Generated Orthographic-like Projection for Panoramic Images

YU-HSUAN, HSIEH
Computer Science Department
National Taipei University
New Taipei, Taiwan

CHEN-HAO, WU
Computer Science Department
National Taipei University
New Taipei, Taiwan

CHIEH-LIN, TSAI
Computer Science Department
National Taipei University
New Taipei, Taiwan

CHI-CHEN WEI
Computer Science Department
National Taipei University
New Taipei, Taiwan

YU-LUN, LIN
Computer Science Department
National Taipei University
New Taipei, Taiwan

***Abstract*—In this paper, we present our work on processing both 2D and 3D panoramic images. We have developed a method to effectively transform these images into an orthogonal-like representation. Our approach not only preserves the visual integrity of the original panoramic images but also enhances the perceptual coherence in the orthogonal view. This advancement offers new perspectives in the visualization and analysis of panoramic scenes, making it a valuable tool for various applications in fields such as virtual reality, architectural visualization, and geographic information systems.**

***Index Terms*—Panoramic Images, 2D Image Processing, 3D Image Processing, Orthogonal Representation, Image Transformation, Computer Vision, Image Analysis**

I. Introduction

Panoramic photography has become increasingly popular for capturing wide-angle scenes, providing viewers with an immersive experience. However, a common challenge associated with panoramic images is distortion. This distortion is primarily caused by the projection methods used during image capture and stitching processes. These distortions can significantly affect the visual quality and spatial accuracy of the images, making it difficult to achieve an accurate orthogonal representation.

In modern times, there is often a need to photograph large structures or objects, such as buildings. When capturing these images, several problems arise. Using panoramic photography can result in significant distortion. Capturing multiple parallel shots and stitching them together can also be problematic due to the need for movement and the challenge of maintaining stability. Additionally, to capture the entire object, the photographer often needs to be far away from the subject, which is not always feasible.

To address these issues, we propose a method that allows for capturing images from a close distance without the need for movement, while achieving a visually pleasing orthogonal-like representation.

In this paper, we address these issues by proposing two methods to transform distorted panoramic images into orthogonal-like representations. The first method is a straightforward 2D image processing technique that applies basic corrections directly to the image. Although simple, this approach can effectively reduce some common distortions.

The second method leverages 3D image processing techniques, using depth information obtained through learning-based methods. Specifically, we employ the MiDaS pretrained model for depth estimation, as it is trained on general photographs that exhibit common perspective distortions. By utilizing MiDaS, we can accurately determine the 3D depth positions of architectural elements within the images, which is crucial for effective distortion correction. This approach involves reprojecting the images based on the depth data, allowing for more sophisticated correction and producing a more accurate orthogonal representation.

By employing these two methods, we aim to enhance the visual quality and usability of panoramic images. Our work offers valuable solutions for various applications, such as virtual reality, architectural visualization, and geographic information systems. By addressing the common problem of panoramic image distortion, we contribute to the field of image processing and open new possibilities for the utilization of panoramic imagery.

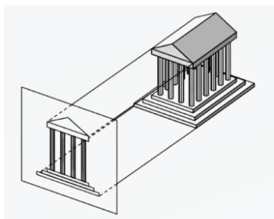
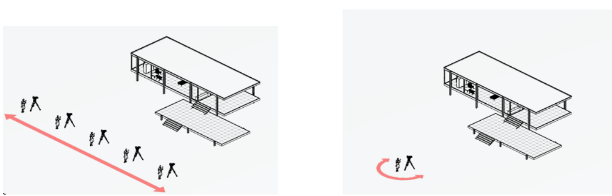


Fig. 1. Orthogonal-like



(a) (b)

Fig. 2. Parallel and Single-view

II. Related Work

Rec-tangling Panoramic Images via Warping: This work focuses on transforming panoramic images into rectangular shapes rather than irregular forms. The primary technique employed in this study is seam carving, combined with edge detection to rectify distortions caused by seam carving, such as converting bent lines back to straight ones. However, this method is not suitable for our scenario because the resulting images exhibit slight quality degradation, which significantly impacts the accuracy of depth calculations. Therefore, we opted for a cropping method to maintain high image quality.

Kaiming He, Huiwen Chang, Jian Sun, 2013. Rec-tangling Panoramic Images via Warping. A Line-Structure-Preserving Approach to Image Resizing: This paper offers a novel perspective on image resizing, particularly when it is necessary to preserve line structures within the image. By integrating content-aware techniques and line structure analysis, it manages to maintain the visual quality of images while resizing. Initially, we considered employing this content-aware approach; however, as our research progressed, we found that incorporating 3D data would be more effective. Consequently, we did not use this method.

Che-Han Chang, Yung-Yu Chuang, 2012. A Line-Structure-Preserving Approach to Image Resizing, National Taiwan University. Make3D: Learning 3D Scene Structure from a Single Still Image: This work introduces a method for learning 3D scene structure from a single still image, aiming to create precise and visually pleasing 3D models. Despite its age, it provided valuable insights into understanding 3D scenes. However, more advanced techniques and machine learning models for depth estimation have since emerged, which we utilized for better results.

Ashutosh Saxena, Min Sun, Andrew Y. Ng, 2009. Make3D: Learning 3D Scene Structure from a Single Still Image. Correction of Barrel Distortion in Fish-eye Lens Images Using Image-Based Estimation of Distortion Parameters: This paper focuses on correcting the barrel distortion in fish-eye lens images. We referred to this work because we initially considered converting our input into fish-eye images. However, we later realized that fish-eye images did not provide as much useful information as anticipated, and their depth processing was sub-optimal. Therefore, we decided to continue with our current version.

III. Method

A. Pre-processing

- Video Capture
  To further enhance the field of view (FOV) and improve the overall quality of the orthogonal representation, we incorporate video capture techniques. By capturing video footage from a fixed point, we can achieve a larger FOV compared to traditional panoramic photography methods.

The video is recorded as if capturing a typical panoramic projection but over a continuous time period. After recording, the video is segmented into individual frames based on time intervals. Each frame is then processed independently to extract the necessary information for creating the orthogonal representation.

By combining frames from different time points, we can construct a comprehensive image with a wider FOV and reduced distortion. This method leverages the temporal aspect of video to gather more data and enhance the final output's quality and accuracy.

- Stitching

Fig. 3. Stitching

For the stitching process, we utilize the built-in stitching function provided by OpenCV. This function uses feature detection and homograph estimation to align and merge multiple images into a seamless panoramic view. By leveraging OpenCV's robust algorithms, we ensure accurate feature matching and transformation, resulting in high-quality stitched images with minimal visible seams and artifacts.

B. 2D Solutions

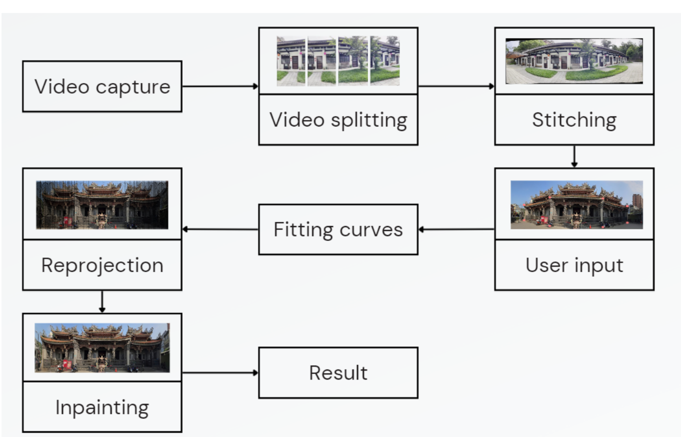


Fig. 4. 2D Method Flow

In our investigation, we observed a distinct pattern in the distortion of buildings in panoramic images. Structures that are originally rectangular often appear warped, with the top and bottom edges curving into a smile-like parabolic shape. To address this issue, we developed a method to identify these characteristic points and correct the distortion.

Fig. 5. Distort

- We will allow users to input the positions of the upper and lower boundaries of the points they identify. We hope they can input multiple points effectively and fit a curve to these points. This curve will then be used to determine the position of the main body of the building.
- Next, we will perform forward mapping on each point in the image. This methodology involves fitting two curves to the identified upper and lower boundaries and calculating the normal vectors for these curves. Subsequently, we project these normal vectors onto the horizontal line to determine the corrected positions of the points, thereby achieving horizontal alignment. Pixel processing is conducted based on their relative positions to the fitted curves. Specifically, pixels located above the upper curve are adjusted according to the normal vector of the upper curve, while pixels located below the lower curve are adjusted based on the normal vector of the lower curve. For pixels situated between the upper and lower curves, adjustments are made based on their relative distances to the two curves.

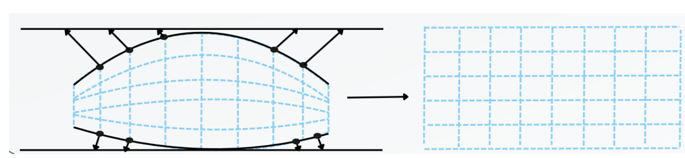


Fig. 6. Distorted Re-Projection

- Finally, we implemented an inpainting method to address the gaps that occur during the processing. These gaps tend to become larger as they move farther from the center of the image. Our approach involves identifying nearby representative values and starting the inpainting process from the outer edges of each gap, gradually filling them inward. By doing so, we effectively restore the missing regions and achieve a coherent final result.

Fig. 7. Forwarding Mapping

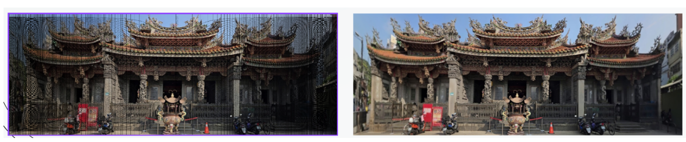


Fig. 8. Inpaniting

C. 3D Solutions

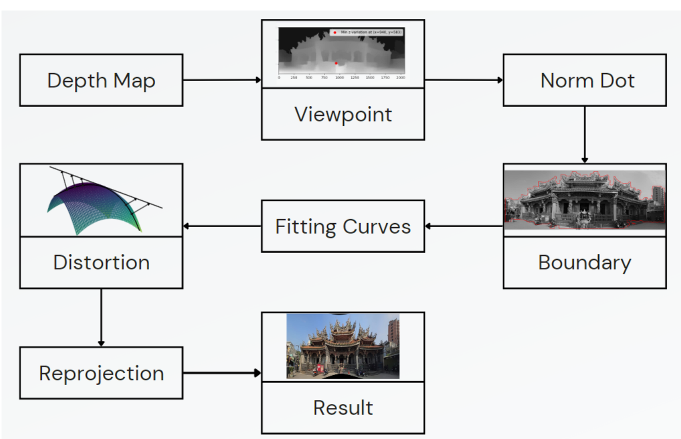


Fig. 9. 3D Method Flow

Our 3D approach leverages depth prediction using the MiDaS pre-trained model to enhance the accuracy and quality of our panoramic image processing. This model generates precise depth maps from single images, which are crucial for our subsequent steps.

Initially, we identify the viewpoint of the image by analyzing the depth map. This step is essential to determine the main viewing direction and perspective. Following this, we locate the boundaries within the image, which are the significant edges and features susceptible to distortion.

With the viewpoint and boundaries identified, we calculate the potential distortion planes. By fitting curves to these boundaries, we can model the distortion patterns present in the image. This modeling enables us to understand and address the distortions more effectively.

We perform re-projection to correct these distortions. By applying the calculated distortion planes, we realign the image to its original, undistorted form. This re projection ensures that the final image maintains the structural integrity and visual accuracy of the original scene.

- Depth Estimation with MiDaS: We utilize the MiDaS depth estimation model because it is trained on general photographs that exhibit common perspective distortions. Given that our post-processed and stitched images share similar perspective distortions, MiDaS is particularly suitable for accurately determining the 3D depth positions of the buildings of interest. By using MiDaS, we can effectively gauge the depth information of the architectural elements within the image, laying the foundation for subsequent steps.

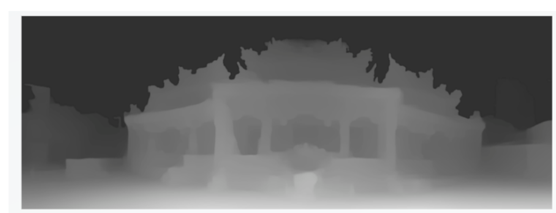


Fig. 10. Depth Map

- Depth Map Processing: Next, we perform calculations on the depth map data. We begin by identifying the position of buildings within the scene. Considering the depth map as a 3D surface, we determine the projection direction of the surface's normal vectors. This information allows us to distinguish between buildings, the sky, and the ground effectively. Additionally, we analyze the local surfaces across the entire depth map to determine their inclination. The areas near the viewpoint exhibit the least inclination, enabling us to pinpoint the viewpoint as the location with the minimal overall distortion in the image.

Fig. 11. Building area

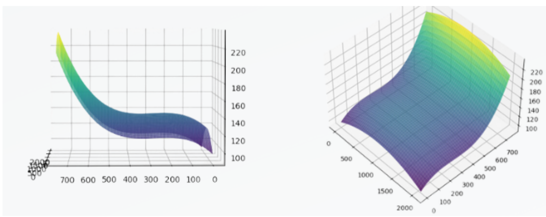


Fig. 12. Fitting plane

- Distorted Image Plane: We estimate the distorted image plane by sampling points within the building's area. During this step, we remove sections with abrupt depth changes to avoid inaccuracies. We fit a surface based on the viewpoint's curvature to model the entire surface. This involves adjusting the image for different segments and removing outliers. By doing so, we can accurately estimate the distortion of the image plane for each segment.

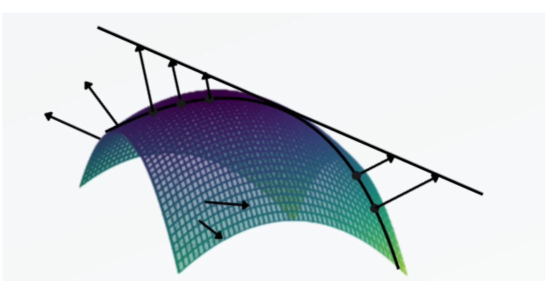


Fig. 13. 3D Model Re-projection

Finally, based on the estimated image plane, we flatten the distorted surfaces. We achieve this by projecting the plane onto the depth position of the viewpoint and using normal vectors for the projection. By comparing the depth map data with the expected normal vectors, we derive the final corrected image.

IV. Results

A. 2D Solutions

Fig. 14. 2D Result_1

Fig. 15. 2D Result_2

Fig. 16. 2D Result_3

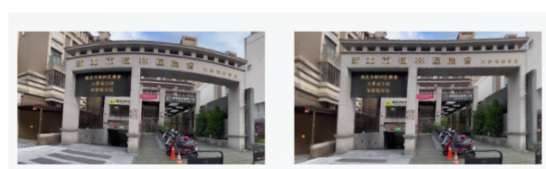


Fig. 17. 2D Result_4

In our 2D method, we compared the results of our algorithm with the original images. We successfully transformed the distorted parts of the images, caused by stitching, into an orthogonal-like perspective. Vertical elements, such as columns, remained correctly aligned without any misplacement. Our results demonstrate that the algorithm effectively achieves the intended outcomes.

By addressing the distortions, our method approximates an orthogonal projection. However, it is important to note that areas of the building not captured due to the original perspective limitations cannot be reconstructed with new data through our algorithm. Therefore, while our method significantly improves the visual representation by reducing distortion, it inherently remains an approximation of true orthogonal projection due to the inherent limitations of the input data.

B. 3D Solutions

Fig. 18. 3D Result

By analyzing the distortion of the image plane, we can achieve an accurate representation of the building's proportions. However, due to the limitations of depth data being relative and often suffering from occlusion issues, while it can improve the performance over 2D methods, it cannot consistently provide stable outputs.

References

[1] M. Tancik et al., "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis," Available: https://www.matthewtancik.com/nerf.

[2] "Neuralangelo," NVIDIA Research, Available: https://research.nvidia.com/labs/dir/neuralangelo/.

[3] "Make3D: Single Image to 3D," Cornell University, Available: http://make3d.cs.cornell.edu/data.html.

[4] "Structure from Motion (SfM)," Available: https://arxiv.org/abs/1906.03539.

[5] S. Chang et al., "Content-Aware Image Resizing," Available: https://www.csie.ntu.edu.tw/~cyy/publications/papers/Chang2012ALS.pdf.

[6] S. Iizuka, "Image Inpainting for Irregular Holes Using Partial Convolutions," Available: https://github.com/satoshiiizuka/siggraph2017_inpainting.

[7] "360MonoDepth: Monocular Depth Estimation from a Single 360° Image," Available: https://manurare.github.io/360monodepth/.

[8] T. Xue et al., "Learning to Calibrate Straight Lines for Fisheye Image Rectification," CVPR 2019, Available: https://openaccess.thecvf.com/content_CVPR_2019/papers/Xue_Learning_to_Calibrate_Straight_Lines_for_Fisheye_Image_Rectification_CVPR_2019_paper.pdf.

[9] K. He et al., "Rectangling Panoramic Images via Warping," SIGGRAPH 2013, Available: https://kaiminghe.github.io/publications/sig13pano.pdf.

[10] "COLMAP: Structure-from-Motion and Multi-View Stereo," Available: https://github.com/colmap/colmap.

[11] S. L. Ramasinghe et al., "Adabins: Depth Estimation using Adaptive Bins," Available: https://github.com/shariqfarooq123/AdaBins/blob/main/README.md.

[12] "Content-Aware Retargeting," Available: https://github.com/tobygameac/Content-Aware-Retargeting.

[13] "Context-Aware Saliency Detection," Available: https://github.com/Somefive/context-aware-saliency-detection.

[14] "Image Super Resolution," Available: https://github.com/idealo/image-super-resolution?tab=readme-ov-file.

[15] "LaMa: Image Inpainting," Available: https://github.com/advimman/lama.