# Learning Residual Elastic Warps for Image Stitching under Dirichlet Boundary Condition

Minsu Kim[1], Yongjun Lee[2], Woo Kyoung Han[1], and Kyong Hwan Jin[2*]

[1]DGIST, Republic of Korea   [2]Korea University, Republic of Korea

[1]{axin.kim, cjss7894}@dgist.ac.kr   [2]{lyj9805, kyong_jin}@korea.ac.kr

## Abstract

*Trendy suggestions for learning-based elastic warps enable the deep image stitchings to align images exposed to large parallax errors. Despite the remarkable alignments, the methods struggle with occasional holes or discontinuity between overlapping and non-overlapping regions of a target image as the applied training strategy mostly focuses on overlap region alignment. As a result, they require additional modules such as seam finder and image inpainting for hiding discontinuity and filling holes, respectively. In this work, we suggest Recurrent Elastic Warps (REwarp) that address the problem with Dirichlet boundary condition and boost performances by residual learning for recurrent misalign correction. Specifically, REwarp predicts a homography and a Thin-plate Spline (TPS) under the boundary constraint for discontinuity and hole-free image stitching. Our experiments show the favorable aligns and the competitive computational costs of REwarp compared to the existing stitching methods. Our source code is available at* https://github.com/minshu-kim/REwarp.*

## 1. Introduction

Estimation of geometric transformations for image stitching requires descriptions of the spatial correspondences between given scenes. To calculate such relationships, existing methods are categorized into two branches: feature-based methods [5, 7, 8, 13, 15, 16, 18, 28] and learning-based [10, 11, 21–25] methods. Feature-based approaches detect key points or lines to match the textures of images in order to estimate the optimal transformation. In contrast, learning-based frameworks compute a 4D cost volume [27] which computes receptive field-wise correlations and provides a prior like dense matching between given scenes.

The two branches have clearly distinct limitations. The recent feature-based frameworks [7, 15] leverage line and
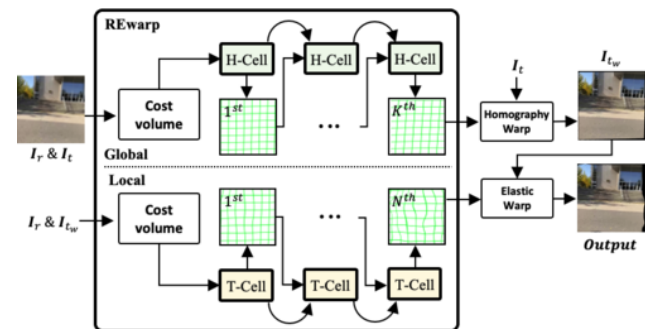


Figure 1. **Overview of our approach.** REwarp sequentially estimates homography and thin-plate spline with two recurrent neural networks. This combination enables elastic image alignment for parallax-tolerant image stitching.

key point matching and estimate transformation minimizing the error of the matched features keeping geometric structures. However, the approaches cause frequent failures and performance degradation in challenging environments like low resolution and few textured scenes.

On the other hand, the recent parallax-tolerant deep image stitchings [10, 22] are robust to few textured images owing to the CNNs with wide receptive fields and the prior of dense correspondences from correlation-based cost volumes [27]. The methods are implemented by either a parametric combination (homography and *Thin-plate Spline* (TPS) [22]) or a *Warp Field* [10]. Because the TPS-based image deformation is a global warp, sometimes it shows limited correction of local misalignment under large parallax errors. In contrast, the flow-based image stitching shows warps with high flexibility for alignment of wide parallax. However, because the method is exposed to artifacts or holes (Fig. 8), it requires modules to address the problems. Furthermore, the occasional discontinuity (Fig. 7) of deep elastic warps between overlap and non-overlap regions requires seam masks [9, 14, 17] to hide the unpleasant boundary.

To address the problems, we suggest Recurrent Elastic warps (REwarp) with Dirichlet boundary condition that

---

*Corresponding author.

(a) **Observation on overlap region.** $1, 2, 4, 6^{th}$ iterations of global(top) and the residual(bottom) alignment

(b) **Image Stitching with our approach.**

Figure 2. Visual Demonstration of our approach.

learns discontinuity-free residual elastic warps. For each inference of recurrent estimations, REwarp predicts an elastic warp parameterized by a homography and a Thin-plate spline. Our experiments show that REwarp addresses the generation of holes and discontinuity of learning-based elastic warps with remarkable image alignment.

In summary, our main contributions are as follows:

- We propose a novel deep image stitching that corrects residual misalignment and varying degrees of parallax errors using elastic warps.

- Our approach shows favorable performance without additional hole-filling and discontinuity correction modules required in the previous methods of deep elastic warp for image stitching.

- We investigate the computation costs of REwarp and our baselines to show that REwarp is a lightweight and failure-tolerant model. The properties will shed light on real-time deep image stitching.

## 2. Related Work

**Elastic Warp for image stitching** As a global alignment by single homography lacks the consideration of local misalignment or parallax, there have been various studies and trials to overcome such a limitation. Gao *et al.* [5] suggested a method that determines two globally dominant frames and estimates dual homography. Lin *et al.* [18] proposed smoothly varying affine fields. Zaragoza *et al.* [28] suggested moving Direct Linear Transformation (DLT) [6] to infer as-projective-as-possible (APAP) spatially weighted homography. Li *et al.* [13] proposed robust elastic warps (robust ELA) that compute *thin plate splinte* (TPS) to finely optimize pixel-wise deformations. Joo *et al.* [8] suggested a line-guided moving DLT for structure-preserving image stitching. Liao *et al.* [15] proposed single-

perspective warps that leverage dual-feature (point + line) for structure-preserving image stitching. Lee *et al.* [12] introduced residual warping for large parallax alignment using multiple homography. Jia *et al.* [7] suggested a line-point-consistence (LPC) constraint for advanced geometric structure-preserving image alignment. Kweon *et al.* [10] suggests Pixel-Wise warping Module (PWM) that learns warp field for wide-parallax image alignment. Recently, Nie *et al.* [22] proposed UDIS++ that contains a deep seamless compositor, a homography estimator, and a TPS estimator for deep image stitching. Despite the successful extends of deep image stitchings to elastic warps [10, 22, 25], there are still challenging hurdles on completeness of aligned image composition. They struggle with discontinuity between overlap and non-overlap regions in target images. In this work, we address the problem and suggest novel deep residual warps for image stitching.

## 3. Problem Formulation

Parallax is a local misalignment caused by lost information on depth planes. As depicted in Fig. 4, the misalignment in projection plane $\boldsymbol{\pi}_2$ arises from the disparity between the homography projection ($\mathbf{H} : \mathbf{x}_1 \mapsto \mathbf{x}_2'$) and the direct projection from real world to an image plane $\boldsymbol{\pi}_2$ ($\mathbf{A}_{\boldsymbol{\pi}_2} : \mathbf{x} \mapsto \mathbf{x}_2$) [6]. The mismatched depth ($\mathbf{x} - \mathbf{x}_{\boldsymbol{\pi},1}$) causes an intrinsic limitation to align such parallax error ($\vec{\boldsymbol{\epsilon}}$), which is represented as

$$\vec{\boldsymbol{\epsilon}} = \mathbf{x}_2 - \mathbf{x}_2' = \mathbf{A}_{\boldsymbol{\pi}_2}\mathbf{x} - \mathbf{H}\mathbf{x}_1, \qquad (1)$$
$$= \mathbf{A}_{\boldsymbol{\pi}_2}\mathbf{A}_{\boldsymbol{\pi}_1}^{-1}\mathbf{x}_1 - \mathbf{H}\mathbf{x}_1. \qquad (2)$$

To minimize the error $\vec{\boldsymbol{\epsilon}}$ on the plane, our network aims to predict a warp field $\widehat{\mathbf{F}}$ [10] to compensate $\vec{\boldsymbol{\epsilon}}$ and a homography $\widehat{\mathbf{H}}$ to globally align two images. Note that we denote $\widehat{(\cdot)}$ as an estimated warp used for image stitching. The overall formulations for alignment of parallax errors are the same
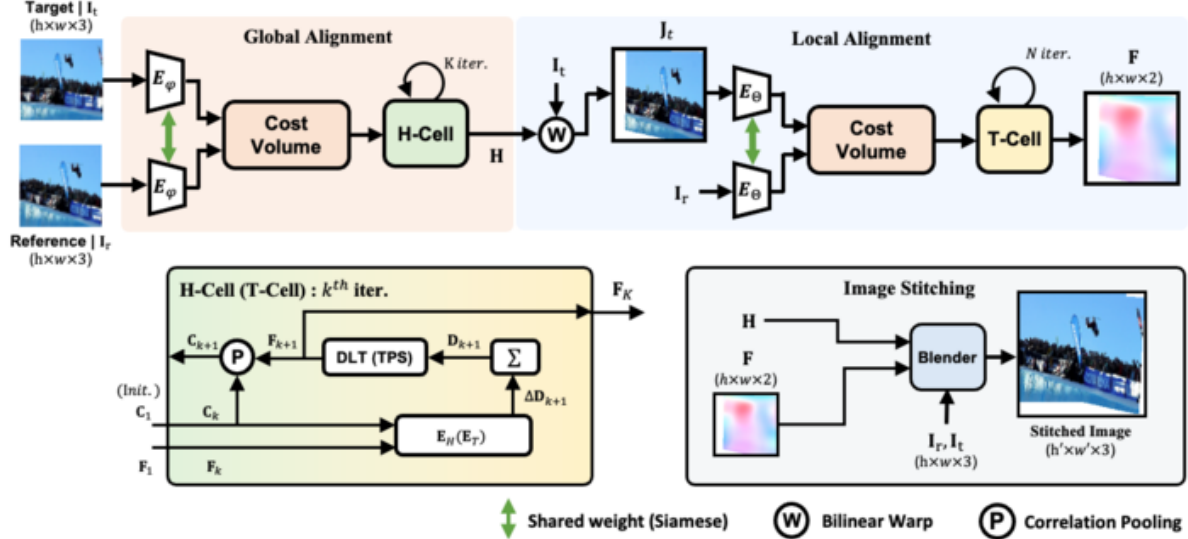
Figure 3. **Architectural Overview of REwarp.** Our model consists of transformation estimators for global alignment and local residual alignment. After the estimation of both parametric warps, we ensemble them and composite an elastic warp field. In our H/T-Cell, an accumulator ($\Sigma$) sums estimated displacements ($\Delta \mathbf{D}_k$) from a regressor that consists of multi-layer CNNs ($\mathbf{E}_H$ or $\mathbf{E}_T$).
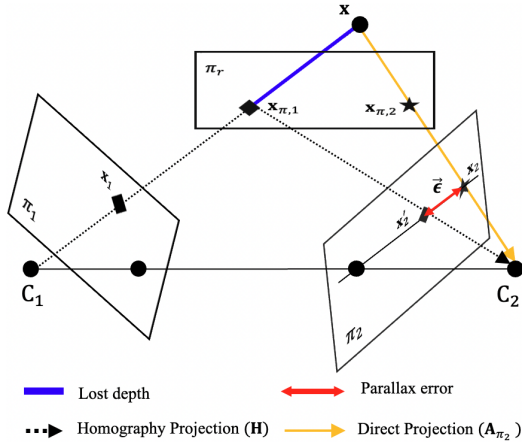


Figure 4. **A parallax error on point $\mathbf{x}_2$.** $\pi$ denotes a projection plane, $\mathbf{C}$ is a view point, $\mathbf{x}'$ is a point projected by a homography.

as follows:

$$
\begin{aligned}
\min \vec{\boldsymbol{\epsilon}} \; &\equiv \; \min ||\mathbf{x}_2 - (\widehat{\mathbf{H}} \cdot \mathbf{x}_1 + \widehat{\mathbf{F}}[\mathbf{x}'_2])||, \quad (3) \\
&= \; \min ||\mathbf{x}_2 - (\mathbf{x}'_2 + \widehat{\mathbf{F}}[\mathbf{x}'_2])||, \\
&= \; \min ||\mathbf{x}_2 - \widehat{\mathbf{x}}_2||,
\end{aligned}
$$

where $\mathbf{x}_2$ denotes an ideal point, $\mathbf{x}'_2$ is a point exposed to a prallax error $\vec{\boldsymbol{\epsilon}}$ by the homography projection.

The estimated grid $\widehat{\mathbf{X}}_2 := \{\mathbf{x}_2 | \mathbf{x}_2 = \mathbf{x}'_2 + \vec{\boldsymbol{\epsilon}}\}$ is used for the image warping, which is defined as

$$
W(\mathcal{I}, \mathcal{X}) := \mathcal{I}[\mathcal{X}], \quad (4)
$$

where $\mathbf{I} \in \mathcal{I}$ and $\mathbf{X} \in \mathcal{X}$ is a set of images ($\in \mathbb{R}^{h \times w}$) and a warped grid of a stitched frame ($\in \mathbb{R}^{h' \times w'}$) respectively. To satisfy our formulations, REwarp estimates an accumulated

displacement vector $\mathbf{D}^G$ and $\mathbf{D}^L$ for homography [6] and TPS computation, respectively. Specifically, we compute a warped grid $\widehat{\mathbf{X}}'_2$ using $\mathbf{D}^G$ and a warp field $\widehat{\mathbf{F}} \in \mathbb{R}^{h \times w \times 2}$ using $\mathbf{D}^L$ as

$$
\widehat{\mathbf{X}}'_2 = \mathcal{S}_{\mathbf{V}}(\mathbf{D}^G) \cdot \mathbf{X}_1, \quad (5)
$$

$$
\widehat{\mathbf{X}}_2 = \widehat{\mathbf{X}}'_2 + \mathcal{S}_{\mathbf{P}_r}(\mathbf{D}^L; \mathbf{X}^o), \quad (6)
$$

where $\mathcal{S}_{\mathbf{V}}$ and $\mathcal{S}_{\mathbf{P}_r}$ denote a Direct Linear Transformation [6] and a warp field evaluation (Sec. 4.1), respectively. $\mathbf{V} := \{0, h\} \times \{0, w\}$ is a vector containing four corner coordinates, $\mathbf{P}_r := \{0, \frac{1}{12}, ..., \frac{11}{12}\}^2 \in \mathbb{R}^{12 \times 12}$ is a control point grid for $\mathbf{I}_r$ (or a pre-computed uniform grid). $\mathbf{X}^o$ means a coordinate set consisting of overlapping regions given by an estimated homography $\widehat{\mathbf{H}} = \mathcal{S}_{\mathbf{V}}(\mathbf{D}^G)$.

## 4. Method

REwarp consists of CNN encoders ($\mathbf{E}_\varphi, \mathbf{E}_\Theta$), an H-Cell ($f^G$) and a T-Cell ($f^L$). Both cells take in a 4D cost volume to estimate a homography and a TPS representation, respectively. Specifically, H-Cell and T-Cell predict a residual displacement vector $\Delta \mathbf{D}^G$ and $\Delta \mathbf{D}^L$ of four corners $\mathbf{V}$ [3] and a control point grid $\mathbf{P}_r$ [1], respectively. The estimated residual vectors are accumulated and converged to the optimal vectors given $K$ (H-Cell) or $N$ (T-Cell) number of iterations as

$$
\mathbf{D}^G = \sum_{k=1}^{K} \Delta \mathbf{D}_k^G, \;\; \mathbf{D}^L = \sum_{n=1}^{N} \Delta \mathbf{D}_n^L, \quad (7)
$$

$$
\text{where} \;\; [\Delta \mathbf{D}_1^G, \ldots, \Delta \mathbf{D}_K^G] = f^G(\mathbf{C}, K; \mathbf{I}_r, \mathbf{I}_t, \varphi),
$$

$$
[\Delta \mathbf{D}_1^L, \ldots, \Delta \mathbf{D}_N^L] = \lambda \cdot f^L(\mathbf{C}, N; \mathbf{I}_r, \mathbf{J}_t, \Theta),
$$

**Algorithm 1** Operations in H-Cell and T-Cell.

---

**Input :** $\mathbf{C}, k_{\text{iter}}$             ▷ Cost Volume, Init., # Iter.
**Output :** $\widehat{\mathbf{X}}$                              ▷ Elastic Grid

---

1: **procedure** CELL($\mathbf{C}, k_{\text{iter}}$)
2:      $\mathbf{F}_1, \delta_1 \leftarrow \vec{\mathbf{0}}$
3:      $k \leftarrow 1$
4:
5:      **while** $k \leq k_{\text{iter}}$ **do**
6:          ## Prior Preparation
7:          $\mathbf{C}_k \leftarrow f_c(\mathbf{C}; \mathbf{X}_k)$                 ▷ $f_c$ := Pooling [26]
8:          $\mathbf{s}_{k+1} = [\mathbf{C}_k, \mathbf{F}_k]$
9:
10:          ## Update & Predictions
11:          **if** H-Cell **then**
12:              $\Delta\mathbf{D}_{k+1}^G \leftarrow \mathbf{E}_H(\mathbf{s}_{k+1})$
13:              $\mathbf{D}_{k+1}^G \leftarrow \sum_{j=1}^{k+1} \Delta\mathbf{D}_j^G$
14:              $\mathbf{X}_{k+1} \leftarrow \mathcal{S}_\mathbf{V}(\mathbf{D}_{k+1}^G)$          ▷ $\mathcal{S}_\mathbf{V}$ := DLT
15:          **end if**
16:
17:          **if** T-Cell **then**
18:              $\Delta\mathbf{D}_{k+1}^L \leftarrow \mathbf{E}_T(\mathbf{s}_{k+1})$
19:              $\mathbf{D}_{k+1}^L \leftarrow \sum_{j=1}^{k+1} \Delta\mathbf{D}_j^L$
20:              $\mathbf{F}_{k+1} = \mathcal{S}_{\mathbf{P}_r}(\mathbf{D}^L)_{k+1}$     ▷ $\mathcal{S}_{\mathbf{P}_r}$ := F Eval. (Sec. 4.1)
21:              $\mathbf{X}_{k+1}[\mathbf{x}^o] = \mathbf{X}_k[\mathbf{x}^o] + \mathbf{F}_{k+1}[\mathbf{x}^o]$
22:          **end if**
23:          $k \leftarrow k + 1$
24:      **end while**
25: **end procedure**

---

$\mathbf{J}_t$ denotes a target image warped by a homography $\mathbf{H}$ computed from $\mathbf{D}^G$. As shown in the above equation, a vector $\mathbf{D}^L$ for a control point grid is estimated from a globally aligned image $\mathbf{J}_t$ and a reference image $\mathbf{I}_r$ to correct the remaining parallax errors. More details and pipelines of our REwarp are provided in Algorithm 1.

### 4.1. Warp Field Evaluation

We obtain a warp field ($\mathbf{F}$) via *Thin-plate Spline* Transformation as

$$\mathbf{F} = \mathcal{S}_{\mathbf{P}_r}(\mathbf{D}^L) = \mathbf{X}_t - \mathbf{U}, \qquad (8)$$

where $\mathbf{U} := [0, h) \times [0, w)$ is an uniform grid, $\mathbf{X}_t \triangleq \{\mathbf{x}_t | \mathbf{x}_t \in \mathbb{R}^2\}$ is a warped grid by TPS transformation.

**TPS Transformation** An explicit representation TPS provides a warp field ($\mathbf{F}$) given control points $\mathbf{P}_r$ and $\mathbf{P}_t \in \mathbb{R}^{12 \times 12}$. In this work, we assume that control points $\mathbf{P}_t$ in $\mathbf{I}_t$ are presented as $\widehat{\mathbf{P}}_t = \mathbf{P}_r + \mathbf{D}^L$ under the assumption that $\mathbf{P}_r$ is equal to a fixed uniform grid to ease the formulation for image alignment. Specifically, TPS calculates the affine weights ($\vec{v} := [a, b, c]$) and kernel weights ($\vec{w} := [w_1, w_2, \cdots, w_M]$) for each $x$ and $y$ axis as

$$\begin{bmatrix} \vec{w}_x & \vec{v}_x \\ \vec{w}_y & \vec{v}_y \end{bmatrix}^T = \mathbf{L}^{-1} \begin{bmatrix} \widehat{\mathbf{P}}_{tx}^T & 0 & 0 & 0 \\ \widehat{\mathbf{P}}_{ty}^T & 0 & 0 & 0 \end{bmatrix}^T, \qquad (9)$$



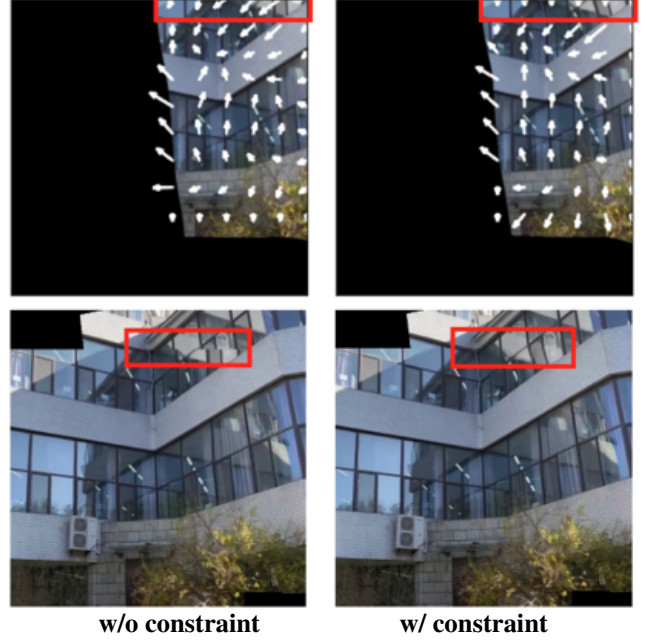**w/o constraint**           **w/ constraint**

Figure 5. **Discontinuity Removal.** The white arrows denote displacement vectors of control points to be applied to overlap region warping. As in the red box, our constraint removes discontinuity by relaxing the displacements in the edge.

where $\mathbf{L}$ is a deterministic matrix corresponding the pairs of control points [1], $[\widehat{\mathbf{P}}_{tx}; \widehat{\mathbf{P}}_{ty}] = \widehat{\mathbf{P}}_t$. The TPS derives each coordinate of the target grid ($\mathbf{X}_t \triangleq \{\mathbf{x}_t | \mathbf{x}_t \in \mathbb{R}^2\}$) with the calculated weights and reference grid $\mathbf{X}_r \triangleq \{\mathbf{x}_r | \mathbf{x}_r \in \mathbb{R}^2\}$ in a range of $[0, h) \times [0, w)$, as

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ y_t \end{bmatrix} = \begin{bmatrix} \vec{v}_x \cdot [1, \mathbf{x}_r] + \sum_m w_{x,m} B(||(\mathbf{x}_r - \mathbf{P}_{r,m})||) \\ \vec{v}_y \cdot [1, \mathbf{x}_r] + \sum_m w_{y,m} B(||(\mathbf{x}_r - \mathbf{P}_{r,m})||) \end{bmatrix},$$

where $B(z) = z^2 log z^2$ is a radial basis function, $w_{A,m}$ denotes a $A$ ($x$ or $y$) coordinate and $m^{th}$ kernel weight of M kernels.

**Dirichlet Boundary Condition** Though the recent remarkable deep parallax-tolerant warps [10, 22] can align paired images with wide-parallax errors, the methods struggle with discontinuity between overlap and non-overlap regions. To address the problem, we suggest a *Dirichlet Boundary Condition* for TPS-based deep image stitching. Specifically, we design REwarp to fix the edge of the control point grid ($\mathbf{P}_t \in \mathbb{R}^{12 \times 12}$) and leverage the estimated displacement vectors ($\Delta\mathbf{D}^L \in \mathbb{R}^{10 \times 10}$). After that, we plug the vectors into the prepared control point grid. As shown in Fig. 5, this constraint prevents REwarp from inconsistent discontinuity because TPS coefficients in Eq. (9) provide almost no warp in the edges of the overlap region owing to the property of radial basis function.
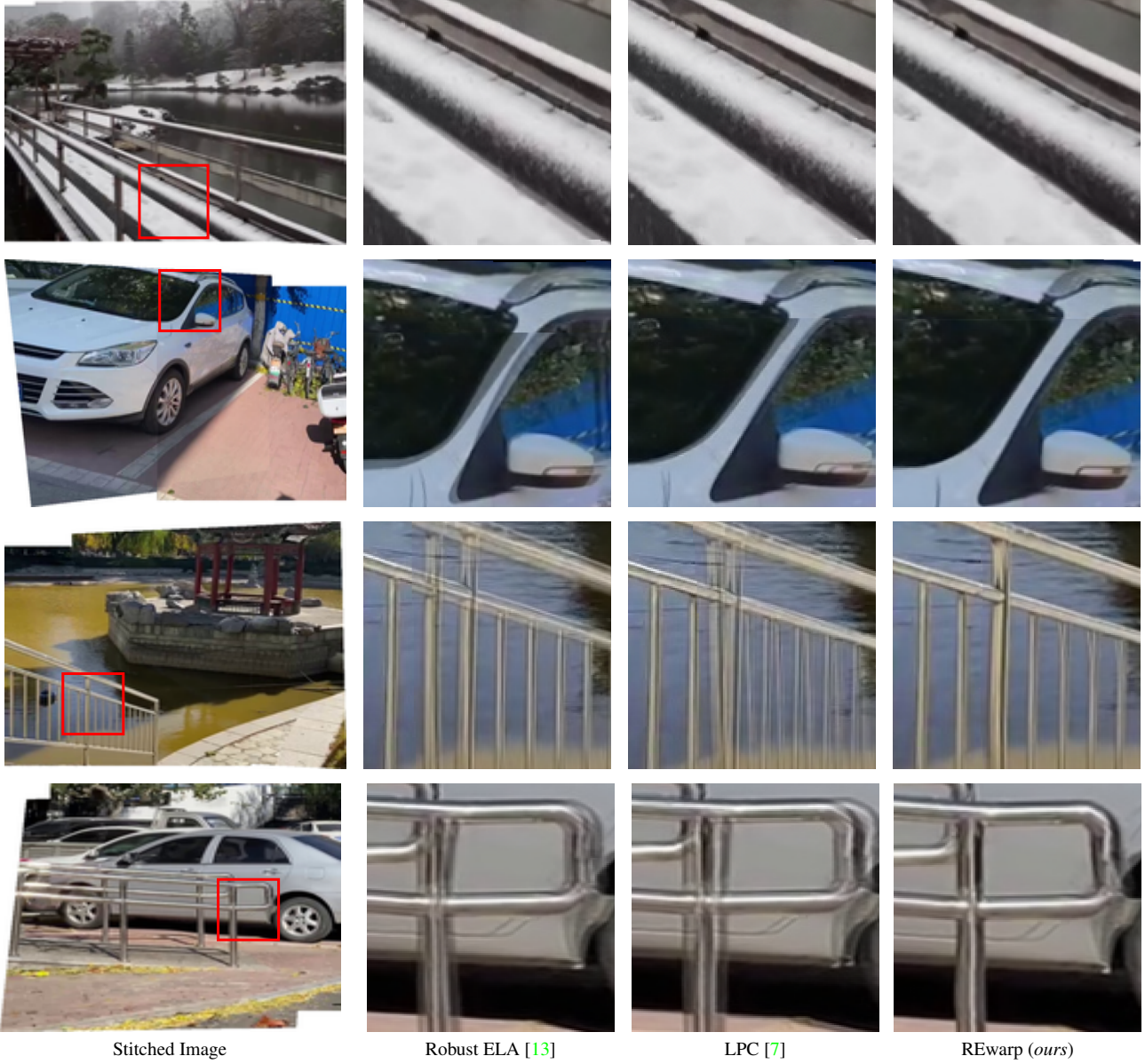
| Stitched Image | Robust ELA [13] | LPC [7] | REwarp (*ours*) |

Figure 6. **Qualitative comparisons to the other methods.**

## 4.2. Training

**H-Cell** We minimize photometric errors of overlapping region as

$$(\mathcal{L}_H) \ : \ \min_{\Theta_G} \sum_{k=1}^{K} \alpha^{K-k} \cdot \sum_{j \in \mathcal{J}_k} \|\mathbf{I}_r[\mathbf{x}_r^j] - \mathbf{I}_t[\mathbf{x}_t^j]\|_1, \quad (10)$$

where $\Theta_G$ denotes parameters in H-Cell. $\mathcal{J}_k$ denotes a valid region coordinate in the $k^{th}$ iteration's overlap region ($\mathcal{J} \triangleq \{j | j \in \mathbb{R}^2\}$), $\mathbf{x}_A^j$ is an overlap region coordinate in the plane of A. $\alpha$ is a weight ($= 0.85$). $K$ is the total iteration number of the H-Cell set as 6 in our implementation.

**T-Cell** We freeze H-Cell and train T-Cell to make REwarp strictly focus on learning the correction of residual parallax.

Similar to Eq. (10), we minimize L1 Loss as

$$(\mathcal{L}_T) \ : \ \min_{\Theta_L} \sum_{n=1}^{N} \alpha^{N-n} \cdot \sum_{q \in \mathcal{Q}_n} \|(\mathbf{I}_r[\mathbf{x}_r^q] - \mathbf{J}_t[\mathbf{x}_t^q])\|_1,$$

$$\text{where} \quad \mathbf{J}_t = W(\mathbf{I}_t, \widehat{\mathbf{H}} \cdot \mathbf{X}_t),$$

where $\mathbf{X}_t \in \mathbb{R}^{h \times w \times 2}$ is a grid of a target image, $\mathcal{Q}_n \triangleq \{q | q \in \mathbb{R}^2\}$ denotes a query set of $n^{th}$ overlapping region, $N$ is the total number of iterations for T-Cell that is set as 3 in our implementation. The optimization of TPS coefficient estimators is guided by the photometric loss, i.e., the criteria is to maximize the correlation of overlapped regions.

| | Overlap Ratio | | | Average (↑) | Failures (↓) | Time (ms) |
|---|---|---|---|---|---|---|
| | $\sim 30\%$ (↑) | $31 \sim 60\%$ (↑) | $61\% \sim$ (↑) | | | |
| SIFT [20] + RANSAC [4] | 18.32 | 21.68 | 22.30 | 21.48 | 1.27% | 111 |
| UDIS [21] | 19.61 | 20.15 | 19.88 | 19.97 | 0% | - |
| IHN [2] | 20.09 | 21.73 | 23.27 | 22.99 | 0% | 38 |
| APAP [28] | 21.28 | 22.30 | 23.54 | 22.69 | 12.30% | 574 |
| SPW [15] | 20.74 | 21.71 | 22.45 | 21.95 | 85.08% | 383 |
| LPC [7] | 17.97 | 21.04 | 21.59 | 20.82 | 42.13% | 1395 |
| Robust ELA [13] | 21.84 | 22.91 | 24.29 | 23.48 | 0.72% | 79 |
| REwarp (*ours*) | 22.11 | 24.55 | 26.08 | 24.84 | 0% | 50 |

Table 1. **Quantitative Comparisons on Image Alignment.** Red (or Blue) indicates the best (or second-best) performance.

| Method | Iteration (Num.) | mPSNR (dB) | Time (ms) |
|---|---|---|---|
| H-Cell | 1 | 19.15 | 32 |
| | 3 | 23.81 | 36 |
| | 6 | **24.84** | 50 |
| | 10 | 24.79 | 67 |
| T-Cell | 1 | 24.49 | 47 |
| | 3 | **24.84** | 50 |
| | 6 | 24.80 | 61 |
| | 10 | 22.75 | 82 |

(a) Number of Iterations.

| Method | mPSNR (dB) | Parameter (M) |
|---|---|---|
| **H** | 22.31 | 1.0 |
| **H + F** | 24.43 | 1.7 |
| **H+TPS (Ours)** | **24.84** | 1.8 |

(b) Implementation approach for Elastic Warp.

Table 2. **Ablation Study on Method specifications.**

# 5. Experiment

In our experiments, we use the UDIS-D [21] dataset that includes real images exposed to various degrees of parallax errors. We follow the splitting method of the pioneering work [21].

## 5.1. Implementation Details

To optimize REwarp, we apply 2-stage learning. First, we train H-Cell during 100 epochs and 8 mini-batch. After then, we set the epoch for T-Cell as 100 with 8 mini-batch. **Common Configurations** We use AdamW optimizer [19] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rates are initialized as $1 \times 10^{-4}$ and exponentially decayed by a factor of 0.98 for every epoch. We employ a $12 \times 12$ control point grid. Note that because we constrain the grid with Dirichlet Boundary Condition, REwarp estimates $11 \times 11$ grid and adds it to the $12 \times 12$ zero-initialized grid as described in Sec. 4.1.

## 5.2. Evaluation

**Qualitative Comparison** Fig. 6 shows the comparisons of REwarp to the previous parallax-tolerant stitchings [7, 13]. The first column contains REwarp's stitched results. The zoom-in comparisons among Robust ELA, LPC, and REwarp are provided in the second to the last columns, respectively. For fair comparisons, we average overlapping regions to composite stitched images. As shown in the figures, we notice that REwarp resolves the hole generations and discontinuity between overlap and non-overlap regions without any additional prevention methods [10]. Under the natural scenes like the first, and last rows, the other baselines with REwarp show reasonable alignment. However, we see that our approach empirically provides more stable results. In this experiment, while Robust ELA and LPC show frequent failures under the images with few textures like the third row (stair), REwarp shows relaxed favorable results. The experiences support the performance gains in the next quantitative comparisons.

**Quantitative Comparison** In Tab. 1, we report photometric errors using mPSNR (valid region-only PSNR) considering that measurement with PSNR contains the contribution of invalid black regions. The reported times are measured by averaging repeated 100 inferences under NVIDIA RTX 3090 24GB (or AMD Ryzen7 4800H). The 'Failures' column provides the ratio of evaluation failures including no overlapping regions and huge image resolutions caused by unreasonably estimated warps. To observe the performances with the difficulty of image stitching, we additionally investigate the overlapping regions. Note that stitching under a low overlap ratio is a more challenging configuration compared to a high ratio. In the provided table, we see that the computation of REwarp is competitive for the existing image stitchings. In addition, the comparison between IHN and REwarp implies the effectiveness of our warps showing a 1.85dB performance gain. The comparison between Robust ELA and REwarp informs that the residual learning for misalign correction provides remark-
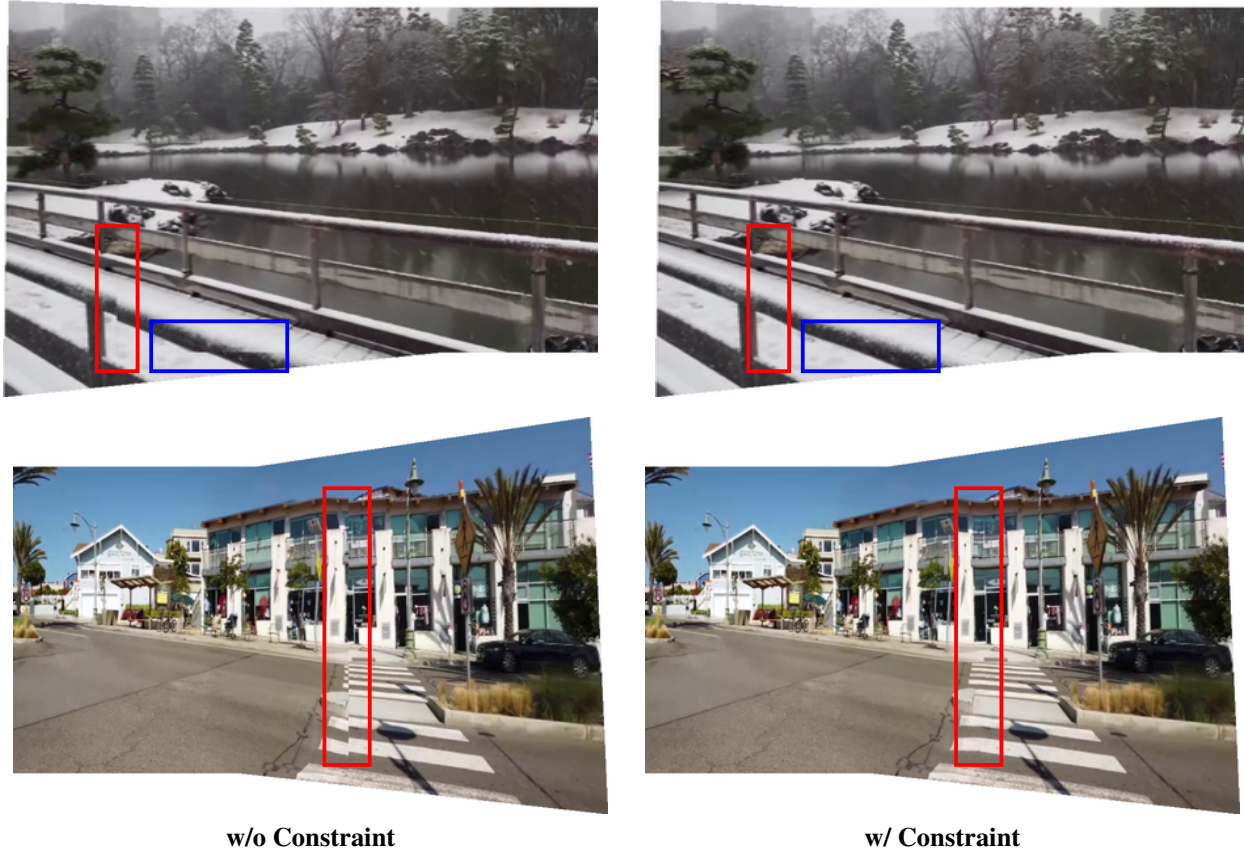
**w/o Constraint**　　　　　　　　　　　　　**w/ Constraint**

Figure 7. **Ablation Study on Dirichlet Boundary Condition.**

able performance gains. On the other hand, the failure ratios of structure-preserving image stitching [7, 15] indicate that their constraints may boost failure cases and cause harm to the overall aligning performances.

## 5.3. Ablation Study

We explore the contributions of three methods of RE-warp including iterative estimation, implementation approach for elastic warps, and boundary constraint. As a default model configuration, we fix the number of H-Cell and T-Cell iterations as 6 and 3, respectively.

**Iterative Estimation** In Tab. 2a, we investigate the relationship between the number of iterations and model performance. As shown in the table, H-Cell stably converges to a global optimum under model capacity while T-Cell shows divergence at 10 iterations. The result implies that the warp field's high degree provides limited model capacity for the extreme number of iterations.

**Implementation Approach for Elastic warp** In Tab. 2b, we explore methods that can be used for the implementation of elastic warp. As an approach for $\mathbf{H}$, we use our H-Cell. To prepare the estimator for $\mathbf{F}$, we simply revise the regressor of T-Cell to a regressor estimating $\mathbf{F} \in \mathbb{R}^{h \times w}$. Our investigation shows the effectiveness of the TPS-based im-

age alignment achieving a 0.41dB mPSNR gain compared to the flow-based warp.

**Boundary Constraint** For empirical observation that our constraint using Dirichlet Boundary Condition prevents RE-warp from the discontinuity between overlap and non-overlap regions, we explore qualitative comparisons in Fig. 7. In this study, we use linear blending considering that the guarantee of continuity makes the stitched image exposed to misalignment. As shown in the left column of Fig. 7, the flow addition in the overlap region of the target image ($\mathbf{I}_t$) causes discontinuity. However, our simple constraint resolves such a problem.

## 6. Discussion

**Property of Flow and TPS-based Alignment** Although flow-based deep image stitching enables highly flexible warps, it struggles with artifacts in low-frequency textures and discontinuity between overlap and non-overlap regions as in Fig. 8. In contrast, the TPS-based method provides relatively rigid flexibility compared to the method. Despite the property, TPS's non-linear and smooth warps with our boundary condition are free from the limitations of flow-based stitching satisfying most alignments for paral-
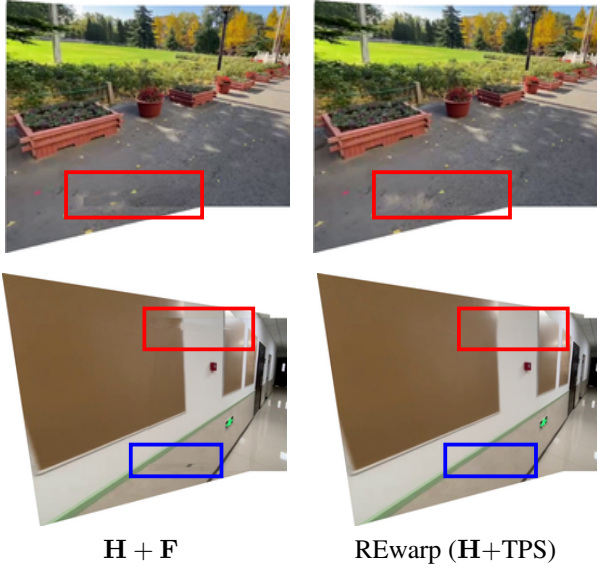
H + F  REwarp (H+TPS)

Figure 8. **Qualitative Comparison of Flow-based and TPS-based deep image stitching.**



LPC [7]



REwarp (ours)

Figure 9. **Comparison on Multiple image stitching.**



Figure 10. **Failure Case.**

lax. Though our constraint resolves the discontinuity, it may cause visually unnatural image stitching. Future works addressing the problem will be promising.

**Multiple image stitching** Sequential inferences with REwarp enable multiple image stitching. In Fig. 9, we demonstrate a qualitative comparison of multi-view stitching between LPC [7] and *ours*. To evaluate the image, we determine a reference image $\mathbf{I}_r$ and two target images $\mathbf{I}_{t_1}, \mathbf{I}_{t_2}$. Then we sequentially estimate transformations between three frames $\mathbf{I}_r$, $\mathbf{I}_{t_1}$ and $\mathbf{I}_r$, $\mathbf{I}_{t_2}$. Despite the relaxed discontinuity of our constraint, REwarp may be exposed to the discontinuity as in the upper crop of trucks in each figure, slightly.

**Under Large Parallax** In Fig. 10, we explore REwarp's stitching with images exposed to large parallax errors. As shown in the figure, it mostly guarantees continuity while providing visually unnatural image composition. Because TPS is a global transformation with smooth warps, our constraint may cause unnecessary warps in other local regions. To overcome this, a more advanced TPS warps with higher flexibility and C2 continuity in boundary would be required.

# 7. Conclusion

We proposed a Recurrent Elastic warp (REwarp) for parallax-tolerant deep image stitching, which recursively and sequentially estimates global and residual warps for misalign correction. REwarp resolves the hole generation and discontinuity issues of the previous work that leverages deep elastic warps and successfully demonstrated its image stitching on traditional real images and UDIS-D dataset quantitatively achieving 1.36dB mPSNR gain compared to Robust ELA. Furthermore, REwarp's fast computation and perfect success ratio in our experiment validated the potential usage for real-time image stitching.

# References

[1] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989. 3, 4

[2] Si-Yuan Cao, Jianxin Hu, Zehua Sheng, and Hui-Liang Shen. Iterative deep homography estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1879–1888, 2022. 6

[3] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation. *arXiv preprint arXiv:1606.03798*, 2016. 3

[4] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 6

[5] Junhong Gao, Seon Joo Kim, and Michael S Brown. Constructing image panoramas using dual-homography warping. In *CVPR 2011*, pages 49–56. IEEE, 2011. 1, 2

[6] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003. 2, 3

[7] Qi Jia, ZhengJun Li, Xin Fan, Haotian Zhao, Shiyu Teng, Xinchen Ye, and Longin Jan Latecki. Leveraging line-point consistence to preserve structures for wide parallax image stitching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12186–12195, 2021. 1, 2, 5, 6, 7, 8

[8] Kyungdon Joo, Namil Kim, Tae-Hyun Oh, and In So Kweon. Line meets as-projective-as-possible image stitching with moving DLT. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1175–1179. IEEE, 2015. 1, 2

[9] Vivek Kwatra, Arno Schödl, Irfan Essa, Greg Turk, and Aaron Bobick. Graphcut textures: Image and video synthesis using graph cuts. *Acm transactions on graphics (tog)*, 22(3):277–286, 2003. 1

[10] Hyeokjun Kweon, Hyeonseong Kim, Yoonsu Kang, Youngho Yoon, Wooseong Jeong, and Kuk-Jin Yoon. Pixel-wise deep image stitching. *arXiv preprint arXiv:2112.06171*, 2021. 1, 2, 4, 6

[11] Wei-Sheng Lai, Orazio Gallo, Jinwei Gu, Deqing Sun, Ming-Hsuan Yang, and Jan Kautz. Video stitching for linear camera arrays. *arXiv preprint arXiv:1907.13622*, 2019. 1

[12] Kyu-Yul Lee and Jae-Young Sim. Warping residual based image stitching for large parallax. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8198–8206, 2020. 2

[13] Jing Li, Zhengming Wang, Shiming Lai, Yongping Zhai, and Maojun Zhang. Parallax-tolerant image stitching based on robust elastic warping. *IEEE Transactions on multimedia*, 20(7):1672–1687, 2017. 1, 2, 5, 6

[14] Jiaxue Li and Yicong Zhou. Automatic color image stitching using quaternion rank-1 alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19720–19729, 2022. 1

[15] Tianli Liao and Nan Li. Single-perspective warps in natural image stitching. *IEEE transactions on image processing*, 29:724–735, 2019. 1, 2, 6, 7

[16] Tianli Liao and Nan Li. Natural image stitching using depth maps. *arXiv preprint arXiv:2202.06276*, 2022. 1

[17] Kaimo Lin, Nianjuan Jiang, Loong-Fah Cheong, Minh Do, and Jiangbo Lu. Seagull: Seam-guided local alignment for parallax-tolerant image stitching. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14*, pages 370–385. Springer, 2016. 1

[18] Wen-Yan Lin, Siying Liu, Yasuyuki Matsushita, Tian-Tsong Ng, and Loong-Fah Cheong. Smoothly varying affine stitching. In *CVPR 2011*, pages 345–352. IEEE, 2011. 1, 2

[19] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[20] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004. 6

[21] Lang Nie, Chunyu Lin, Kang Liao, Shuaicheng Liu, and Yao Zhao. Unsupervised deep image stitching: Reconstructing stitched features to images. *IEEE Transactions on Image Processing*, 30:6184–6197, 2021. 1, 6

[22] Lang Nie, Chunyu Lin, Kang Liao, Shuaicheng Liu, and Yao Zhao. Learning thin-plate spline motion and seamless composition for parallax-tolerant unsupervised deep image stitching. *arXiv preprint arXiv:2302.08207*, 2023. 1, 2, 4

[23] Lang Nie, Chunyu Lin, Kang Liao, and Yao Zhao. Learning edge-preserved image stitching from large-baseline deep homography. *arXiv preprint arXiv:2012.06194*, 2020. 1

[24] Dae-Young Song, Geonsoo Lee, HeeKyung Lee, Gi-Mun Um, and Donghyeon Cho. Weakly-supervised stitching network for real-world panoramic image generation. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVI*, pages 54–71. Springer, 2022. 1

[25] Dae-Young Song, Gi-Mun Um, Hee Kyung Lee, and Donghyeon Cho. End-to-end image stitching network via multi-homography estimation. *IEEE Signal Processing Letters*, 28:763–767, 2021. 1, 2

[26] Zachary Teed and Jia Deng. RAFT: Recurrent all-pairs field transforms for optical flow. In *European conference on computer vision*, pages 402–419. Springer, 2020. 4

[27] Jia Xu, René Ranftl, and Vladlen Koltun. Accurate optical flow via direct cost volume processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1289–1297, 2017. 1

[28] Julio Zaragoza, Tat-Jun Chin, Michael S Brown, and David Suter. As-projective-as-possible image stitching with moving DLT. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2339–2346, 2013. 1, 2, 6