BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY
Excel@FIT 2025
http://excel.fit.vutbr.cz

Texture reconstruction from multiple views

Bc. David Kedra*

Abstract

High-resolution acquisition of large, flat surface textures, such as maps, posters, paintings, or historical documents, is often limited by camera resolution, lighting conditions, or physical constraints that prevent obtaining a single high-quality photo. This work addresses the problem of reconstructing a seamless texture from multiple views captured under varying conditions and viewpoints. A pipeline is proposed that performs both hard and soft alignment of input views to a low-resolution reference image, followed by data tiling and the reconstruction (enhancement) of a reference image, which represents the entire surface, using a trained deep learning model. The results demonstrate that the output textures are significantly cleaner and sharper than the original reference images, successfully restoring fine details, suppressing visual inconsistencies, and eliminating seams and lighting imbalances. This approach can be applied in cultural heritage digitization, archival imaging, and visual documentation where preserving both fidelity and detail is crucial.

*xkedra00@vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Capturing clean, high-resolution textures of large flat surfaces (maps, posters, artworks) is difficult due to camera limitations, lighting, and physical constraints. A single image is rarely sufficient. Multiple overlapping views may be taken, but these suffer from seams, blur, and exposure differences when naively stitched.

The task is to reconstruct a sharp, seamless, and visually consistent texture image from several misaligned and variably lit photographs. The output should be well-aligned, preserve details, remove seams, and correct lighting, while remaining artifact-free and faithful to the original. Its quality can be evaluated by measuring the similarity to the ground truth (GT) texture using metrics like PSNR or SSIM.

Related methods typically focus on 3D geometry-aware super-resolution using known camera poses and depth maps (e.g., [1]) or by projecting observed images onto a 3D surface mesh (e.g., [2]). In contrast, my method relies on homography and optical flow alignment, which is closer to the approach presented in [3]. This paper also highlights the constraints of image stitching techniques, such as geometry misalignment and inconsistent appearance across input images. Research has shown that existing methods do not directly deal with the specific setup for multi-view texture reconstruction of flat geometric surfaces.

The proposed solution first aligns input images to a reference image using homography calculations and SEA-RAFT [4] robust optical flow. A convolutional neural network, based on encoder-decoder architectures such as U-Net, UNet++, and a multi-view adaptation of the EDSR model [5], is then used to reconstruct a clean texture from the aligned data, processed patch by patch. For high-quality feature extraction, various encoders are explored, including ResNet [6] and Mix Vision Transformer (MiT) [7] backbones. The patch-based technique ensures high-resolution processing and allows having outputs in any required large resolution. The majority of the training data is synthetic, generated in Blender and aligned using extracted UV maps, apart from the described alignment of real data.

Contributions:

- A working pipeline for seamless reference image (texture) reconstruction from multiple images, ideally covering the entire surface area, captured under varying conditions.
- Introduced a hybrid alignment approach combining geometric homography and flow-based SEA-RAFT methods to accurately align input views to a low-resolution reference.
- Designed and trained a deep learning model on generated synthetic images and real photos to

effectively enhance fine details in a reference image, remove seams, and correct lighting inconsistencies by leveraging the complementary information from input views.

2. Method overview

This section describes the used approach for reconstructing high-resolution texture images by enhancing reference images using other views, organized into key phases of the pipeline.

The input data are sets of Q images taken from different viewpoints above the surface. Individual images vary in lighting, view area, and distance. A dataset of 150 synthetic scenes, each containing $Q = 30$ images (3840 × 2160), was generated in Blender. Each set has a uniquely deformed surface and a high-resolution (averaging 10k pixels per side) texture mapped on it. Additionally, 30 real photo sets were taken with a smartphone (resolution 4624 × 3468) to ensure the model can handle diverse distortions. In total, 180 sets were used for training and validation.

Reference views are generated through a realistic texture degradation process. This helps the reconstruction to guide towards an image perfectly aligned with a texture and to closely match its reference colors, which allows evaluating their similarity. Moreover, it simulates real-world imperfections, such as Gaussian blur, motion blur, chromatic aberration, Bayer pattern, noise, and JPEG artifacts, which are caused by lenses, sensors, or post-processing systems.

The core of the reconstruction process involves aligning all views to a reference image. Synthetic data are aligned using UV maps from Blender, while real images are aligned using homography and optical flow via the SEA-RAFT(L) model. To manage memory consumption and keep the original pixel-level resolution, images are processed in 1024 × 1024 patches.

Aligned images are split into 512 × 512 patches to reduce computational load, improve efficiency, increase the training dataset size, and avoid the need to downsample images. Only non-empty patches are saved.

Each data sample consists of one texture patch, one reference image patch, and N other view patches selected from aligned views at the specific patch position. N is set to 5. The model is trained using batched samples and several encoder-decoder architectures, including multi-view model inspired by single-image EDSR, ResNet34+U-Net, MiT+U-Net, and ResNet101+UNet++. Current evaluation was performed on two sets: *Scene A*, which contains photos of a map printed on A2-sized paper, and *Scene B*, which contains photos of a poster measuring 59 cm × 46 cm. Both sets were captured using a Samsung Galaxy A52s 5G device. PSNR measures pixel intensity differences, while SSIM evaluates structural and perceptual image quality.

3. Conclusions

This work introduces a method for reconstructing high-resolution textures from multiple views, handling alignment and lighting inconsistencies. Models trained on both synthetic and real-world data successfully enhance reference images and recover details. The approach can be applied in cultural heritage preservation or used as a mobile camera-based scanner. Future work may focus on improving model performance, building better architectures, or refining the selection of the most informative patches for reconstruction.

Acknowledgements

I would like to thank my supervisor Ing. Michal Hradiš, Ph.D. for his help, support, and valuable advice throughout this project.

References

[1] Ri Cheng, Yuqi Sun, Bo Yan, Weimin Tan, and Chenxi Ma. Geometry-aware reference synthesis for multi-view image super-resolution, 2022.

[2] Audrey Richard, Ian Cherabier, Martin R. Oswald, Vagia Tsiminaki, Marc Pollefeys, and Konrad Schindler. Learned multi-view texture super-resolution, 2020.

[3] Jiaqin Jiang, Li Li, Bin Tan, Lunhao Duan, and Jian Yao. A multi-view references image super-resolution framework for generating the large-fov and high-resolution image. *Journal of Visual Communication and Image Representation*, 100:104123, 2024.

[4] Yihan Wang, Lahav Lipson, and Jia Deng. Sea-raft: Simple, efficient, accurate raft for optical flow, 2024.

[5] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution, 2017.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[7] Enze Xie, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. Segformer: Simple and efficient design for semantic segmentation with transformers, 2021.