Realistic Compositing of Planar Posters into Live-Action Video

Alex Li (Juntao)*
Simon Fraser University

Zhaoyue Yuan*
Simon Fraser University

Zeyou Peng*
Simon Fraser University

S. Mahdi H. Miangoleh
Simon Fraser University

Yağız Aksoy
Simon Fraser University

Figure 1: Input poster, first frame of input video and the first frame of result video

ACM Reference Format:
Alex Li (Juntao), Zhaoyue Yuan, Zeyou Peng, S. Mahdi H. Miangoleh, and Yağız Aksoy. 2024. Realistic Compositing of Planar Posters into Live-Action Video. In *Proceedings of ACM Conference (Conference'17).* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The original idea starts with a question: can we replace the poster on the building with our own poster? It leads to several more interesting questions: Will companies try fitting their advertisement on the wall digitally before actually implementing it? What if customers want to see how the poster act in different real-live conditions? So we formulate our project as: Realistic compositing of Planar posters into Live-action Video.

Our project aims to explore and enhance the realm of tracking and compositing by seamlessly integrating planar posters into dynamic live-action sequences. The ultimate goal is to get an output where the input poster maintains a consistent presence, accurately replace and track the designated quadrilateral planar region within the video, and match the changes in perspective, location, and lighting/shading conditions. This realistic compositing strategy can be used in areas such as advertising demos, film production, social media filters or augmented reality, where planar digital contents are coexisting with live-action elements. An example of the input poster and first frame of the input video is shown in figure 1 on the left, and our first frame of the composed output video is on the right.

The whole project can be divided into two parts: tracking and shading. For the tracking portion, our approach includes both the traditional perspective tracking methods and the surface normal-based tracking method, depending on the actual feature in the input video. These methods provides a robust and accurate way to follow the planar region's orientation and position as it moves or changes in the 3D space.

Shading and lighting play main roles in ensuring that the composited poster's integration appears natural and fusion in the video. We employ shading techniques to match the poster's lighting to the dynamic conditions of the video, preserving the scene's inherent shadows, highlights, and texture nuances, which advertisement or movies have to achieve. Surface normal is estimated by the *Omnidata* model, intrinsic composition, re-scaling, lighting coefficients are applied to the poster and the final albedo harmonization and compositing organize these effects. The result is a fusion of the poster with the input video, where the poster inherits the video's ambient and direct lighting conditions, reacts to environmental changes, and consistently belongs to the space it occupies.

2 PIPELINE, METHOD, FORMULATION

The pipeline shown in Figure 2 starts with the input of a poster image and a mp4 video that contains a surface (so that the poster can be stick on). The first step is to identify the type of the video, we divided it into two separate cases.

The first case is the perspective case, if the planar surface in the video contains a clear quadrilateral region and it stays in the boarder

*Denotes equal contribution.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

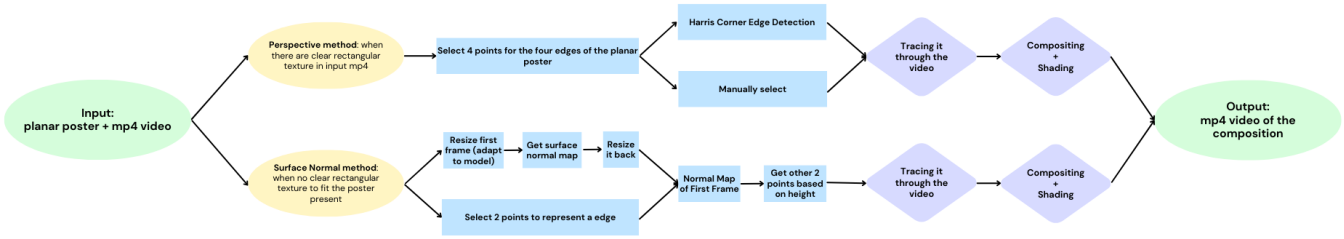

Figure 2: Pipeline throughout the project, separated into two cases

of the video constantly, we can apply the perspective method. The first step is to locate the quadrilateral region in the first frame of the input video. We are selecting the four points manually by clicking the four corners of the quadrilateral, if you do not want to manually select the corners, Harris corner detection is another choice, during the point selecting process, several additional tools are provided such as a magnifier, distance and area calculator, these four points are later used to create a perspective transformation matrix. After this, the next step is to set parameters for the *Lucas-Kanade* optical flow method. Optical flow is used here to track the motion of the selected points between frames. We then enters the main loop which reads each frame of the video in sequence. Each frame is converted to grayscale and *calcOpticalFlowPyrLK* in Cv2 computes the new points of the selected region of the poster based on the changes in different frames. These new points will be tracked and proceeds with the transformation and overlay. Furthermore, when the region is confirmed in the inter-loop of each frame, the poster is warped using perspective transformation matrix M to fit the quadrilateral defined by the corners, plus the OpenCv2 function *warpPerspective*. The mask is then created as the same size as the frame, and a filled polygon is drawn using corner points, the current frame and the warped poster are combined using bitwise operations. By keep updating the to corner points for the next iteration, we get our output video with the poster stays in the quadrilateral region realistically.

The second case is the surface normal case, if the planar surface in the video is "infinite", which a quadrilateral region cannot be select or detected directly, we are using the normal method. The first step involves extracting the first frame from the video and resizing it to a size suitable for the model to process, then the we get the normal map from *Omnidata* such as Figure 3. The normal map is then resized back, to match the resolution of the video frame. We can now select 2 points based on the normal map we get, to define one edge of our quadrilateral region, and input the height of the poster. By evaluating the edge, the height perpendicular to it, and the normal values in this region, we can define the position of the poster in the first frame of the video. After this, we apply similar steps as the perspective method does, for looping and tracking the poster's position by optical flow throughout the video.

In the mean time of looping through the frames for both perspective and normal method, we apply shading. By getting the combined image of the poster and first frame of the video, we apply the following pipeline of methods to this combined image. Firstly, we get the surface normal map from *Omnidata* and use it to determine how light interacts with the surfaces. Then Intrinsic Decomposition is

Figure 3: Example of the normal visualization of the image.

applied to separate the image into its shading (captures the effect of lighting on the scene) and reluctance components. From this, we then get the lighting coefficients which characterize the light's behavior in the scene and use it to ensure that the new shading on the foreground object matches the background lighting. Besides that, albedo harmonization is applied to adjust the composited object's albedo, so that its colors are consistent with the lighting of the scene. The final image is composited by blending the re-shaded poster with the background while respecting the original shading, the normal, and the depth of the scene. With this final image, we can get the shaded poster (based on the shadings of the first frame) and use it as the input poster image throughout the frames loop (since the lighting in the scene of our video is assumed to be not changing).

By combining the points selection, shading and tracing, we get our final output image.

3 RESULTS/EVALUATION/LIMITATION

Several results are shown below in Figure 4, since our output is a live-action, we just display the poster, the original first frame and the first frame of the output for comparison and visualization.

Results for most of our examples are nice and smooth, but it's completeness and perfectness still depend on the input video's texture. For edge cases like the texture of soft materials such as clothes or flags (which might not to be completely planar), or the texture of glasses (might be transparent and show details through it, or consists of complex shadings), it is hard to find a nice normal map for the corner selecting and the tracing process.

4 CONCLUSION/APPLICATION

Our project uses advanced tools from computer vision to perfectly blend posters into live-action video scenes. By calculating surface normals or perspective, and tracking movement, we ensure that the poster's position and orientation throughout the video. We also used intrinsic decomposition and other shading tools to adjust the lighting on the poster so it matches the scene perfectly. These techniques help maintain the scene's natural look and improve how the poster fits visually. Our project can be used for advertising, film and movie editing, AR and VR, or further develop for some more straightforward and effective content creation and composition in the future.

REFERENCES

Shariq Farooq Bhat, Reiner Birkl, Diana Wofk, Peter Wonka, and Matthias Müller. 2023. Zoedepth: Zero-shot transfer by combining relative and metric depth. *arXiv preprint arXiv:2302.12288* (2023).

Chris Careaga and Yağız Aksoy. 2023. Intrinsic Image Decomposition via Ordinal Shading. *ACM Transactions on Graphics* 43, 1 (2023), 1–24.

Chris Careaga, S Mahdi H Miangoleh, and Yağız Aksoy. 2023. Intrinsic Harmonization for Illumination-Aware Compositing. *arXiv preprint arXiv:2312.03698* (2023).

Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. 2021. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10786–10796.

S Mahdi H Miangoleh, Sebastian Dille, Long Mai, Sylvain Paris, and Yagiz Aksoy. 2021. Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9685–9694.

[Bhat et al. 2023] [Careaga and Aksoy 2023] [Eftekhar et al. 2021] [Miangoleh et al. 2021] [Careaga et al. 2023]


SFU
DEPARTMENT NAME
WE
LIKE
THE WAY
YOU
THINK
UBC
Robson Square
800 Robson Street
i'm lovin'it
don't stare too long
You'll miss it
Egg McMuffin'@McDonald's

Figure 4: Five pairs of our demo visualization, contains the original poster, the first frame of the original video, the first frame of the result video (since video cannot be shown here, check our demo videos!)