Image Based Rendering

Krisha Chawan[1], Khushi Chowdhary[2], Hrishikesh Dalvi[3], Jesleena Gonsalves[4]

[1,2,3]Student at NMIMS Mukesh Patel, School of Technology Management and Engineering, Mumbai, Maharashtra, India
[4]Assistant Professor, at NMIMS Mukesh Patel, School of Technology Management and Engineering, Mumbai, Maharashtra, India

Abstract
Image-Based Rendering (IBR) is known to be a developing technology in the field of computer graphics due to its high potential for applications in various scenarios. This technique utilizes existing images as the main data source and captures the essence of it to generate realistic visual representations of a scene in three-dimension. For several years, research has been carried out for various techniques which may or may not require the geometric specifications of objects in a scene. Hence, we introduce a method for each type of classification.

Keywords Image-Based Rendering, geometry, photorealistic, techniques

1. Introduction

At present-day, the field of computer graphics has continuously been advancing with incorporating virtual environments. In order to connect the traditional computer graphics with the objective of virtual experiments, image-based rendering has been employed to aid the process. Image-based rendering involves the generation of photorealistic or non-photorealistic images from two dimensional images of a scene which leads to the creation of a three-dimensional model.

Understanding virtual reality technology has been a key component in the field of computer graphics as virtual environments provides a platform for individuals to experience visuals with enhanced efficiency and fidelity. This can be easily achieved through image-based rendering techniques which can help model realistic virtual scenes. Virtual scenes in 3D environments enhances the tactile, auditory and visual experiences for the user, providing an immersive platform to indulge in.

Image-based rendering leverages real world images which provides us with realistic visual scenes and helps the technology achieve a higher level of realism in the field of computer graphics. These can be further applied in the field of gaming and medical imaging which will be later introduced and discussed.

2. Rendering without geometry

In this section of the review paper, we'll discuss and analyze the representative techniques for rendering different unknown scene geometry.

The base of these techniques lies on the use of Plenoptic function of Adelson and Bergen. It is a parameterized function that describes things in the surrounding which are visible from a given position in space. It represents the intensity of each light ray as a function of visual wavelength, angle, time and viewing position and is related to various concepts such as "the structure of ambient light" or "visual pyramid" developed by different scientists. The complete plenoptic function is a 7-dimensional function with observing locations (V_x, V_y, V_z), the angle of incidence as (θ, ϕ), wavelength as λ and time t. Thus, the 5-Dimensional plenoptic function can be written as:

$P_5 = P(V_x, V_y, V_z, \theta, \phi)$

With the help of plenoptic function, the various problems related to IBR can be redefined such as synthesis of novel views becomes more accurate by exploiting the angular information provided in the plenoptic function.

Light field and Lumigraph

The 5d plenoptic function can be simplifies to 4D plenoptic function,

P4 = P(u, v, s, t), where (u,v) and (s,t) are two parallel planes of bounding box.

This 4D structure is called Lumigraph, encapsulating appearance of a scene across different viewpoints. This provides advantage in visual perception in scenarios based on viewer's angle. Sets of images or color values for each point and each incoming light direction is stored by lumigraphs.

Light field are representations of light traversing in direction through each point in space. This includes both intensity of light and its direction. Light fields are used for simulating light effects and depth of field. In light field system, rig is designed to obtain sample images. Light field is pre-filtered before rendering to reduce aliasing effect. Aliasing is a visual artifact in computer graphics and digital imaging that occurs when high-frequency patterns or details in an image are misrepresented or distorted during the process of sampling or rendering.

Amount of data in light field rendering is reduced by vector quantization scheme. On other hand, lumigraph constructs set of images taken from viewpoints.

Concentric mosaic

Plenoptic function becomes simpler as we have more constraints on camera location n (Vx, Vy, Vz). For capturing all viewpoints, we need a complete 5D plenoptic function. We limit the movement of the camera to follow circular paths in a plane, and generate mosaics that consist of concentric circles. Each circular path corresponds to a distinct manifold mosaic.

We can show that concentric mosaics represents plenoptic function parameterized by three variables

1. Rotation angle
2. Radius (distance of camera to rotation axis)
3. Vertical elevation or field of view

At rendering time, novel views are produced by efficiently combining the relevant captured rays. Even though the rendered images have vertical distortions, depth correction can reduce them. Similar to panoramas, concentric mosaics do not require recuperating geometric and photometric scene models. Furthermore, concentric mosaics offer a far better user experience since they let the user freely navigate a circular area while noticing noticeable lighting and parallax changes. In contrast to a Lightfield or Lumigraph, concentric mosaics possess significantly reduced file sizes since they solely construct a 3D plenoptic function. Concentric mosaics are easy to capture and have good space and computational efficiency.

1. Rendering Novel view

1.1 Rays in capture view

Since concentric mosaics have captured the majority of rays in the capture plane, we can render any novel ray in the plane given a collection of concentric mosaics. When rendering, all we have to do is locate and determine where the previously recorded concentric mosaics' rays of the novel view—that is, a horizontal line image in the plane—are, or we can bilinearly interpolate them from nearby mosaics.

1.2 Rays off the plane

Only a small subset of the rays off the plane is stored in concentric mosaics because a slit image rather than a regular image is captured. This is the reason why concentric mosaics have a much smaller file size than a Lightfield or Lumigraph.

2. Depth Correction

2.1 Full perspective depth correction

When distances between the camera optical center and points are known, full perspective correction is used. Using optical flow, pixel values of rays can be warped to render view. There are 2 problems with this approachgeometrics are not known for real scenes and hole -filling problems still exists.

2.2 Weak perspective approximation

For real scenes, pixels from vertical line have same depth hence we use weak perspective approximation. For this only a estimation of depth value for each vertical line is needed and then scaling of whole line is done. With increase in amount of depth variation in each vertical line, the distance from scene objects from camera decreases.

3. Rendering with geometry

In this section of the review paper, we'll discuss and analyze the representative techniques for rendering with known scene geometry.

Image-based rendering with geometry involves different techniques which utilize image information as well as geometric representations of a scene. For example, the application of depth maps or point clouds can help to improve the accuracy and the realism of rendered images. It is to be noted that different method of IBR uses different forms of three-dimensional information.

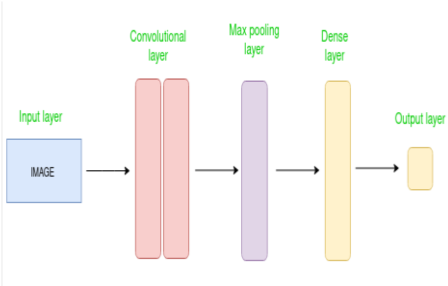
Generally, the precise three-dimensional information of the scene can be used to input the color of each pixel and allocate it to a specific fixed position in three-dimensional space. Accordingly, the color of the space point can be projected onto the place of the image with respect to the camera parameters of the novel perspective. However, this causes problems to arise because of mainly two reasons. Firstly, some pixels of the input image may not be able to find the appropriate spatial positions. Secondly, in case they find the correct spatial position, the correspondence of spatial locations to different pixels in the same image can lead to inconsistency in the rendering of the image.

Thus, in order to solve the above problems, several different algorithms have been developed to resolve the issues arising due to the special cases.

Rendering Methods with Incomplete Depth Map

Rendering methods with incomplete depth map are important for cases where a complete depth map cannot be obtained due to lack of feasibility, increased costs, etc. In brief, depth maps are images which contain information regarding the distance of the surfaces of objects in a scene from a viewpoint. These are extremely useful in applications of interior designing, accurate mapping, object detection and many more. Approaches and techniques used to render images with incomplete depth maps have been further discussed and analyzed.

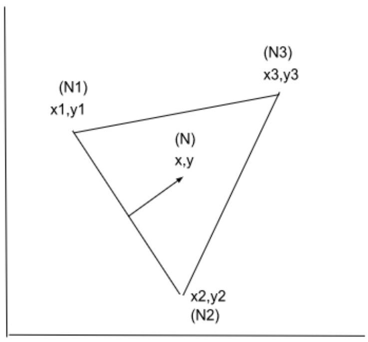
Depth inpainting methods have an objective to fill in the incomplete parts in a depth map by estimating the depths or using images and videos as data. Moreover, the use of deep neural networks has been the primary technique for filling gaps in depth maps. It is a two staged process where first, information such as edges, textures and down-sampled images are extrapolated and a refinement network is leveraged to generate a complete picture from the respective features. A particular network called Convolutional Neural Network (CNNs) have become a prominent approach for depth inpainting where the networks are trained to operate on large datasets to create a bridge between incomplete depth maps and their equivalent complete map.



A CNN consists of multiple layers as shown in the figure. The convolutional layer filters the input image and extracts the important features. It then calculates the dot product between the filters and the input which is later down sampled by the pooling layers to reduce the dimensionality of the extracted data and increase its accuracy. It reduces the computational load and advances the robustness of learned features in orientation and scale. These networks specially trained to use optimization techniques such as gradient descent which in turn gradually help to build up a hierarchy of patterns that make the networks suitable for tasks involving spatial relationships such as image recognition and image rendering.

Image-Based Rendering combining local and global geometry

Combining local and global geometry is a common approach to attain a realistic and visually appealing results of different images.



Techniques like texturing, shading and bump mapping are utilized to specify the surface details of individual objects. For example, a shading model, Phong, provides an increased accurate interpolation-based approach which can be employed for rendering shapes. It provides a better approximation of the shading of any smooth surface.

The Phong model can be employed by firstly determining the average unit normal vector at each polygon vertex. Hence, for n polygons, the average would be $\frac{\sum Ni}{|\sum Ni|}$ where i is initialized from 1 to N. Then, linearly interpolate the vertex normal over the different surfaces of the polygon.

$$N = \frac{(y - y2)}{(y1 - y2)} N1 + \frac{(y1 - y)}{(y1 - y2)} N2$$

Simply apply the illumination model along each scan line in order to determine the projected pixel intensity of the surface points.
It takes in consideration of the interaction of light such as effects of shadows, refractions and reflections. Global illumination methods such as ray tracing are used to understand and stimulate the complex relationship of light with surfaces of objects in a scene. Ray tracing explores the path of individual light rays whilst interacting with different surfaces. This technique helps to combine the local and global geometry by tracing light rays which considers the local shading effects as well as reflections and refractions. Thus, by combining local and global geometry for rendering images, individuals can enhance the realism of pictures by representing the detailed interplay of light and surfaces in a three-dimensional scene. These techniques are especially required in real-time applications such as video games.
However, ray tracing is computationally expensive when recreating complex scenes. Thus, achieving high frame rates in video games may be difficult due to the complex calculations employed in the ray tracing method. Additionally, it requires powerful hardware support including ray tracing capable GPU to handle the various complex computational demands of ray-tracing in real-time applications.

4. Rendering with Implicit Geometry

To render new views, a set of methods based on positional correspondences over a limited number of pictures is used. The word implicit in this class refers to the fact that 3D information is only calculated using standard projection computations; geometry is not directly available. Direct manipulation of these positional correspondences, which are often point characteristics, is the basis for computing new perspectives.

View Interpolation

The view interpolation approach developed by Chen and Williams can reconstruct any perspective given two input pictures with intense optical flow between them. This approach works well when two input views are close by, such that visibility ambiguity does not cause a severe difficulty. If not, flow fields must be restricted in order to stop fold overs. Furthermore, the overlapping areas of two photos may get excessively tiny when the views are far apart. The method developed by Chen and Williams performs best when all of the input images have the same look direction and the output images are limited to gaze angles of less than 90°. It might be challenging to establish flow fields for view interpolation, especially when dealing with actual pictures. It is necessary to use computer vision techniques like stereo or feature correspondence. The existing depth values may be used to produce flow fields for synthetic pictures.

View Morphing

From two input photos, Seitz and Dyer's view morphing approach reconstructs any perspective on the line linking two optical centers of the original cameras. If the camera movement associated with an intermediate view is perpendicular to the camera's viewing direction, then the intermediate view is just an exact linear combination of the two views. A pre-warp stage can be used to correct two input pictures such that the corresponding scan lines are parallel if the two input images are not parallel. Consequently, the intermediate pictures can be un-rectified using a post-warp stage. Scharstein expands this concept to include motion of a camera in an aircraft. However, he makes the assumption that the camera's parameters are known.

Joint View Triangulation

Pixel matching and visibility reasoning are the main issues with view interpolation. When source pictures are not calibrated, visibility reasoning becomes particularly challenging as no relative depth information is available to forecast occlusion in novel viewpoints. Planar patch building and quasi-dense matching are the two pre-processing phases of JVT. Interest point extraction and matching (using zero-mean normalized cross correlation) make up quasi-dense matching. An initial list of correspondences ordered by the correlation score is produced in this stage. In order to find more matches in close proximity of the point correspondence, this list is sorted starting with the highest score. To make sure there are no duplicate points in the final list, the uniqueness constraint is applied. Planar patch constructing assumes piecewise smoothness of the scene in its second stage. It's also done to get rid of any inconsistencies. The JVT algorithm uses quasi-dense matching and planar patch construction to generate Delaunay triangulations on source images, ensuring one-to-one correspondence in vertices and edges. Patches are added in raster style, labelled as matched or unmatched, and view interpolation is performed.

5. Conclusion

The field of Image-Based Rendering (IBR) has witnessed significant advancements, exploring diverse avenues in rendering techniques that leverage both geometry and non-geometry-based approaches. Rendering with geometry, categorized by methods like rendering with incomplete depth map and using local or global geometry, has long been a cornerstone in computer graphics. However, the limitations posed by complex scenes and intricate geometry have fueled the exploration of rendering without geometry, ushering in a new era of possibilities.

Non-geometry-based rendering techniques, such as light field rendering, have demonstrated their effectiveness in overcoming geometric complexities and capturing the richness of visual scenes from images. These approaches, including plenoptic function-based models and view interpolation methods, offer promising solutions for realistic rendering without relying on explicit geometric representations.

As research in IBR continues to evolve, the integration of geometry and non-geometry-based rendering techniques promises to yield more robust and versatile solutions. The synergy between these approaches opens paths for addressing the challenges linked with real-world scenes, allowing more accurate, efficient, and aesthetically pleasing rendering across various applications. In the ever-expanding landscape of computer graphics, IBR stands as a testament to the ongoing pursuit of realism and innovation in visual representation, shaping the future of how we interact with and perceive virtual environments.

Acknowledgements

We would like to express our sincere gratitude and appreciation to all those who contributed to the successful completion of this review paper on Image-Based Rendering. This collaborative effort has been a culmination of the hard work, expertise, and dedication of a diverse team of individuals.

References

1. GeeksforGeeks. (2023, December 20). Introduction to Convolution Neural Network. https://www.geeksforgeeks.org/introduction-convolution-neural-network/
2. GeeksforGeeks. (2023a, April 11). Phong Shading Computer Graphics. https://www.geeksforgeeks.org/phong-shading-computer-graphics/
3. Depth inpainting via vision transformer. (2021, October 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9585857
4. Chan, S. C. (2021). Plenoptic function. In Springer eBooks (pp. 992–998). https://doi.org/10.1007/978-3-030-63416-2_7
5. Yue, G., Hou, C., Gu, K., Zhou, T., & Zhai, G. (2019). Combining local and global measures for DIBR-Synthesized image quality evaluation. IEEE Transactions on Image Processing, 28(4), 2075–2088. https://doi.org/10.1109/tip.2018.2875913
6. Cinematic Rendering for medical imaging. (n.d.). https://www.siemens-healthineers.com/digital-health-solutions/cinematic-rendering
7. Shum, H., & Kang, S. B. (2000). Review of image-based rendering techniques. Proceedings of SPIE. https://doi.org/10.1117/12.386541
8. Chang, Y., & Wang, G. (2019). A review on image-based rendering. Virtual Reality & Intelligent Hardware, 1(1), 39–54. https://doi.org/10.3724/sp.j.2096-5796.2018.0004
9. Kang, S. B., Li, Y., Tong, X., & Shum, H. (2006). Image-Based rendering. Foundations and Trends in Computer Graphics and Vision, 2(3), 173–258. https://doi.org/10.1561/0600000012
10. Ollila, M. (2003). Simplistic dynamic image based rendering. www.academia.edu. https://www.academia.edu/91818948/Simplistic_dynamic_image_based_rendering?uc-g-sw=6190476
11. Prakash, S. (2023). Hybrid image-based rendering for free-view synthesis. www.academia.edu. https://www.academia.edu/73668563/Hybrid_Image_based_Rendering_for_Free_view_Synthesis?uc-g-sw=6190476
12. Oliveira, M. M. (2002). Image-Based Modeling and Rendering Techniques: A survey. ResearchGate. https://www.researchgate.net/publication/220162289_Image-Based_Modeling_and_Rendering_Techniques_A_Survey
13. Pereira, F., Da Silva, E. a. B., Lafruit, G., Chellappa, R., & Theodoridis, S. (2018). Plenoptic imaging: Representation and processing. In Elsevier eBooks (pp. 75–111). https://doi.org/10.1016/b978-0-12-811889-4.00002-6
14. Shum, H., & He, L. (1999). Rendering with concentric mosaics. Microsoft Research. https://doi.org/10.1145/311535.311573
15. The geometry-image representation tradeoff for rendering. (2000). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/899212
16. Anonymous. (2024, January 10). Computer Graphics: Rendering, geometry, and image Manipulation | Stanford Online. Stanford Online. https://online.stanford.edu/courses/cs248a-computer-graphics-rendering-geometry-and-image-manipulation
17. Baboud, L. (2006). Rendering Geometry with Relief Textures. https://inria.hal.science/inria-00510228