Parallax-aware Image Stitching based on Homographic Decomposition

Simon Seibt[1], Michael Arold[1], Bartosz von Rymon Lipinski[1], Uwe Wienkopf[2], and Marc Erich Latoschik[3]

[1] Game Tech Lab, Faculty of Computer Science, Nuremberg Institute of Technology, Nuremberg, Germany
simon.seibt@th-nuernberg.de
[2] Institute for Applied Computer Science, Nuremberg Institute of Technology, Nuremberg, Germany
[3] Human-Computer Interaction Group, Institute of Computer Science, University of Wuerzburg, Wuerzburg, Germany

Abstract. Image stitching plays a crucial role for various computer vision applications, like panoramic photography, video production, medical imaging and satellite imagery. It makes it possible to align two images captured at different views onto a single image with a wider field of view. However, for 3D scenes with high depth complexity and images captured from two different positions, the resulting image pair may exhibit significant parallaxes. Stitching images with multiple or large apparent motion shifts remains a challenging task, and existing methods often fail in such cases. In this paper, a novel image stitching pipeline is introduced, addressing the aforementioned challenge: First, iterative dense feature matching is performed, which results in a multi-homography decomposition. Then, this output is used to compute a per-pixel multidimensional weight map of the estimated homographies for image alignment via weighted warping. Additionally, the homographic image space decomposition is exploited using combinatorial analysis to identify parallaxes, resulting in a parallax-aware overlapping region: Parallax-free overlapping areas only require weighted warping and blending. For parallax areas, these operations are omitted to avoid ghosting artifacts. Instead, histogram- and mask-based color mapping is performed to ensure visual color consistency. The presented experiments demonstrate that the proposed method provides superior results regarding precision and handling of parallaxes.

Keywords: Image Stitching · Parallaxes · Feature Matching.

1 Introduction

Image stitching is an important technique in computer vision that combines two or more images with overlapping areas into a single high-resolution, wide-field image. It is used in various media applications, including generation of panoramic images for virtual tours, computation of high-resolution photo mosaics from multiple satellite images, and in medical imaging procedures [18,19,27,33].

Typically, conventional image stitching methods are executed in three stages [25]: (1) Extraction and matching of features between an image pair, (2) Estimation of a

global homography based on the feature matches, and (3) Perspective transformation and blending of the target image onto the reference image. One of the most challenging steps here is image warping using one global estimated homography. A homographic relation can be used to describe feature correspondences for points, which lie on the same plane in 3D space [7]. However, if the captured 3D scene is not planar, i.e. including foreground objects at different scene depths, and the camera baseline between an image pair is large, then parallaxes can be observed [11]. In such cases, stitching results based on planar transformation models, such as the aforementioned global homography approach, often exhibit visual artifacts, like distortions and ghosting.

To address the issue of parallax artifacts in stitching of a single image pair, various warping methods were developed previously. For example, some approaches divide an image into regular pixel cells, which are then warped using different geometric models, such as [5,11,12,29]. In order to additionally optimize warping, energy minimization frameworks were utilized [12,31]. Also, local alignment techniques were introduced to register specific regions of the image while hiding artifacts in misaligned regions through seam-cutting methods, such as [6,16,28,30]. However, images with high depth complexity and large parallaxes still represent a challenge as neighboring pixels in the reference image may not have corresponding adjacent pixels in the target image.

In this paper, a new image stitching pipeline is introduced that addresses the aforementioned problem. The main contribution of this work comprises an accurate and robust stitching method for image pairs, which exhibit complex structures and multiple depth layers, while avoiding visual artifacts caused by parallaxes: The presented solution utilizes results from the work of Seibt et al. [24], *dense feature matching* (DFM), for detection of robust and accurate feature correspondences between image pairs. DFM targets at images of real scenes with significant depth complexities, offering high precision and recall values.

The corresponding pipeline is based on a *homographic decomposition* of the image space, providing the following advantages for image stitching in terms of alignment accuracy and parallax handling: (a) Conventional feature matching is extended to an iterative rematching process. The search for correct feature matches is thus re-executed per iteration with an individually estimated homographic transformation. By utilizing the resulting homographic decomposition, DFM positionally refines matching feature points in the target image and extrapolates additional points that could not be matched using standard “one-shot matching“ methods. The result of the rematching process is a precise set of dense feature matches, each associated with a homography matrix, recovering multiple tangent planes in a 3D scene. (b) Additionally, Delaunay triangulation of the dense feature point set is used to determine the overlapping region of an image pair. (c) Through combinatorial analysis of the underlying per-vertex homography configurations in the overlapping region, ”parallax awareness“ is handled and “critical image areas” (i.e. containing occlusions) are identified by so-called *inhomogeneous* triangles. The overlapping sub-region without visible parallaxes is classified as *homogeneous*.

Subsequently, these differently classified regions are processed in the proposed stitching pipeline as follows: (1) For both, the homogeneous (overlapping) region and non-overlapping region of the target image, a multidimensional weight map is computed. It is used to control the contributions of the multi-homography decomposition re-

sults in order to achieve an accurate and robust perspective transformation for warping. The concept of utilizing a stitching weight map was inspired by Gao at el. [5]: In their approach, two-dimensional weights were used for dual homography warping. (2) For the inhomogeneous (overlapping) region, weight-based transformations and blending would result in ghosting artifacts due to parallaxes. Instead, a histogram- and mask-based color mapping is executed. (3) After perspective transformation, non-overlapping sub-areas in the target image may result in undesirable intersections with overlapping sub-areas in the reference image. These intersecting sub-areas are cut out in order to prevent additional visual artifacts. (4) In the last step, the perspective transformation on the cropped target image is performed utilizing the weight map. Additionally, the overlapping homogeneous region is blended to obtain smooth visual transitions.

2 Related Work

Over the years, various stitching methods were proposed, which can be broadly categorized into three types: Adaptive warping methods, shape-preserving warping methods and seam-based methods. For the adaptive warping methods, Gao et al. [5] introduced a dual homography method that can only handle scenes with a distant and a ground plane. The method adaptively blends the homographies estimated for the two planes according to positions of clustered feature points. Zaragoza et al. [29] estimated optimal homographies for each regular grid cell using moving direct linear transformations (MDLT). They assigned higher weight to feature points closer to the target cell based on alignment error. Liu and Chin [17] introduced an extension of [29] by inserting appropriate point correspondences in automatically identified misaligned regions to improve the underlying warp. Zhang et al. [31] minimized distortions in warped images by incorporating a scale-preserving term and a line-preserving term. Li et al. [12] proposed an analytical warping function based on thinplate splines with radial basis functions to approximate the projection bias. Lee and Sim [10] described an epipolar geometry-based video stitching method. It is suitable for handling large parallaxes, requiring temporal motion information of foreground objects. Lee and Sim [11] also proposed a solution that uses warping of residual vectors to distinguish matching features from different depth planes, resulting in more naturally stitched images. Liao and Li [14] introduced a method that simultaneously emphasizes different features under single-perspective warps, including alignment, distortion, and saliency. Chen et al. [4] proposed a method for complex image pairs with moving pedestrians. It includes a structure preservation module based on grid constraints and a composite ghost removal module using YOLOv4, ensuring unique preservation of each pedestrian after stitching. Jia et al. [9] introduced the characteristic number to match co-planar local sub-regions, and additionally considered global co-linear structures using a separate objective function. While [5] is only capable of handling scenes with two planar regions, methods such as [12] [29] [31] can warp backgrounds containing multiple planar regions. However, these methods often rely on the assumption of continuous scene depths with minimal parallaxes. That poses a challenge for aligning foreground objects with large parallaxes, which have abrupt depth changes relative to the background.

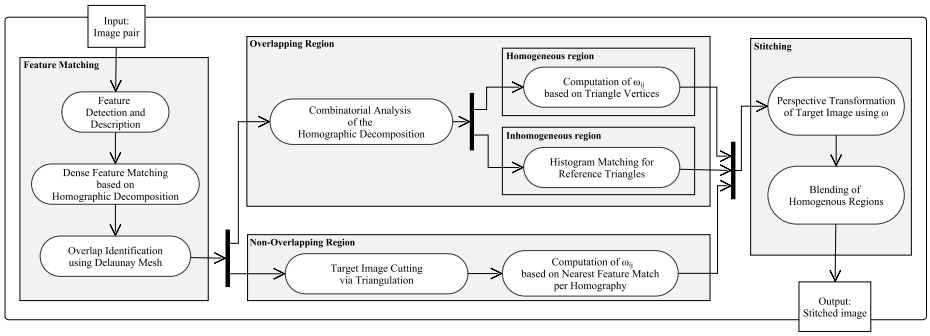


Fig. 1: UML-based activity diagram of the presented image stitching pipeline: main stages in grey, sub-activities in white color.

Shape-preserving warping-based methods are commonly introduced to mitigate perspective distortions in non-overlapping regions of an image pair. Chang et al. [2] presented a warping method that combines a projective and a similarity transformation. It smoothly extrapolates the projective transformation of overlapping regions to non-overlapping regions, gradually transitioning from projection to similarity across the image. Chen et al. [3] estimated the scale and rotation for each image and designed an objective function for warping based on a global similarity prior. Lin et al. [15] proposed a homography linearization method that also smoothly extrapolates warps from overlapping to non-overlapping regions. Li et al. [13] described a quasi-homography warp to solve the line bending problem by linear scaling of the horizontal homography component.

Seam-based methods produce visually appealing image stitching results, but typically lack geometric precision due to local alignment: Gao et al. [6] introduced a seam-cutting loss to measure the discontinuity between warped target and reference images. They estimated multiple homographies using RANSAC and selected the optimal one with the minimum seam-cutting loss. Lin et al. [16] improved stitching performance using iterative warp and seam estimation. Zhang et al. [30] used local homographies to align specific image regions and applied content preserving warping for further refinement. Misalignment artifacts were hidden via seam cutting. Xue et al. [28] introduced a method that integrates point-line features as alignment primitives. They used a pixel difference evaluation model to iteratively compute and update their smoothing term to find the most suitable seam.

Previous deep-learning-based stitching methods [20] [32] [21] face significant challenges, like limitation to mostly synthetic training datasets, unsatisfactory performance for real-world images and issues with preservation of the original image resolutions.

3 Parallax-aware Image Stitching Pipeline

In the following sub-sections the four main stages of the presented stitching pipeline are described, including all corresponding processing steps (overview in Fig. 1).

3.1 Dense Feature Matching

In the first pipeline stage, DFM is executed to generate a precise and dense set of feature matches between an image pair. This makes it possible to identify the overlapping pixel region and to obtain a homographic decomposition of the image space: DFM performs *iterative rematching* to give mismatched feature points a further chance to be considered in subsequent processing steps, resulting in potentially larger matching sets and increased recall values. During rematching an individual homography is estimated per iteration, searching for feature matches that correspond to the same plane in 3D space. In practice, a single homography can span multiple surfaces. So, feature points are clustered per iteration to improve the homography estimation by recalculating it for each cluster. The result of rematching is the *homographic decomposition*, i.e. a set of feature point pair clusters, each associated with a distinct homography matrix (illustration in Fig. 2 (c)).

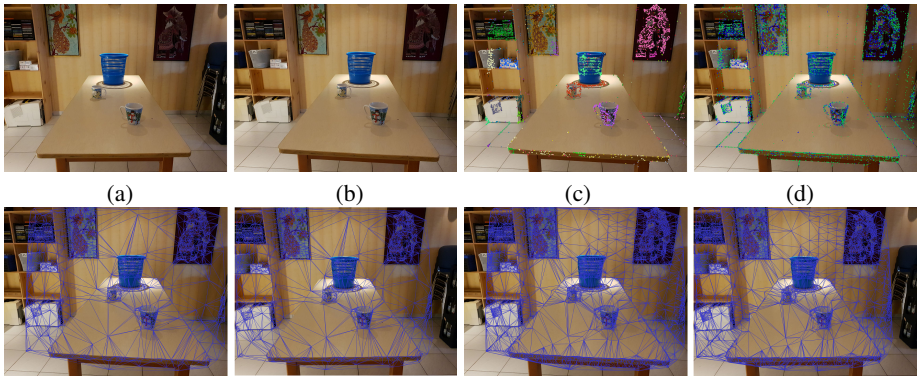
Additionally, rematching involves incremental Delaunay triangulation of the reference feature set and a mapping of the resulting mesh to the target image. As described in [24] further DFM processing steps utilize this mesh as follows: *Delaunay outlier detection* removes false positive matches by detecting reference-to-target triangle edge intersections, which are often caused by repeated patterns. *Focused matching* simulates "visual focusing" by executing local rematching per Delaunay triangle, resulting in the detection of additional detailed feature points. Additionally, matching accuracy is increased using positional *feature refinement* by taking advantage of neighboring homography candidates per feature point. Finally, *feature extrapolation* is used to detect additional feature points by utilizing multiple local homography candidates per Delaunay triangle.

The overall DFM result is a *refined homographic decomposition* (see Fig. 2 (d)). The triangulation results are illustrated in Fig. 2 (e-h). Fig. 2 (e) and 2 (f) depict the outcome of the initial iterative rematching. Fig. 2 (g) and 2 (h) show densified meshes resulting from execution of further DFM processing steps, mentioned previously. As can be seen, the meshes have undergone substantial refinement, resulting in a better approximation of geometric structures in the scene.

3.2 Overlapping Region

The next step of the stitching pipeline builds on the final DFM Delaunay triangulation of the reference image feature set and the corresponding mapping to the target image. Pixels, which are located within the triangle mesh and its mapping, respectively, define the overlapping region for stitching. Further processing of this region requires segmentation of the corresponding Delaunay mesh into *homogeneous* and *inhomogeneous* triangles: Homogeneity – in context of a feature point mesh – is defined for a triangle t with vertices $v_i \in V_t$ as follows [24]: Let H be the set of homography matrices resulting from homographic decomposition, $h_i \in H$ the (initially) associated homography with feature point vertex v_i. Then, t is homogeneous, if:

$$\exists h_{hom} \in H, \forall i \in \{1,2,3\} : h_{hom} * v_i \approx h_i * v_i. \tag{1}$$



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 2: (a,b) Input image pair; (c) Clustered features in the reference image after DFM's iterative rematching. Each color represents one cluster with an individual homography; (d) Matched features after further DFM processing, incl. focused matching, refinement and extrapolation: initially matched source features in blue (cf. image c), further features in green; (e) Triangulation of matched features after initial iterative rematching; (f) Mapping of the mesh in image e to target image; (g,h) Densified triangulations using the final feature matching set after further DFM processing.

Homographic decompositions commonly have high variances in homography-to-vertex associations with initially one homography per vertex. The *homogenization* process is executed in order to relax the detection of inhomogeneous triangles for feature detection. In this work, it is used to improve segmentation of the overlapping region for stitching: The homogenization uses the connectivity information of the Delaunay feature set mesh to search for local equivalently transforming homography matrices. Each feature point is successively transformed from the reference image to the target image using homographies of neighboring features. Each time, the reprojection error is calculated and compared against the threshold parameter of the RANSAC algorithm (used during rematching). If the reprojection error of a neighboring homography is smaller, then it is (additionally) assigned to the current feature point. Notice that in the resulting (homogenized) *multi-homographic decomposition* a feature point can be associated with multiple locally equivalent transforming matrices. Consequently, an ablation of the homogenization process would prevent robust detection of inhomogeneous triangles (parallax areas). This would inevitably result in ghosting artifacts after final blending. The homogenization process is shown in Algorithm 1 and illustrated in Fig. 3 (a).

The next pipeline step aims at a seamless overlapping region for stitching. Hence, the multi-homographic decomposition is used to compute a multidimensional warping weight map ω for the homogeneous sub-region: For each pixel in the reference image, a point-triangle intersection test is executed based on the Delaunay mesh. On a hit, the three vertices of the corresponding enclosing triangle are selected and the Euclidean distance d_i to each vertex is calculated. Then, each initially per-vertex associated homography H_i for a pixel position (x, y) is weighted using the following equation:

(a) (b) (c) (d)

Fig. 3: (a) Segmented overlapping region of a reference image with homogeneous (blue) and inhomogeneous triangles (red); (b) Visualization of a multidimensional per-pixel weight map for an overlapping homogeneous region of a target image. Each homography is represented by a unique color. The pixel color in the overlapping homogeneous region is a combination of colors according to the weighted homographies; (c) Stitching result of an overlapping region. The homogeneous region was transformed and blended. The inhomogeneous region was unprocessed; (d) The same stitching result with color mapping applied to the inhomogeneous region.

$$\omega(x, y, i) = \frac{1/d_i^2}{\sum_{k=1}^{n} 1/d_k^2} \tag{2}$$

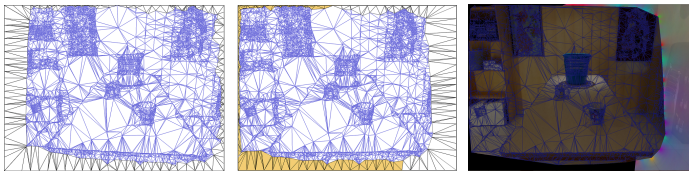
where $n = 3$, corresponding to the simplex dimension of a triangle. Using reciprocal squared distances results in a smoother transition between the homography clusters during weighted warping [5]. An example visualization of the multidimensional weight map for an overlapping homogeneous region is shown in Fig. 3 (b).

As introduced, inhomogeneous regions represent "critical image areas", typically exhibiting parallaxes. These cannot be aligned properly. To avoid ghosting artifacts, computation of the weight matrix, weighted transformation and blending are omitted for this region. In order to prevent visual color discrepancies between the blended overlapping region and the non-blended parallax region, color histograms are precomputed for both, the inhomogeneous area of the reference image and the separately blended homogeneous area. Subsequently, histogram-matching-based color mapping is performed

Algorithm 1 Homogenization (of overlapping region)

Input: Source feature point set P, corresponding target feature point set P', global homography set H, threshold parameter ϵ of RANSAC
Output: P and P' associated with multiple $h \in H$

for each $p \in P$ and corresponding $p' \in P'$ **do**
  for each $h \in H \mid$ neighbor of p has homography h **do**
    Transform p into the target image: p_T := $h * p$
    if $d(p', p_T) < \epsilon$ **then**
      Assign h to local homography set of p
    end if
  end for
end for



(a) (b) (c)

Fig. 4: (a) Triangulation of a reference feature set (blue triangles) with additional border points (black triangles) considered for cutting. (b) Triangulation mesh mapped to the target feature set including border points. Orange target triangles are cropped. (c) Visualization of a multidimensional per-pixel weight map for a non-overlapping region in a target image (border pixel area). Each estimated homography is associated with an individual color. The final visualization color of a pixel results from the weighted mixture of all associated homography colors. Black pixels illustrate the cut out border areas of the target image.

per color channel: Pixels of the inhomogeneous region are modulated using histogram equalization w.r.t. the histogram of the homogeneous region [1]. Fig. 3 (c-d) shows an example color mapping result in an inhomogeneous region.

3.3 Non-Overlapping Region

In this stage, the non-overlapping region is processed, preparing it for later perspective transformation regarding correct alignment and prevention of blending artifacts.

First, triangular sub-regions in the non-overlapping target image area, which would intersect with the overlapping region of the reference image after perspective transformation, are identified: Auxiliary points are placed at the image border rectangle (per image) using an equidistant distribution with a density corresponding to the number of edge vertices of the Delaunay mesh. These border vertices are triangulated incrementally with the existing mesh of the overlapping region in the reference image. The resulting mesh is then mapped to the target image, and corresponding triangles in the reference and target images are compared: If a triangle area in the target image is smaller than the corresponding triangle area in the reference image, then it conveys less visual information for stitching. In this case, this target triangle's pixel area is cut out. The target image cutting process is illustrated in Fig. 4 (a-b).

The second step is the extension of the multidimensional per-pixel weight map ω to the non-overlapping region for later perspective transformation: For each feature point cluster resulting from DFM's iterative rematching (cf. sub-section 3.1) and each pixel position (x', y') in the cropped target image, the closest feature point is determined. Then, for every cluster homography H_j – based on the pixel distance d_j between (x', y') and the closest feature point – its weight $\omega(x', y', j)$ is calculated according to equation 2. Here, number n refers the total number of estimated homographies (representing n many weight matrices). The closer a selected pixel is to a feature cluster, the higher is the weight assigned to its respective homography. Incorporating all homographies with their corresponding weights results in robust perspective transformation for

the non-overlapping region. An example visualization of the extended multidimensional weight map is shown in Fig. 4 (c).

3.4 Stitching

The last pipeline stage comprises final stitching computations, including perspective transformation of the cropped target image. For each target pixel, the final homography is computed using the multidimensional per-pixel weight map ω as follows:

$$H(x,y) = \sum_{k=1}^{n} \omega(x,y,k) H_k \;\mid\; \sum_{k=1}^{n} \omega(x,y,k) = 1 \tag{3}$$

Perspective transformation warping using multiple weighted homographies may result in small pixel gaps in the resulting image due to scaling and projective distortions, respectively. This gap-filling problem is solved using backward warping in combination with inverse bilinear interpolation. Finally, pixels in the homogeneous region of the reference image, which overlap with the transformed pixels from the target image, are uniformly blended to achieve a visually smooth transition in the stitching result.

4 Experiments and Discussion

The underlying software prototype was developed in C++ using the OpenCV 4 library. The implemented stitching pipeline was tested with ten image pairs, each with challenging varying parallax dimensions. The evaluation set includes three image pairs of the newly captured "Cellar Room" dataset, including different camera baselines and viewing directions. Additionally, image pairs "Propeller", "Building", "Seattle", "Backyard", "Adobe" and "Garden" were picked from the "Parallax-tolerant Image Stitching" dataset [30]. And lastly, image pair "Dwarves" is from the "Middlebury Stereo" datasets [8,23].

4.1 Evaluation

The first evaluation part includes the comparison of visual results w.r.t. other three well-known stitching methods: Global homography (GH), APAP [29] and ELA [12]. GH represents a "basic stitching algorithm", aligning image pairs using a single dominant homography. Our implementation of GH additionally incorporates USAC [22] for a more reliable outlier removal during feature matching. APAP and ELA are adaptive image alignment methods (Source codes were provided by the respective authors). APAP uses moving direct linear transformation (MLDT) to compute homographies for grid cells using a spatial distance-based weighting scheme. ELA aligns grid cells by approximating the warping error analytically. In Fig. 5 all image stitching results are presented. Comparisons to approaches with a different methodological focus were avoided, like shape-preserving warping [2,3,13,15] and seam-cutting [6,16,28,30], respectively.

The second part of the evaluation focuses on a quantitative analysis: The methods and test image pairs used in this part are the same as those shown in Fig. 5. Firstly,

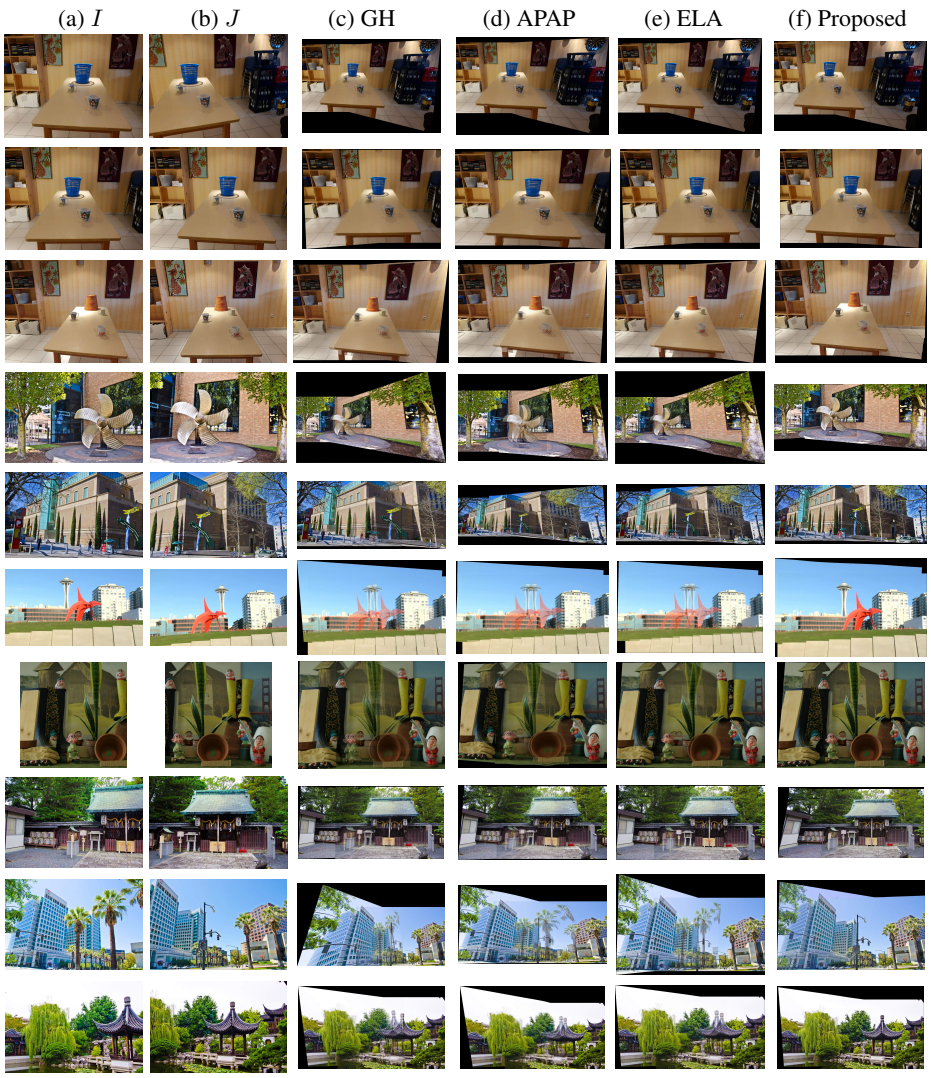


Fig. 5: Image stitching results of the proposed pipeline, including comparisons to other three methods: GH (described in text), APAP [29] and ELA [12]. The scenes are arranged from top to bottom in the following order: Three different scenes of "Cellar Room", one scene of "Propeller", "Building", "Seattle", "Dwarves", "Backyard", "Adobe" and "Garden".

the image alignment performance of the overlapping region was evaluated using the structural similarity index measure (SSIM) [26]. The overlapping region was extracted using DFM's Delaunay triangulation (cf. sub-sections 3.1 and 3.2). The computed SSIM scores for the overlapping region are presented in Fig. 6. Secondly, the non-overlapping region, which is susceptible to perspective distortion, was also evaluated as follows

(Overall results, incl. time measurements, are in the next sub-section): To simulate ground truth for the used image pairs, their overlapping regions were cropped in the image pairs. Subsequently, each cropped area was uniformely subdivided into one simulated overlapping and two (left and right) non-overlapping regions. The new simulated image pairs were again stitched using the presented methods, and the perspective-transformed non-overlapping regions were evaluated against ground truth using SSIM.

4.2 Results

In comparison to the other evaluated methods, the proposed pipeline achieves significantly improved stitching quality for all scenes regarding the handling of parallaxes: In the first image pair of "Cellar Room" only small parallaxes are noticeable (mainly due to the table in the foreground). So, no significant parallax artifacts are visible in any of the methods' stitching results. However, the propsed method is the only one that exhibits precise alignment of foreground and background objects, including accurately overlapping joints of the floor flows. The second and third image pairs of "Cellar Room" contain larger parallaxes. In contract to the proposed method, GH, APAP and ELA produce significant misalignment artifacts in the overlapping region. The image pair of "Propeller" exhibits noticeable parallaxes, particularly around the rotor blades. Again, stitching results of GH, APAP and ELA contain clearly recognizable visual defects in the parallax regions. Additionally, they fail to compute a proper perspective transformation for the non-overlapping region of the target image, as the right tree is slightly skewed. For the image pair of "Building", ELA and the proposed method provide the most plausible perspective transformation for the non-overlapping region. However, all other methods produce parallax artifacts at the trees, including corresponding misalignments of the overlapping region. The image pair of "Seattle" has the largest parallaxes in the presented evaluation. Only the proposed method handles the inhomogeneous region appropriately. The image pairs of "Dwarves" and "Backyard" have a high depth complexity and various minor parallaxes. Due to the variety of objects with different scene depths, GH, APAP and ELA fail to align both, foreground and background objects simultaneously. However, for the image pair of "Backyard", GH, APAP and ELA provide a better perspective transformation than the proposed method (Notice the slightly sheared cottage on the left). The perspective transformation of the image pair of "Adobe" is plausible for all methods except GH. But, APAP and ELA fail to prevent ghosting artifacts at the palm trees, caused by larger parallaxes. For the image pair of "Garden", all perspective transformation of the non-overlapping region provide similar visual qualities. Nevertheless, GH, APAP, or ELA result again in ghosting artifacts, particularly at the foreground garden pavilion. Additionally, most of the the trees in the background appear blurred compared to the proposed method.

The results of the quantitative evaluation can be summarized as follows: The presented method significantly improves image alignment of the overlapping region in all scenes compared to the other evaluated methods. GH, APAP, and ELA have average SSIM scores of 0.44, 0.54, and 0.61 and harmonic SSIM scores of 0.40, 0.48, and 0.56, respectively. The proposed method achieves the best results, i.e. an average SSIM score of 0.85 (ca. 39% to 93% increase) and a harmonic SSIM score of also 0.85 (ca. 52% to 113% increase). Moreover, the proposed solution also shows improvement in the SSIM

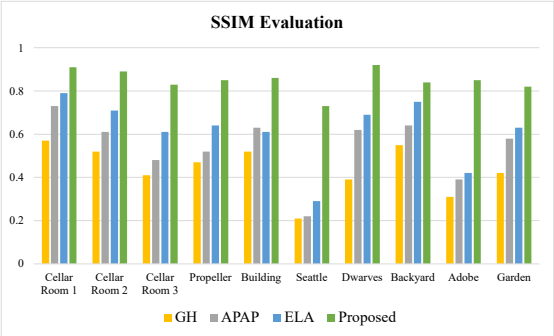


Fig. 6: Image alignment evaluations using SSIM metric.

score for the non-overlapping region: Here, GH, APAP, and ELA have average SSIM scores of 0.33, 0.47, and 0.57, with harmonic SSIM scores of 0.29, 0.46, and 0.55, respectively. The presented solution achieves an average SSIM score of 0.63 (ca. 11% to 91% increase) and a harmonic SSIM score of 0.61 (ca. 11% to 110% increase). Regarding runtime measurements, the described processing pipeline takes on average 27s for feature matching, 14s for the overlapping region, 11s for the non-overlapping region and 15s for stitching, resulting in a total time of 67s. The other evaluated methods are on average faster: GH takes 23s, APAP 21s and ELA 29s for the total processing time. However, their visual quality is significantly worse compared to the proposed method.

The presented solution has some limitations: The robustness of the stitching process is primarily determined by the quality of the DFM results, particularly the homographic decomposition. In this evaluation, the given default settings of DFM were used, as this directly led to satisfactory results. But generally, incorrect parameter settings may cause errors, like undetected parallax regions, typically leading to ghosting artifacts. In such cases, manual re-parameterization of DFM would be necessary for improving quality.

5 Conclusions

In this paper, a novel image stitching method for image pairs with high depth complexity and larger parallaxes is presented. The first stage of the proposed stitching pipeline utilizes dense feature matching to generate a homographic decomposition of the image space, including precise and dense feature correspondences between image pairs. Using Delaunay triangulation of the matching set makes it possible to identify the overlapping region for stitching. In the second stage, the homogenization algorithm is used to compute a “parallax-tolerant” overlapping region. This is realized by detection and segmentation of “inhomogeneous” sub-regions, which are typically caused by parallax effects. Inhomogeneous image parts are not considered for perspective transformation, nor for blending, during stitching. For perspective transformation of the parallax-free overlapping region, a multidimensional per-pixel weight map of the target image is computed. Additionally, color mapping assures uniform blending results for the inhomogeneous image region. In the third pipeline stage, the non-overlapping region is partially cropped

to exclude border areas that would produce visual artifacts after perspective warping. Finally, the multidimensional weight map is extended to provide robust weighting for perspective transformation even in the non-overlapping image region. The presented results on challenging real-world stitching datasets demonstrate that the proposed method achieves accurate and robust image alignments, minimizing ghosting artifacts. It outperforms the other evaluated methods in term of visual quality and structural similarity. Future work includes multi-stitching of several images and performance optimizations.

References

1. Acharya, T., Ray, A.K.: Image Processing - Principles and Applications (2005)
2. Chang, C.H., Sato, Y., Chuang, Y.Y.: Shape-preserving half-projective warps for image stitching. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3254–3261 (2014)
3. Chen, Y.S., Chuang, Y.Y.: Natural image stitching with the global similarity prior. In: European Conference on Computer Vision. pp. 186–201 (2016)
4. Chen, Y., Xue, W., Chen, S.: Large parallax image stitching via structure preservation and multi-matching. In: Neural Computing for Advanced Applications. pp. 177–191 (2022)
5. Gao, J., Kim, S.J., Brown, M.S.: Constructing image panoramas using dual-homography warping. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 49–56 (2011)
6. Gao, J., Li, Y., Chin, T.J., Brown, M.S.: Seam-driven image stitching. In: Eurographics (2013)
7. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision (2003)
8. Hirschmuller, H., Scharstein, D.: Evaluation of cost functions for stereo matching. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2007)
9. Jia, Q., Li, Z., Fan, X., Zhao, H., Teng, S., Ye, X., Latecki, L.J.: Leveraging line-point consistence to preserve structures for wide parallax image stitching. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 12181–12190 (2021)
10. Lee, K.Y., Sim, J.Y.: Stitching for multi-view videos with large parallax based on adaptive pixel warping. IEEE Access (2018)
11. Lee, K.Y., Sim, J.Y.: Warping residual based image stitching for large parallax. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 8195–8203 (2020)
12. Li, J., Wang, Z., Lai, S., Zhai, Y., Zhang, M.: Parallax-tolerant image stitching based on robust elastic warping. IEEE Transactions on Multimedia pp. 1672–1687 (2018)
13. Li, N., Xu, Y., Wang, C.: Quasi-homography warps in image stitching. IEEE Transactions on Multimedia pp. 1365–1375 (2017)
14. Liao, T., Li, N.: Single-perspective warps in natural image stitching. IEEE Transactions on Image Processing p. 724–735 (2020)
15. Lin, C.C., Pankanti, S.U., Ramamurthy, K.N., Aravkin, A.Y.: Adaptive as-natural-as-possible image stitching. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1155–1163 (2015)
16. Lin, K., Jiang, N., Cheong, L.F., Do, M.N., Lu, J.: Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In: European Conference on Computer Vision (2016)
17. Liu, W.X., Chin, T.J.: Correspondence insertion for as-projective-as-possible image stitching. ArXiv (2016)
18. Lyu, W., Zhou, Z., Chen, L., Zhou, Y.: A survey on image and video stitching. Virtual Reality and Intelligent Hardware pp. 55–83 (2019)

19. Megha, V., Rajkumar, K.K.: Automatic satellite image stitching based on speeded up robust feature. In: International Conference on Artificial Intelligence and Machine Vision. pp. 1–6 (2021)
20. Nie, L., Lin, C., Liao, K., Liu, M., Zhao, Y.: A view-free image stitching network based on global homography. Journal of Visual Communication and Image Representation (2020)
21. Nie, L., Lin, C., Liao, K., Liu, S., Zhao, Y.: Unsupervised deep image stitching: Reconstructing stitched features to images. IEEE Transactions on Image Processing (2021)
22. Raguram, R., Chum, O., Pollefeys, M., Matas, J., Frahm, J.M.: Usac: a universal framework for random sample consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence (2013)
23. Scharstein, D., Pal, C.: Learning conditional random fields for stereo. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–8 (2007)
24. Seibt, S., Von Rymon Lipinski, B., Latoschik, M.E.: Dense feature matching based on homographic decomposition. IEEE Access **10** (2022)
25. Szeliski, R.: Image alignment and stitching: A tutorial. Foundations and Trends in Computer Graphics and Vision (2006)
26. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing pp. 600–612 (2004)
27. Win, K.P., Kitjaidure, Y., Hamamoto, K.: Automatic Stitching of Medical Images Using Feature Based Approach. Advances in Science, Technology and Engineering Systems Journal pp. 127–133 (2019)
28. Xue, W., Xie, W., Zhang, Y., Chen, S.: Stable linear structures and seam measurements for parallax image stitching. IEEE Transactions on Circuits and Systems for Video Technology pp. 253–261 (2022)
29. Zaragoza, J., Chin, T.J., Brown, M.S., Suter, D.: As-projective-as-possible image stitching with moving dlt. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 2339–2346 (2013)
30. Zhang, F., Liu, F.: Parallax-tolerant image stitching. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 3262–3269 (2014)
31. Zhang, G., He, Y., Chen, W., Jia, J., Bao, H.: Multi-viewpoint panorama construction with wide-baseline images. IEEE Transactions on Image Processing pp. 3099–3111 (2016)
32. Zhao, Q., Ma, Y., Zhu, C., Yao, C., Feng, B., Dai, F.: Image stitching via deep homography estimation. Neurocomputing (2021)
33. Zhao, Q., Wan, L., Feng, W., Zhang, J., Wong, T.T.: Cube2video: Navigate between cubic panoramas in real-time. IEEE Transactions on Multimedia pp. 1745–1754 (2013)