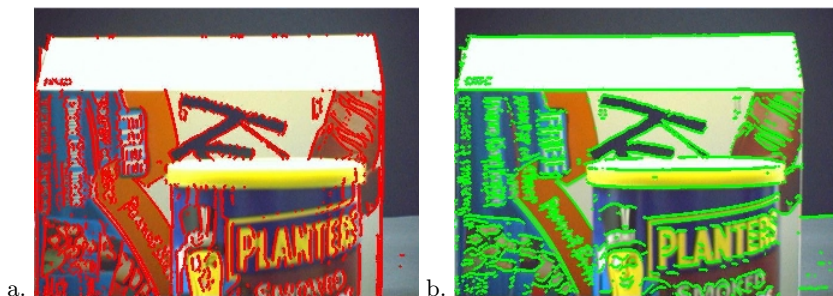


Fig. 4. a. A sample image with the pixels corresponding to visible recovered points from the horizontal camera motion shown in red. Notice that many points were recovered near vertical edges, and few were recovered near horizontal edges. b. The same image, but with visible recovered points shown in green. For vertical camera motion, there is a bias towards recovery near horizontal edges.

the feature points that are visible in that image¹ and linearly interpolating the $1/z$ values within each triangle. This corresponds to fitting planar surfaces to each of triangles in the scene. To see this, consider the equation for a plane in space, $ax + by + cz = 1$. Dividing each side of this equation by z we obtain $a(x/z) + b(y/z) + c = (1/z)$, that is $1/z$ is an affine function of x/z and y/z which implies that $a'u + b'v + c' = (1/z)$ where (u, v) corresponds to the projection of the point (x, y, z) in the image. We conclude that for a plane $1/z$ should be an affine function of the image coordinates u, v .²

Since the recovered scene points correspond to edges in the image, this construction has the agreeable property that the triangles tessellate the image without crossing image intensity discontinuities, which means that the resulting tri-

¹ It is a simple matter to record where various extracted features appear in the spatiotemporal volume used in the EPI analysis. We use this information to ensure that only the features that are known to be visible in a particular viewpoint are used to construct the depth map associated with that view.

² Hardware based texture mapping schemes also exploit this fact to their advantage.

angles have a strong tendency to correspond to facets of actual surfaces in the scene. Even triangles that do not correspond to surface facets encompass homogeneous regions in the image, such as blank walls, which can be morphed to the novel view in a convincing manner. The scheme also has an adaptive sampling property in that regions of the image that have a lot of intensity discontinuities are tessellated quite finely while other regions of the images that are less interesting are covered with fewer facets. This is appropriate from the perspective of view synthesis since we can usually get away with less refined interpolation in portions of the image that correspond to untextured areas. Figure 5 shows the results of a typical triangulation.

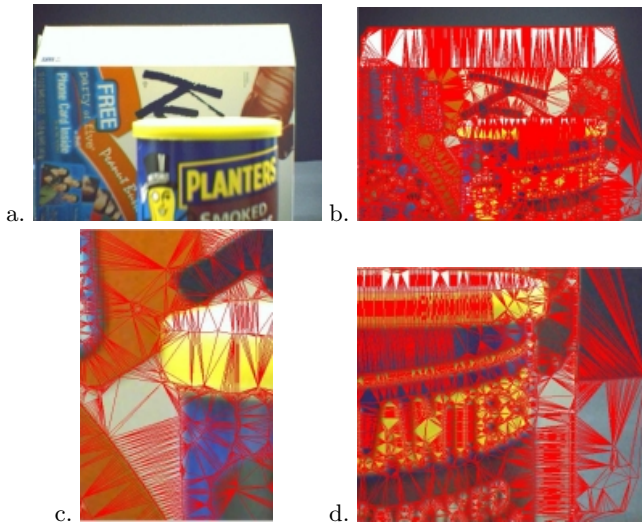


Fig. 5. a. An image taken from the intersection of a vertical and horizontal path. b. The triangulation of the projections of the visible recovered points. Notice that the vertices of these triangles lie along intensity edges. c. A close-up view of the left side of the canister. d. A close-up view of the right side of the canister. Notice that texture-free areas are covered by large triangles, while more detailed image areas have dense triangulations.

At the end of this stage, the system has constructed a set of triangular facets for each of the input images and the locations of the vertices of these triangles in the scene are known. At this point the system can generate novel views of the scene by supplying these triangles to a standard rendering pipeline and using the original images as texture maps. The rendering system correctly accounts for the parallax induced as a result of the motion of the virtual viewpoint and hidden surface removal reproduces the majority of the occlusion and disocclusion events that would be observed as the camera moves.

Note that in this framework any of the original images could be morphed to the novel viewpoint. In the current implementation the simple expedient of choosing the closest original viewpoint as a basis for morphing is employed with the idea that minimizing the difference in position between novel and original viewpoints will minimize the errors introduced by the viewpoint morphing operation. Results obtained by applying this simple morphing operation to a sample scene are shown in Figure 7. In most situations, this simple scheme produces acceptable results since it correctly reproduces the motion of the most perceptually salient features in the image, the intensity edges.

Unlike plenoptic sampling techniques, this approach only requires us to store the images taken at the intersections of the scanpaths and the associated depth maps. However, we are still able to produce convincing synthetic images from a reasonable range of viewpoints.



Fig. 6. 6 Original images of a doll house scene used in our view synthesis experiments



Fig. 7. Synthetic images produced by using the interpolated depth maps to morph the original images to novel vantage points

3 Reasoning about 3D Structure

While the procedure described in the previous section produces depth maps that can be used to perform novel view synthesis, it cannot predict the appearance of regions which are not visible from a given input view. In order to correctly reproduce the appearance of these regions we must combine information from multiple images into a composite description of the scene.

When accurate range scans are available, several excellent techniques are available for merging this information into a coherent surface [3,7,20]. Unfortunately, these techniques cannot be directly applied in this case since the interpolated range maps may underestimate the depth of the scene in a given image. We propose an alternative approach based on the following observation: while the individual interpolated depth maps may underestimate the depth of the surface in places, they do accurately predict the *freespace* in the scene. The following theorem states this observation more concisely:

Freespace Theorem : Suppose that three recovered space points P , Q , and R project to the pixel locations p , q , and r respectively in a given input image, and suppose further that $\triangle pqr$ is one of the Delaunay triangles formed in that image. Then the tetrahedron formed by the camera center and the space triangle $\triangle PQR$ must consist entirely of free space, i.e. no surface in the scene intersects this tetrahedron.

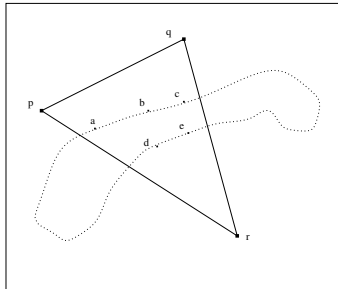


Fig. 8. If the space triangle $\triangle PQR$ is occluded by another surface in the scene when viewed from a particular vantage point then the corresponding triangle in that image, $\triangle pqr$, would contain points corresponding to the boundary of the occluding surface. This would preclude the triangle $\triangle pqr$ from the Delaunay triangulation of the image.

Proof: Assume, to the contrary, that there *is* a surface within this tetrahedron. Then the surface cannot occlude any of the points P , Q , or R (or else the occluded point would not be visible in the image). Therefore the boundary of the occluding surface must lie at least partially inside the tetrahedron. Points from the boundary would then show up as edges in either the EPI generated by horizontal camera motion or the EPI generated by vertical camera motion, and so we would expect this to cause at least one point to be recovered that has its

projection inside $\triangle pqr$ (see Figure 8). This contradicts our assumption that p , q , and r are part of a single Delaunay triangle, since the interiors of Delaunay triangles cannot contain other vertices³. Hence, the tetrahedron must consist entirely of free space. ♣

Note that this argument rests on the assumption that we have a procedure that is capable of accurately reconstructing the depths of the majority of the edge features in an image. If the reconstruction system only returned the depths of isolated corner features in the image, this property might not hold.

We can use the freespace theorem to draw conclusions about the 3D structure of the scene. First consider the star shaped freespace volumes defined by the interpolated depth maps associated with each of the input viewpoints. Then consider the union of these freespace regions as shown in Figure 9. Notice that the union provides a more accurate approximation of the structure of the scene than any of the original depth maps. It is also important to notice that the input depth maps need not be particularly accurate; they may underestimate the scene depth in places but these errors can be discovered and corrected through the freespace union procedure.

Given the coordinates of a point in space, P , we can easily test whether that point is in the union of the freespace volumes by projecting the point into each of the original viewpoints and determining whether the depth of the point P with respect to the image center is less than the depth of the corresponding entry in the original interpolated depth map for that image.

This procedure can be represented by an indicator function $\Phi(P) : R^3 \rightarrow [0, 1]$ which returns 0 whenever the point P lies within the union of the freespace volumes and 1 otherwise. This function can be thought of as an *implicit* representation of the total freespace volume.

Armed with this function, $\Phi(P)$, we can determine where a ray in space intersects the freespace boundary by sampling this function at various points along the ray in search of a transition. Once a transition has been found it can be localized rapidly using standard bisection search techniques [14].

This ray intersection procedure can be used to determine the depth from any given vantage point to the freespace boundary as shown in Figure 9. It can also be used to determine whether or not a particular point is visible from a given camera position since this amounts to testing whether the ray between the given point and the camera center lies entirely in freespace.

The boundary of the freespace volume can be thought of as a fair approximation for the surface geometry in the sense that it will correspond to a surface which is consistent with all of the available information. Notice, however, that the actual surface may contain regions that are not visible from any of the camera positions. The freespace union method will still produce reasonable results in these situations.

³ In fact, the defining property of Delaunay triangulations is that the interior of the *circumcircle* of every triangle does not contain any other vertices; however, we do not require such a strong property for our argument.

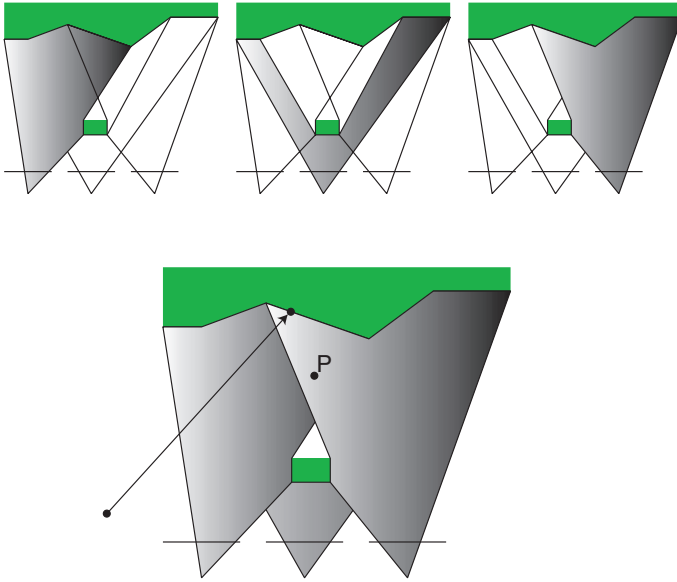


Fig. 9. The first row of figures depicts the freespace volumes associated with the triangulations of each of the 3 input images. The union of these volumes provides a more accurate approximation of the 3D structure of the scene than any of the original depth maps. It is a simple matter to construct a function $\Phi(P)$ which indicates whether a particular point P lies within the union of the freespace volumes.

One could, in principle, construct an explicit representation for the freespace boundary from the indicator function $\Phi(P)$ by invoking an isosurface algorithm such as Marching Cubes [17]. However, this would involve deciding on an appropriate discretization of the scene volume. Coarse discretizations would result in meshes that failed to capture fine details of the surface but fine discretizations would produce exceedingly large meshes. Fortunately, novel views of the scene can be synthesized without the aid of an explicit surface.

4 Novel View Synthesis

The techniques described in the previous section can be employed to refine the conservative depth maps associated with each of the input images. For every pixel in a given input image we can construct the ray passing between the camera center and the pixel center and determine where that ray intersects the freespace boundary. This analysis gives us a refined estimate for the depth of the scene at that pixel.

The resulting depth maps can be further refined by applying a smoothing procedure similar to anisotropic diffusion which takes into account the locations of intensity discontinuities in the original image and preserves salient depth discontinuities in the scene. The ultimate goal of this smoothing procedure is to

produce depth maps that have crisp boundaries corresponding to edges in the intensity image. In this sense, the method is similar in spirit to the scheme proposed by Tao, Sawhney and Kumar [21,22]. However, our approach to producing these depth maps based on an analysis of the freespace volume is completely different.

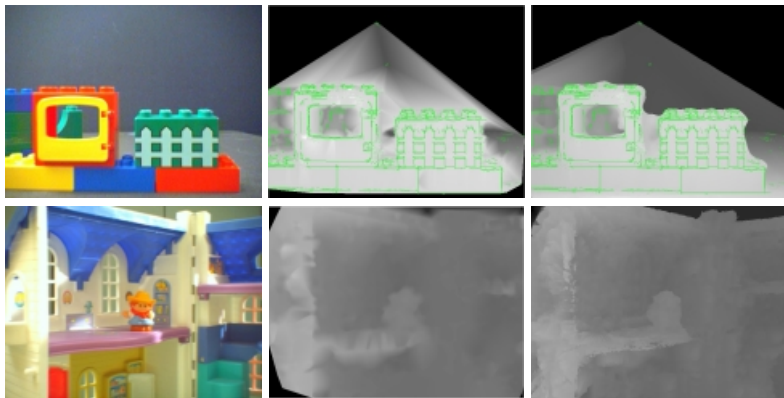


Fig. 10. The first column of figures corresponds to the input images, the second column depicts the original interpolated depth maps and the last column shows the results of applying the freespace analysis procedure to refine the depth maps. Notice that freespace analysis correctly recovers many of the sharp depth discontinuities that were missing in the original interpolated depth maps.

Figure 10 shows the results of applying this procedure to typical input images. The refined depth maps shown in Figure 10 are remarkably accurate considering that they were produced from quasi-sparse depth samples. Note that in both cases the procedure correctly recovers many of the sharp depth discontinuities associated with occluding edges in the scene, discontinuities that were not present in the original interpolated depth maps. When these depth maps are used to morph the corresponding input images to a novel vantage point, these depth discontinuities can produce holes in the synthetic image. This problem is handled by morphing several input views to the new viewpoint and combining the resulting images to fill the gaps. Figures 12 and 13 show some of the results obtained by applying this view synthesis procedure to actual image data.

The scene with the Lego blocks was chosen precisely because the texture free surfaces and occluding edges would pose a challenge for most correlation based stereo algorithms. Notice that the view synthesis procedure correctly reproduces the appearance of the distant green block which is seen through the window of the foreground Lego block. This demonstrates that the system is able to reason correctly about the complex 3D structure of the scene.

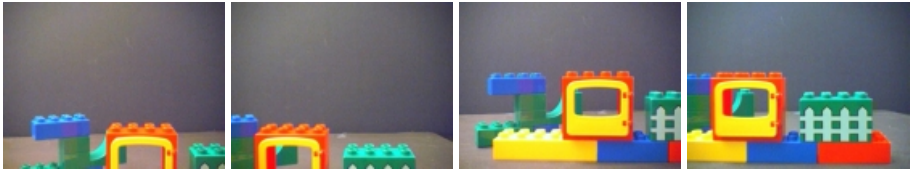


Fig. 11. Original images of a scene of blocks that were used to produce novel views

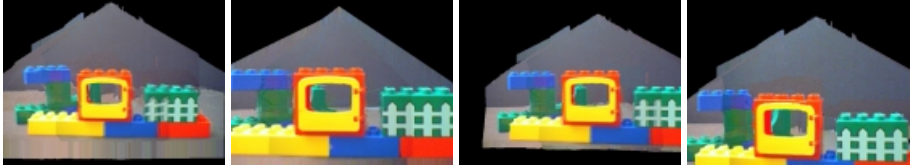


Fig. 12. Synthetic images produced by freespace analysis method.



Fig. 13. Synthetic images produced by freespace analysis method.

Figure 14 shows a side by side comparison of actual images and the corresponding synthetic views. The normalized correlation coefficient between the synthetic views and the actual images is greater 0.94 in both cases.

5 Conclusions

This paper describes an approach to novel view synthesis based on a quasi-sparse set of feature correspondences. This scheme hinges on the observation that human viewers tend to be sensitive to the motion of intensity discontinuities so the system focuses its efforts on recovering accurate estimates for the depths of these features. The resulting quasi-sparse depth maps resemble the output of an edge extraction procedure where the edgels have been augmented with accurate depth information. In each of the input images a Delaunay triangulation of the feature points is used to produce an interpolated depth maps which provide an approximation of the surface structure that can be used to morph the input views to novel vantage points.

We have also presented a principled approach to reasoning about the 3D structure of the scene by analyzing the union of the freespace volumes associated with each of the interpolated depth maps. This freespace union procedure provides a powerful constraint which allows us to produce relatively accurate

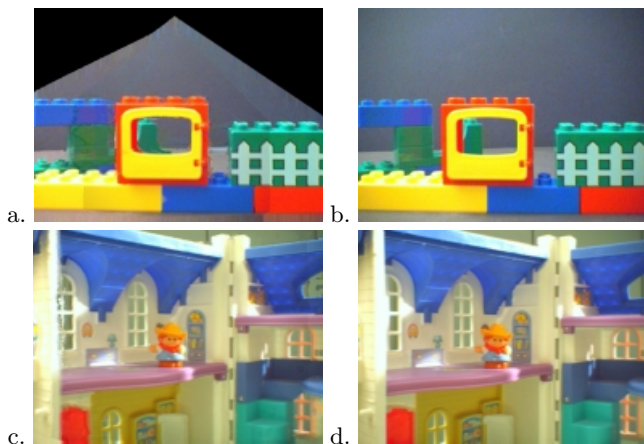


Fig. 14. The first column corresponds to synthetic views and the second to actual images of the scenes. Note that the view synthesis procedure correctly reproduces the motion of the green block which is viewed through the window of the foreground lego block.

depth maps from quasi-sparse feature correspondences in multiple images. Importantly, the arguments made about analyzing the freespace volumes could also be applied to situations where depth information is obtain using other means such as range finders or stereo systems. This analysis allows us to correctly reproduce occlusions and disocclusions in the synthetic views.

5.1 Future Work

The current system makes use of Epipolar Plane Image analysis since it provided a simple and accurate automated method for obtaining the depth of the salient features in the scene. It should be possible to obtain similar information about the scene features from standard video sequences by applying structure from motion techniques. It should also be possible to extend the view synthesis technique so that it could take advantage of correspondences provided by correlation based stereo methods whenever they are available.

Better methods for smoothing the refined depth maps based on information in the intensity images are being investigated. We also intend to consider extensions of the scheme that will allow us to model how the radiance of observed scene points changes as a function of viewing direction so that we can correctly reproduce specularities and other view dependent effects [15].

References

1. R.C. Bolles, H.H. Baker, and D.H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1(1):7-55, 1987.

2. S. E. Chen and L. Williams. View interpolation from image synthesis. In *SIGGRAPH*, pages 279–288, August 1993.
3. Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of SIGGRAPH 96. In Computer Graphics Proceedings, Annual Conference Series*, pages 31–43, New Orleans, LA, August 4-9 1996. ACM SIGGRAPH.
4. Y. Genc and J. Ponce. Parameterized image varieties: A novel approach to the analysis and synthesis of image sequences. In *International Conference on Computer Vision*, pages 11–16, January 1998.
5. Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael Cohen. The lumigraph. In *Proceedings of SIGGRAPH 96. In Computer Graphics Proceedings, Annual Conference Series*, pages 31–43, New Orleans, LA, August 4-9 1996. ACM SIGGRAPH.
6. V. Hlavac, A. Leonardis, and T. Werner. Automatic selection of reference views for image-based scene representations. In *European Conference on Computer Vision*, pages 526–535, 1996.
7. T. Kanade, P.W. Rander, and J. P. Narayanan. Virtualized reality: Constructing virtual worlds from real scenes. *IEEE Multimedia*, 4(1):34–47, 1997.
8. S. Laveau and O.D Faugeras. 3-d scene representation as a collection of images. In *International Conference on Pattern Recognition*, pages 689–691, 1994.
9. Marc Levoy and Pat Hanrahan. Light field rendering. In *Proceedings of SIGGRAPH 96. In Computer Graphics Proceedings, Annual Conference Series*, pages 31–43, New Orleans, LA, August 4-9 1996. ACM SIGGRAPH.
10. Maxime Lhuiller and Long Quan. Image interpolation by joint view triangulation. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 139–145, 1999.
11. Maxime Lhuiller and Long Quan. Edge-constrained joint view triangulation for image interpolation. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 218–224, 2000.
12. Ko Nishino, Yoichi Sato, and Katsui Ikeuchi. Eigen-texture method: Appearance compression based on 3d model. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 618–624, 1999.
13. M. Pollefeys, L Van Gool, and M. Proesmans. Euclidean 3d reconstruction from image sequences with variable focal lengths. In *European Conference on Computer Vision*, pages 31–42, 1996.
14. W. Press, B.Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1988.
15. Y. Sato, M.D. Wheeler, and K. Ikeuchi. Object shape and reflectance modeling from observation. In *Proceedings of SIGGRAPH 97. In Computer Graphics Proceedings, Annual Conference Series*, pages 379–387. ACM SIGGRAPH, August 1997.
16. Steven Seitz and Charles R. Dyer. View morphing. In *Proceedings of SIGGRAPH 96. In Computer Graphics Proceedings, Annual Conference Series*, pages 31–43, New Orleans, LA, August 4-9 1996. ACM SIGGRAPH.
17. J.A. Sethian. *Level Set Methods; Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Material Sciences*. Cambridge University Press, 1996.

18. Jonathan Shade, Steven Gortler, Li wei He, and Richard Szeliski. Layered depth images. In *Proceedings of SIGGRAPH 98. In Computer Graphics Proceedings, Annual Conference Series*, pages 231–242. ACM SIGGRAPH, August 1998.
19. Heung-Yeung Shum and Li-Wei He. Rendering with concentric mosaics. In *SIGGRAPH*, pages 299–306, August 1999.
20. Ioannis Stamos and Peter Allen. 3-d model construction using range and image data. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2000.
21. Hai Tao and Harpreet Sawhney. Global matching criterion and color segmentation based stereo. In *Workshop on the Applications of Computer Vision*, pages 246–253, December 2000.
22. Hai Tao, Harpreet Sawhney, and Rakesh Kumar. A global matching framework for stereo computation. In *International Conference on Computer Vision*, pages 532–539, 2001.
23. T. Werner, R.D. Hersch, and V. Hlavac. Rendering real-world objects using view interpolation. In *International Conference on Computer Vision*, pages 957–962, 1995.
24. Masanobu Yamamoto. *The Image Sequence Analysis of Three-Dimensional Dynamic Scenes*. PhD thesis, Electrotechnical Laboratory, Agency of Industrial Science and Technology, May 1988.