A NEW METHOD FOR IMAG E-BASED REPRESENTATION AND FAST RENDERING OF THE 3D OBJECT OR SCENE

Li Xinxiao, Zheng Nanning

AI and Robotics Institute,
Xi'an Jiaotong University, Xian Shaanxi 710049, China
lxx@aiar.xjtu.edu.cn, nnzheng@xjtu.edu.cn http://www.aiar.xjtu.edu.cn

ABSTRACT

An efficient image-based modeling and rendering approach is presented. This method, which is an extension of layered depth image (LDI), uses a panorama-like data structure to describe the model of an object or scene. The representation has the conventional advantages resting with 2D image-space, such as regularity of pixels' position and epipolar line relationship among multiple views. It is fitful to compress N sample images efficiently, and generate new views in interactive rate.

Keywords: image-based rendering, panoramic image, image warping, layered depth image

INTRODUCTION

In traditional computer graphics, 3D geometry consists of a series of 3D elements (triangle surfaces, voxels). Output images are generated from complicated lighting simulation that calculates the reciprocal reaction between light rays and reflective surfaces. Recently, an image-based rendering technique (IBR) is deeply studied in fields of computer vision and computer graphics. In this technique, a group of sample images are directly employed to synthesize new views. The input comes from not only computer-generated images but also the photographs taken from the real scene, which has utmost complexity.

Chen and Williams [1] use image interpolation to generate inter-views, which deals with variation of the view position and direction. McMillan and Bishop [2] exploit cylinder panoramas to sample the plenoptic function in 3D space, and re-project to new views. Levoy et al [3] present an approach to sample the light density at all positions and in all directions of a bounded space from a list of sample images. A large image list is needed in this approach, which exacerbate the insufficient memory. All the methods either limit the variation of view parameters or cost large storage.

New problems are caused when N sample images are directly used to describe an object or scene. If not properly compressed and simplified, the data are directly proportional to N. For complicated hidden surfaces or a large scene, the number of sample images increase further, and then the cost of storage and computation increase linearly. In practical applications, fields of view (FOV) are overlapped between two neighboring views in order to acquire accurate surfaces. One space point in overlapping portions can be projected into multiple sample pixels, but many of them are redundant because only one pixel with its depth is needed to reconstruct the surface point. How to efficiently compress these redundant pixels is a key problem of IBR technique. William et al [4] bring forward a compression frame called Delta-tree to describe an object or scene. Delta-tree can effectively compress N sample images and fast synthesize new views, but this method is only available to protruding surfaces.

In this paper, a new data structure is put forward to compactly represent an object or scene. The representation called layered depth panorama (LDP) is an extension of LDI that presented by Shade [5]. It uses a simple data structure analogous to that of an 2D image-space to store all sampled surface information. This method not only perfectly describes complicated hidden surfaces, but also efficiently derives new views. We describe the LDP structure and its two prominent advantages: (1) exploiting the regularity of image-space to efficiently warp sample pixels to new views; (2) using the relationship of epipolar lines in muiti-view geometry to avoid the Z-buffering or depth comparison.

DEFINITION

Traditional cylinder panorama is the cylinder view that is acquired by rotating the camera in the fixed point. When several around plane images are stitched into a panorama, the stitching process is independent on pixel depths. After the view position or direction change, we can't get the correct re-projections due to the influence of parallax that relates with depths.

LDP records all space points of an object or scene along view-rays of a panorama-view. Not only the visible but also hidden points to the center of projection (COP) are preserved. It differs from a traditional panorama in that there are more than one point in different depths along a ray corresponding to the pixel position. Figure 1 displays the difference among LDP, traditional panorama and LDI. Figure 1(a) is a traditional panorama, Figure 1(b) and (c) are LDP, and Figure 1(d) is LDI. Showed From the figure above, LDP describes not only the surrounding scene but also the surface of a central object, whereas the traditional panorama describes only the surrounding scene. LDI projects all surface points to a plane image, so decreases the resolution of back-surfaces if their depths have large difference with the front-surface points, however LDP projects all surface points to a panorama-image that more equably samples the surfaces. Correspondingly, new views can be generated from LDP when moving and rotating the camera, while conventional panorama can't generate correct new views when the COP changes. The data structure of LDP is described as following:

LDP{
Camera; // LDP camera parameters
LayeredDepthPixel[Xres,Yres];// Xres,Yres is the width and height of the unfold panorama
}

// For every layer depth pixel
LayeredDepthPixel{
&DepthPixel;
&Next; // pointer to the next layer depth pixel
}

//The color and depth of every depth pixel
DepthPixel{
PixelGray;
Zdepth; // surface point's depth from COP
}

RECONSTRUCTINTG AND RENDERING

Accordingly, we use N sample images to re-construct a LDP, then re-project them to new views. The process includes two transformations that are from plane image to LDP and from LDP to plane image. In this section, we first deduce the two transformations; next the approach of warping pixels without depth comparison is exploited by using epipolar line relationship of multi-view geometry; finally the rendering flow is presented.

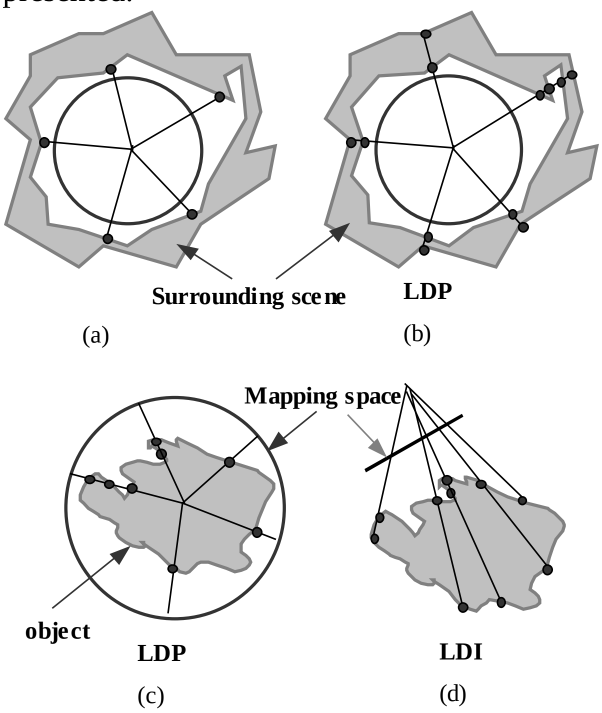


Figure 1. The difference among traditional panorama, LDP and LDI.

BI-DIRECTION TRANSFORMATION

In plane pinhole camera model, each image-pixel corresponds to a traversing ray originating from COP. The mapping relationship between a pixel and its corresponding ray-direction is described by

$$\bar{d}_1 = P\bar{x}_1 \qquad (1)$$

where $\bar{x}_1 = [u, v, 1]$ is the homogeneous coordinate, $\bar{d}_1$ is its corresponding ray-direction, and P is the mapping matrix. Because the product of a scalar factor and P does not change a ray direction, P is independent on common factors. The surface point corresponding to $\bar{x}_1$ is described by

$$\dot{X} = C_1 + t_1\bar{d}_1 = C_1 + t_1 P\bar{x}_1 \qquad (2)$$

where C_1 is COP, and t_1 is the depth from C_1 corresponding to surface point $\bar{x}_1$.

In cylinder coordinate space, the one-to-one relation between a surface point and its ray-direction is represented by

$$\bar{d}_2 = \begin{bmatrix} \sin(2\pi u + k_{cylinder} v) \\ y_0 + (y_1 - y_0) \\ \cos(2\pi u + k_{cylinder} v) \end{bmatrix} = E(\bar{x}_2),$$

$$u, v \in [0,1] \qquad (3)$$

where y_0 and y_1 is the lower-bound and upper-bound of a cylinder, and $k_{cylinder}$ is an oblique parameter of the cylinder space. From (1) to (3), we derive the warping transformation between LDP and a plane image, described by

$$\bar{x}_1 = P^{-1}[\delta(\bar{x}_1)(C_2 - C_1) + E(\bar{x}_2)] \qquad (4)$$

$$\bar{x}_2 = E^{-1}[\delta(\bar{x}_2)(C_1 - C_2) + P\bar{x}_1] \qquad (5)$$

where $\delta(\bar{x}_1) = \frac{1}{t_1}$ or $\delta(\bar{x}_2) = \frac{1}{t_2}$ is called as *general disparity*, Eq. (5) is used for re-constructing LDP, Eq. (4) corresponds with LDP's warping (or re-projecting) to new views. Due to the non-linear transformation $E(\circ)$ in the Eq. (4) and (5), incremental computation is not fitful for cylinder panorama. But we may pre-compute unchanged mid-terms for efficient rendering because only one "image" (LDP) is needed to represent the whole model. The term $E(\bar{x}_2)$ represents ray directions that are pre-computed and stored in LDP pixel's positions, so we use these values in a memory table to derive new views, which is pivotal in interactive visualization applications

WARP WITHOUT DEPTH COMPARISON

Warping transformation is a many-to-one mapping, and output pixels are the ones corresponding to nearest surface points from the COP. With epipolar line relationships in multi-view geometry, depth comparison is avoided to resolve visibility and warp pixels orderly.

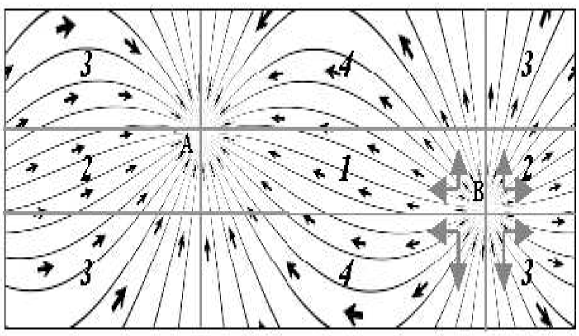


Figure 2. The scan order in the cylinder LDP

Epipolar points are got by way of linking the COP of a new view and the COP of LDP. The positive epipolar (A in Figure 2) is the crossed point between linking line of COPs and the cylinder surface, and the negative epipolar point (B in Figure 2) is the cross point between the extended line and the cylinder surface. Showed in Figure 2, the two epipolar points horizontally and perpendicularly partition the unfolded cylinder surface to four quadrants corresponding to nine areas (when the epipolars are not on the surface of the limited cylinder surface, only 6 or 3 sections are got). In figure 2, the top may be looked as the unlimited extension of the bottom, the right as the extension of the left. At each pixel, an epipolar line departs from the negative epipolar point and converge at the positive epipolar point. As long as the pixels are selected and warped along epipolar lines, the new view can be derived by scanning pixels one by one without depth comparison, and the mapping from the farthest to nearest along a view-ray is automatically satisfied. The gray bold arrow shows the scan order in every quadrant. At every layer of the layered depth pixel, multiple points are scanned from the largest to the smallest depth value.
We describe the re-construction process as follows:

Begin
warp each depth image to LDP
  warp each depth pixel to LDP //Eq. (5)
  If(the warped position overlaps with the former one)
    calculate the difference of warped depths
    If (the difference is larger than ξ *)*
      Add one layer at the pixel's position
    Else
      Average the colors of warped pixels at this depth;
End

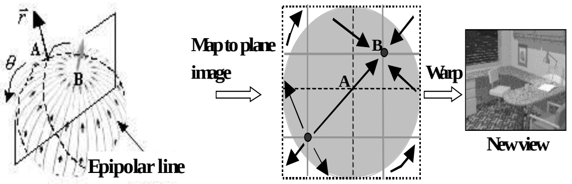
The rendering flow of LDP is as following:
Begin
 Warp new COP to LDP to partition it into quadrants
  Warp each quadrant of LDP to the new plane view
   Warp every layered depth pixel // Eq. (4)
    Warp LDP's pixels in increasing order of depths
 End

FULL-ORIENTATION LDP

In cylinder LDP described above, view-direction isn't permitted in top and bottom, so the top-view and bottom-view is incomplete. To get the whole model along full view directions, the direction extension using a spherical surface instead of a cylindrical surface is fitful. A spherical surface isn't directly correspondent to a plane image as a cylindrical surface is, which is inadvantageous for storage. In order to store LDP to a plane image, we establish the pixel's correspondence that maps each LDP's pixel to one pixel of the plane image. The correspondence is defined as:

LDP's pixels are mapped to pixels in a plane image one by one. The distance from the plane image center to its pixel is proportional to the angle between view-direction of LDP to a LDP's reference view-direction

Warping from LDP to new views is processed orderly without depth comparison by exploiting the relationship of epipolar lines in multi-view geometry. All epipolar lines depart from the negative epipolar point and converge to the positive epipolar point, as is shown in figure 3(a). In figure 3(b), the gray circle is the mapping area from LDP to the plane image, the arrowhead lines hint the directions of epipolar lines that guarantee the correct scan order in that area without depth comparison.



(a) (b) (c)

Figure3. Map from a sphere LDP to a plane image and warp to new views

Another extension uses a cube surface instead of cylinder surface as a LDP, where LDP's center is the cube center. The cube surface is composed of six planes. Warping transforms among the plane images is linear, so incremental computation can be employ to accelerate rendering process. Correspondingly, Eq. (4) changes to

$$\bar{x}_1 = \delta(\bar{x}_1)\left[P^{-1}(C_2 - C_1)\right] + \left(P^{-1}E\right)\left(u \quad v \quad 1\right)^{-1}$$
$$= \delta(\bar{x}_1)\mathrm{A} + \mathrm{B}\left(u \quad v \quad 1\right)^{-1} \tag{6}$$

The first item has one multiplication, and the second item can be computed in row or column order, where

$$\mathrm{B}(u+1 \quad v \quad 1)^{-1} = \mathrm{B}(u \quad v \quad 1)^{-1} + \mathrm{B}(1 \quad 0 \quad 0)^{-1}$$

= the former value + the first row in B

$$\mathrm{B}(u \quad v+1 \quad 1)^{-1} = \mathrm{B}(u \quad v \quad 1)^{-1} + \mathrm{B}(0 \quad 1 \quad 0)^{-1}$$

= the former value + the second row in B

With the epipolar line relationship among the plane views, depth computation and depth comparison can be avoided. McMillan proved the occlusion-compatible visibility algorithm for plane images in [6].

MORE DISCUSSION

Surface points in different depths have different sample resolutions. The more distant are the surface portions from the COP, the lower are their sample resolutions. LDP is superior to LDI resting with its approximately equal sample resolutions of most surfaces. It projects a space point to widely spaced different directions, where LDI projects the same point in a comparatively concentrated directions.

If the COP is on sides of an object or scene, the projected points are centralized in some directions than the other directions, so the depth complexity increases in local portions of the image space, and the sample resolution of back-surfaces is decreased. This situation is effectively improved by setting the COP to the geometry center, which reduces the distant hidden surfaces or back-surfaces.

Acquiring depth values of sample points is the difficult step in IBR technique [2][5]. If the images is computer-generated (volume rendering, geometry rendering etc), sample depths are the final z-buffering values. For the real scene, many optical, supersonic and mechanical methods are used to acquire and infer the depth values of sample images. Borghere et al [7] measure the depth of a man face model at the precision less than 0.1mm with a video camera and some laser spots. Warping algorithm is depth-insensitive, rude depth information can synthesize satisfied visual images, but the depth samples in multiple views need to be identical.

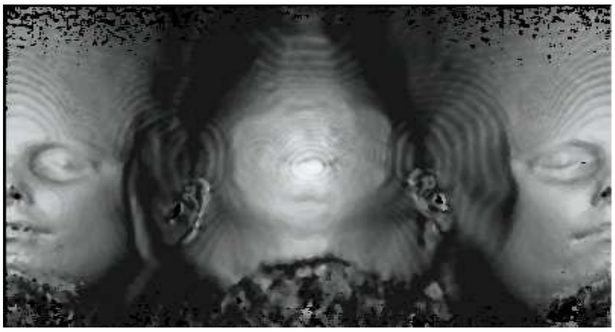
Since LDP is a compact description of surfaces of 3D objects, we may modify the traditional graphics rendering to construct LDP of 3D models. The modified graphics rendering system records all crossed surface points instead of only visible points along view-rays.

EXPERIMENT RESULT

We experiment with two types of data, a head medical CT data and an indoor scene generated from 3DS. From both, we can get the accurate depth images from corresponding graphics rendering. Figure 4(a) is the cylinder

LDP generated from 18 depth images, the greatest depth layer is 4, the average depth layer is 1.06, and the compression ratio from the sample images to LDP is 8.5:1. Figure 4(b) is the new views warped from LDP. The rendering time is less than 30ms.

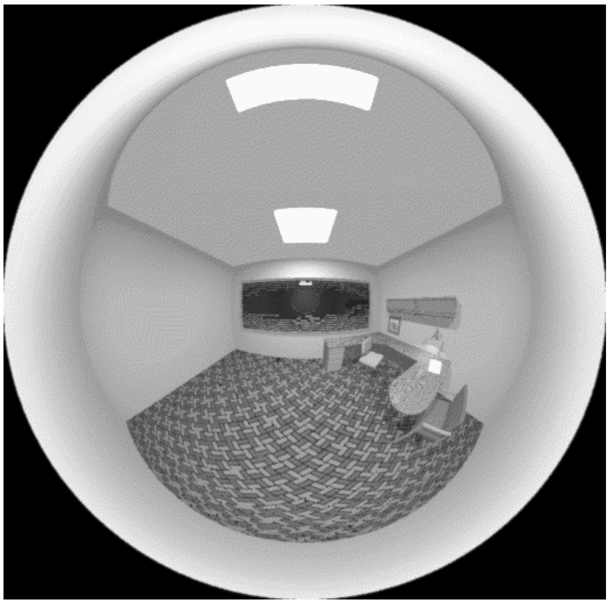
Figure 5 shows the example of a surrounding scene's LDP that is constructed by a previous ray tracing software with a slight modification. Figure 5(a) is the sphere LDP, the greatest depth layer is 5, and the average depth layer is 1.02. Figure 5(b) is the cylinder LDP, the vertical view-angle is 90, the greatest depth layer is 5, and the average depth layer is 1.03. Figure 5(a) and 5(b) seem to be the same as the traditional panorama when observed from the visible layer, but the hidden surfaces are included in the LDP, that is to say, multiple depth pixels is preserved in one pixel position. Fig5(c) is the new view generated from the sphere LDP, whose rendering time is about 56ms.



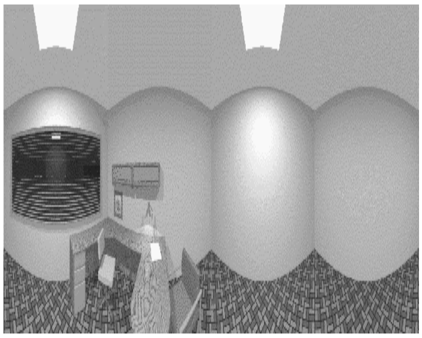
(a)

(b)

Figure 4.LDP constructed from volume rendering and the new view derived from it.



(a)



(b)

(c)

Figure 5. LDP of the surrounding scene and a derived view

CONCLUSION

In this paper, an efficient image-based representation of an object or scene is presented. It not only effectively compresses sample images but also describes surfaces of an object or scene with complicated hidden surfaces. LDP is different from traditional panorama or LDI. Its major advantage is to equally preserve the surface in an image-like data structure and generate new view in the change of camera's position and direction. With characteristics of image-space, the warping transformation is incrementally computed in interactive applications and the visibility are solved without depth comparison.

REFERENCES

[1] S. E. Chen, and L. Willams, *View Interpolation for Image Synthesis*, SIGGRAPH'93, 1993, pp. 279-188.

[2] L. McMillan, G. Bishop, *Plenoptic Modeling: An Image-based Rendering System*, SIGGRAPH'95, 1995, pp. 39-46.

[3] M. Levoy, P. Hanrahan, *Light Field Rendering*, Computer Graphics (SIGGRAPH'96 Conference Proceedings), August 4-9, 1996, pp. 31-42.

[4] J. D. William, L. McMillan, G. Bishop, and H. Funcs, *The Delta Tree: An Object-Centered Approach to Image-based Rendering*, MIT AI LAB Technical Memo 1604, May 1996.

[5] J. Shade, S. J. Gortler, Li-wei He, R. Szeliski, *Layered Depth Images*. SIGGRAPH'98, 1998, pp. 231-242.

[6] L. McMillan, *A List-Priority Rendering Algorithm for Redisplaying Projected Surfaces*, Department of Computer Science, UNC, Technical Report TR95-005, 1995.

[7] N. A. Borghese, G. Ferrigno, et al, *Autoscan: A Flexible and Portable 3D Scanner*, IEEE Computer Graphics & Application, May/June 1998, pp. 38-41.