RecRecNet: Rectangling Rectified Wide-Angle Images by Thin-Plate Spline Model and DoF-based Curriculum Learning

Kang Liao† Lang Nie† Chunyu Lin* Zishuo Zheng Yao Zhao
Institute of Information Science, Beijing Jiaotong University
Beijing Key Laboratory of Advanced Information Science and Network
{kangliao, nielang, cylin, yzhao}@bjtu.edu.cn

Abstract

The wide-angle lens shows appealing applications in VR technologies, but it introduces severe radial distortion into its captured image. To recover the realistic scene, previous works devote to rectifying the content of the wide-angle image. However, such a rectification solution inevitably distorts the image boundary, which changes related geometric distributions and misleads the current vision perception models. In this work, we explore constructing a win-win representation on both content and boundary by contributing a new learning model, i.e., Rectangling Rectification Network (RecRecNet). In particular, we propose a thin-plate spline (TPS) module to formulate the non-linear and non-rigid transformation for rectangling images. By learning the control points on the rectified image, our model can flexibly warp the source structure to the target domain and achieves an end-to-end unsupervised deformation. To relieve the complexity of structure approximation, we then inspire our RecRecNet to learn the gradual deformation rules with a DoF (Degree of Freedom)-based curriculum learning. By increasing the DoF in each curriculum stage, namely, from similarity transformation (4-DoF) to homography transformation (8-DoF), the network is capable of investigating more detailed deformations, offering fast convergence on the final rectangling task. Experiments show the superiority of our solution over the compared methods on both quantitative and qualitative evaluations. The code and dataset are available at https://github.com/KangLiao929/RecRecNet.

1. Introduction

The wide-angle lens has been widely applied in various applications such as computational imaging, virtual reality (VR), and autonomous driving, due to their larger field of view (FoV) than conventional lenses [19, 12, 44, 27]. By employing a specific mapping, the wide-angle lens can produce convex and non-rectilinear images rather than images with straight lines of perspective. However, this mapping mechanism violates the pinhole camera assumption, and the captured image suffers from severe radial distortions. To eliminate the distortion induced by the wide-angle lens, the distortion rectification algorithm is presented. For example, the traditional approaches calibrate the wide-angle lens using the calibration target (e.g., a checkerboard) [48, 19] and specific scene assumption [9, 2]. In recent years, deep learning has brought new inspirations to this field. It helps to automatically rectify the radial distortion in wide-angle images based on learned semantic features. To pursue better performance, the recent works improve the learning framework through diverse features [46, 43], learning paradigm [26], calibration representation [23, 29], and distortion prior [7, 25], etc.

We notice that the above distortion rectification ap-
proaches mainly focus on rectifying the image content. By shrinking the distorted image towards the principal center, the spatial distribution of all pixels has been rearranged and thus generates the rectified result. However, the image boundaries are inevitably distorted by this correction way, making the image exhibited with a visually narrow FoV and irregular shape. Moreover, we found such a deformed boundary can significantly influence the current mainstream vision perception models because most of them are trained on the dataset with rectangular images. As shown in Figure 1, Mask R-CNN [16] is confused by the rectified image, leading to a missing segmentation and detection result near the boundary (the surfboard is neglected). While some works [24] fill the blank region beyond deformed boundaries using outpainting strategy [35, 40], the generated contents are usually fictitious and twist the original semantic characterization.

In this work, we explore a win-win rectification representation for the wide-angle image on both content and boundary without introducing any extra semantic information [15, 32]. For this new representation, the distorted content has been corrected while the image boundary keeps straight, as shown in Figure 1. To this end, we propose a Rectangling Rectification Network (named RecRecNet). Concretely, a thin-plate spline (TPS) motion module is proposed to formulate the non-linear and non-rigid transformation for rectangling the rectified wide-angle image. It can flexibly warp the source structure to the target domain by learning the control points on the rectified image. The control points are located at the whole distribution of an image, while Figure 1 only shows a part of the points for intuitive clarification.

Subsequently, we design a DoF-based curriculum learning to further inspire our RecRecNet to grasp the progressive deformation rules, relieving the burden of complex structure approximation and initializing a better start point for training. The general paradigm of curriculum learning [3] introduces a heuristic and purposeful strategy for training a learning model. The model can converge faster by gradually learning various samples or knowledge based on different stages in a curriculum, mimicking how humans learn. In this regard, we increase the degree of freedom (DoF) of the devised stages/transformations from similarity transformation (4-DoF) to homography transformation (8-DoF). Moreover, the procedure of our curriculum reveals a simple-to-complex order from the straight line, sloped line to curve for the image boundary. As a consequence, our RecRecNet can discover more geometric details and converge faster on the final rectangling stage.

While aiming for a win-win rectification representation, we also provide an in-depth analysis of why the deformed boundary can introduce new features on the original feature maps, which further forms new semantics or cause blind spots for non-salient features. And we demonstrate that this effect commonly exists in other research regions.

We conclude our major contributions as follows:

- We propose a Rectangling Rectification Network (RecRecNet) for a win-win representation of the rectified wide-angle image. To our knowledge, it has never been investigated in previous literature.
- A thin-plate spline (TPS) motion module is proposed to flexibly formulate the non-linear and non-rigid rectangling transformation. We also provide an in-depth analysis of why the deformed image boundary can significantly influence the visual perception models.
- To grasp the progressive deformation rules and relieve the burden of complex structure approximation, a DoF-based curriculum learning is designed to inspire our RecRecNet.

2. Related Work

2.1. Distortion Rectification

The distortion rectification approaches for wide-angle lenses can be classified into traditional and learning-based approaches. The majority of traditional techniques concentrate on detecting the hand-crafted features [30, 21, 9, 2]. However, these techniques frequently fail due to particular constraints like the plumb line and curve. Recently, blind distortion rectification has been studied using deep learning. It can rectify the radial distortion of the wide-angle image without any scene assumptions. For example, the regression-based solutions [34, 4, 46, 29, 25] predict the distortion parameters of input, and then the estimated parameters are used to rectify the distortion offline or online. The reconstruction-based solutions [26, 23, 45, 7] directly learns the pixel-level mapping or displacement field from the distortion image into the rectification output.

Despite promising flexibility and efficiency, these rectification approaches inevitably distort the image boundary during content correction. As a result, the distorted distribution at the boundary confuses the vision perception network and leads to inferior performance. One direct way is to crop the rectified image into a rectangular shape [14], however, it discards informative content and disobeys the original intention of the wide-angle lens.

2.2. Curriculum Learning

Curriculum learning techniques have been effectively exploited across the board in machine learning. Elman et al. [10] provide a conceptual framework for curriculum
learning that emphasizes the value of beginning at a low level and then works way up to more challenging cases. Bengio et al. [3] are the first to formalize curriculum learning as a strategy to gradually raise the level of complexity of the data samples utilized in the training process. Subsequently, most researchers follow this paradigm and apply the curriculum learning into numerous regions [37, 41, 33]. Particularly in computer vision, curriculum learning has been investigated to inspire the learning networks on image classification [6, 31, 13], object detection [22, 42, 38], and semantic segmentation [47, 36, 11], etc.

In this work, we incorporate curriculum learning into a new task. By disassembling the rectangling rectification into three levels based on the DoF of transformation, the proposed curriculum can inspire our network to learn the challenging deformation in a simple-to-complex order and thus facilitates a faster convergence for training.

3. Methodology

3.1. Problem Formulation

Given a rectified wide-angle image $I_{Rec} \in \mathbb{R}^{h \times w \times 3}$ with deformed boundaries, our Rectangling Rectification Network (RecRecNet) aims to construct a win-win representation on both image content and image boundary. RecRecNet learns a flexible local transformation based on the thin-plate spline (TPS) motion model, which predicts the control points on the image boundary and warps the target image by using a mesh grid. It is jointly optimized with the appearance constraint $L_{AP}$, perceptual constraint $L_{PE}$, and inter-grid constraint $L_{IG}$. To further relieve the challenges of structure approximation, we inspire our RecRecNet to learn the gradual deformation rules with a DoF-based curriculum learning.

The general polynomial camera model [19] is widely used in the approximation for the radial distortion of the wide-angle image. We formulate the relationship between the lens’s projection manner and its distortion parameters $K = [k_1, k_2, k_3, \cdots]$ as follows:

$$r(\theta) = \sum_{i=1}^{N} k_i \theta^{2i-1}, N = 1, 2, 3, \cdots, \quad (1)$$

where $r$ is the distance between the principal point and the pixels in the image, and $\theta$ indicates the angle between the optical axis and incident ray. Such a projection can produce a convex and non-rectilinear image, and thus the captured image suffers from severe geometric distortions. Suppose that the coordinates of a pixel in the wide-angle image and
rectified images are \( p = (x, y) \) and \( p_{rec} = (x_{rec}, y_{rec}) \), the rectification solution can be then described as follows:

\[
p_d = \mathcal{R}(p_{rec}, \mathcal{K}) = \left( \frac{u_0}{v_0} \right) + \frac{r(\theta) p_{rec}}{\|p_{rec}\|_2},
\]

where \((u_0, v_0)\) represents the principal point. With the above equation, the wide-angle image can be rectified by using bilinear interpolation. By shrinking the wide-angle image towards the principal center, the spatial distribution of all pixels is rearranged, and thus rectified content is obtained. However, the image boundaries are inevitably distorted by this rectification way.

### 3.2. Rectangling with TPS Transformation

As shown in Eq. 2, the rectification of the wide-angle image is a non-linear and non-rigid mapping function, which cannot be represented by simple transformations. Besides, it is challenging to build an accurate parameterization model to supervise the rectangling process with estimated parameters. From the perspective of energy minimization, we can approximate the rectangling rectification by minimizing the transformation energy with a uniform model:

\[
\epsilon = \varepsilon_T + \lambda \varepsilon_d,
\]

\[
\mathcal{T} : (x_{rec}, y_{rec}) \mapsto (x_{rec'}, y_{rec'}),
\]

where \(\varepsilon\) denotes the total energy of the expected transformation. \((x_{rec}, y_{rec}) \in \Omega_S\) and \((x_{rec'}, y_{rec'}) \in \Omega_T\) represent the point in the domain of source \(S\) and target \(T\), respectively. \(\varepsilon_T\) indicate the data penalty energy and \(\varepsilon_d\) indicate the distortion energy. The above formula with the lowest overall energy is the desired transformation, in which a hyperparameter \(\lambda\) is designed to balance the energies between the data penalty and distortion.

Particularly, we propose to exploit TPS transformation \(\mathcal{T}\) to minimize the aforementioned energy function, which enables an end-to-end unsupervised deformation. TPS transformation provides for the representation of more complex motions since it is flexible and non-linear. It allows warping from one image to the other with minimum distortion, given two sets of control points in the corresponding images. Suppose \( q = [q_1, q_2, \cdots, q_N] \) and \( q' = [q'_1, q'_2, \cdots, q'_N] \) are the source points and the target points, the data term \(\varepsilon_T\) can be calculated by the difference of control points displacement \(\sum_{i=1}^{N} \| \mathcal{T}'(q_i) - q'_i \|^2 \). Having aligned the control points, TPS finds an optimal interpolation transformation with minimal distortion term \(\varepsilon_d\) by:

\[
\min \iint_{\mathbb{R}^2} \left( \frac{\partial^2 \mathcal{T}'}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 \mathcal{T}'}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 \mathcal{T}'}{\partial y^2} \right)^2 \, dx \, dy.
\]

The above equation utilizes second-order derivatives to formulate the distortion deviation of each target point and constrains a cumulative global minimum. Then we can reach a spatial deformation function parameterized by control points as follows:

\[
\mathcal{T}'(q) = A \left[ q \right] + \sum_{i=1}^{N} w_i U (\| q'_i - q \|_2),
\]

where \(q\) is a point located at the rectified wide-angle image. \(A \in \mathbb{R}^{2 \times 3}\) and \(w_i \in \mathbb{R}^{2 \times 1}\) are the transformation parameters [5], which can be derived by minimizing Eq. 3. Besides, \(U(r)\) is a radial basis function that denotes the influence of the control point on \(q\): \(U(r) = r^2 \log r^2\).

In RecRecNet, we define \(N = (U + 1) \times (V + 1)\) control points that can be connected to form a mesh, and the network learns to predict the mesh of rectified wide-angle images. Then TPS transforms the predicted mesh to the regular mesh on the target rectangular image, where the control points are set to be evenly distributed on the image.

For the network, we first choose ResNet50 [17] as the backbone of our RecRecNet to extract the high-level semantic features. Suppose the size of the input rectified image is \(W \times H\), and the size of the output feature map of the backbone is \(W' \times H'\). Then these feature maps are fed to a motion header to predict \((U + 1) \times (V + 1)\) control points on the rectified image. 4 convolutional layers with a kernel size of 3 and BatchNormalization are stacked to aggregate the high-level features, of which all channel dimensions are set to 512. Subsequently, 3 fully connected layers with the following numbers of units 4096, 2048, \((U + 1) \times (V + 1) \times 2\) are used to predict the coordinates of all control points.

### 3.3. DoF-based Curriculum Learning

As mentioned above, we propose to exploit TPS transformation to flexibly approximate the motion from the rectified image to rectangling result by a set of control points. To further relieve the challenges of this transformation, we inspire our RecRecNet with a curriculum learning strategy.

#### 3.3.1 Image Transformation Formulation

Image transformation contains different elements such as translations, scales, rotations, shearing, etc. Any composition of these elements can form a new transformation with the degree of freedom (DoF). In general, the degree of freedom is a collection of independent displacements that completely describes a system’s displaced or deformed position. There are some classical transformations on 2D images: translation transformation (2-DoF), Euclidean/Rigid transformation (translation + rotation, 3-DoF), similarity transformation (translation + rotation + scale, 4-DoF), affine transformation (translation + rotation + scale + shear, 6-
DoF), and projective transformation (two each for translation, rotation, scale, and lines at infinity, 8-DoF).

The image transformation between the rectified wide-angle image and its rectangling image has never been studied and formulated. Compared to distortion rectification, the rectangling transformation seems to be an inverse asymmetric warping due to the more important image boundary. To understand this transformation intuitively, we visualize the pixel displacement field between the rectified wide-angle image and its rectangling image. Specifically, we leverage the state-of-the-art optical flow estimation network RAFT [39] to exhibit the motion difference and name it as rectangling flow (rectified image $\rightarrow$ rectangling image). As shown in Figure 3, compared with the rectification flow (wide-angle image $\rightarrow$ rectified image), we have the following observations: (1) Both rectangling flow and rectification flow show a radial spatial distribution, in which four dominated directions spread towards or backward four image corners. (2) Rectangling flow has an unfixed radial center, which is related to the geometry prior to the radial distortion in the wide-angle image.

Thus, we can conclude the image content highly identifies the distribution of rectangling flow. It is painful to represent this transformation using classical image transformations and their compositions. Nevertheless, such a non-linear and non-rigid transformation can be locally disentangled into different transformation elements such as translation, scale, and shearing. It would be meaningful to inspire the network to learn some basic transformations and gradually transfer them into more challenging cases.

3.3.2 Curriculum Selection

To choose effective curriculums for the rectangling task, we should obey two principles: First, all stages in a curriculum have closely associated knowledge of the target. Second, the stages in a curriculum intrinsically offer a simple-to-complex order. To this end, we design a DoF-based curriculum to inspire RecRecNet, of which we increase the DoF of the devised transformations from similarity transformation (4-DoF) to homography transformation (8-DoF) and eventually to the rectangling transformation.

We argue that the proposed curriculum has the following strengths: (1) As mentioned in Section 3.3.1, the DoF directly expresses the difficulty of the image transformation. As the DoF increases, the transformation shows more complex global mappings and diverse motions. Thus, RecRecNet can learn the gradual deformation rules. (2) For the local geometry, i.e., the shape of the image boundary, our curriculum shows a progressive shift from a straight line, sloped line, to a curve. Such a design can improve the localization ability of control points in TPS transformation and enable a boundary-aware ability for RecRecNet. More details and experiments will be demonstrated in Section 4.4.

3.4. Training Losses

Appearance Loss: We constrain the rectangling image $I_{Rec}^{i,j}$ to be close to the ground truth $I_{GT}^{i,j}$ at the pixel level. The appearance loss $L_{AP}$ can be calculated by the difference between $I_{Rec}^{i,j}$ and $I_{GT}^{i,j}$ with $L_1$ norm.

Perceptual Loss: We then minimize the $L_2$ distance between $I_{Rec}^{i,j}$ and $I_{GT}^{i,j}$ in high-level semantic perception to guarantee the rectangling results perceptually natural:

$$L_{PE} = \frac{1}{W_{i,j}H_{i,j}} \sum_{z=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} \| \phi_{i,j}^{x,y}(I_{Rec}^{i,j}) - \phi_{i,j}^{x,y}(I_{GT}^{i,j}) \|_2. \tag{6}$$

where the difference of rectangling rectified image and ground truth is minimized on the feature map $\phi_{i,j}$, which is derived from the $j$-th convolution (after activation) before the $i$-th max-pooling layer in the VGG19 network.

Inter-grid Mesh Loss: To avoid the content distortion after our rectangling, the predicted mesh should not be largely deformed. Therefore, we design an inter-grid mesh term $L_{IG}$ to constrain the shape of the predicted mesh, which encourages the neighboring grids to transform consistently. The edges of two successive deformed grid $\{\tilde{e}_{i1}, \tilde{e}_{i2}\}$ are supervised to be co-linear as follows:

$$L_{IG} = \frac{1}{M} \sum_{\{\tilde{e}_{i1}, \tilde{e}_{i2}\} \in m_s} (1 - \frac{\langle \tilde{e}_{i1}, \tilde{e}_{i2} \rangle}{\| \tilde{e}_{i1} \| \cdot \| \tilde{e}_{i2} \|}), \tag{7}$$

where $M$ is the number of tuples of two successive edges in a mesh $m_s$. When maximizing the above cosine representation, i.e., the value equals 1, and the corresponding two edges are co-linear. Thus, the loss reaches the minimum and the image content exhibits consistently.

The overall training loss can be obtained by:

$$L = \lambda_{AP} L_{AP} + \lambda_{PE} L_{PE} + \lambda_{IG} L_{IG}. \tag{8}$$

Figure 3. Visualization of the rectangling flow and rectification flow for the rectified wide-angle image. We mark the different radial centers in different rectangling flows with arrows.
4. Experiments

4.1. Implementation Details

**Dataset Establishment:** Following existing methods [34, 46, 26, 45], we first synthesize the wide-angle images by using a $4^{th}$ order polynomial model based on Eq. 1. Then we perform the distortion rectification on the wide-angle image in terms of Eq. 2. Due to non-linear and non-rigid characteristics, forming an accurate transformation model for rectangling the rectified image is difficult. We notice that there is a classical panoramic image rectangling work by He et al. [15] on computer graphics. It makes the stitched image regular by optimizing an energy function with line-preserving mesh deformation. Thus, we perform the same energy function on the rectified image to fit our task. However, the capability to preserve linear structures [15] is limited by line detection. Consequently, some rectangling images have non-negligible distortions. To overcome this issue, we carefully filter all results and repeat the selection process three times, resulting in 5,160 training data from 30,000 source images and 500 test data from 2,000 source images. Each manual operation takes around 10s. More details about the dataset are reported in the supplementary material. We would like to release the dataset to promote the development of the related community.

**Experimental Configuration:** We train RecRecNet with an exponentially decaying learning rate with an initial value of $10^{-4}$ using Adam[20]. The batch size is set to 16 and the total epoch is set to 260. Particularly, RecRecNet is trained with 3 stages in terms of the proposed DoF-based curriculum. The training epochs of the 4-DoF curriculum, 8-DoF curriculum, and final rectangling curriculum are empirically divided into 30, 50, and 180.

4.2. Rectangling Rectification Results

RecRecNet is the first effort to rectangling the rectified wide-angle image. We compare our method with the classical image rectangling work He et al. [15] and the state-of-art outpainting work for rectified image [24]. In the first comparison (Figure 4), we split the test dataset into the complex scene and simple scene based on the image feature. We also show the cropped result that directly discards the content.
around the deformed boundary, which is commonly applied in previous works. Experimental results demonstrate while RecRecNet is trained based on the filtered labels from He et al. [15], it can learn a higher upper bound of performance and generalize to more diverse scenes. By contrast, He et al. [15] fail to rectangling the scene with a complicated layout or monotonous background due to line-preserving rules. In the second comparison (Figure 5), we can observe ROP [24] makes the deformed boundary of the rectified image regular by extrapolating the image content. But the generated results are usually fictitious and twist the original semantic characterization, in which an unpleasant fracture of semantics occurs near the boundary. Without introducing new and ambiguous content, our RecRecNet constructs a win-win representation for the rectified image using a flexible TPS transformation, showing the closest appearances to the rectangling label. Besides, Figure 6 shows RecRecNet can considerably help the downstream vision tasks.

### 4.3. Why Deformed Boundary Makes Perception Deformed

While our task is to rectangling the rectified image, we are still curious about the secret behind the performance degradation of the vision perception models. In this part, we provide an in-depth analysis of this effect. First, two typical failure cases of the perception model are formulated: wrong perception and missing perception. Concretely, we focus on the detection performance of Mask R-CNN [16] and visualize some samples in Figure 6 and Figure 7. We can observe the model fails to detect the objects, especially those located near the boundary. By rectangling the deformed boundary, the perception performance is surprisingly restored.

In experiments, we found the deformed boundary can introduce new features onto the feature maps, which (i) form new semantics (leads to the wrong perception) or (ii) cause blind spots for original features (leads to missing perception). Especially in Figure 7, we exhibit the feature maps of Mask R-CNN with ResNet101 [17] backbone from shallow layers to deep layers. The noticeable line artifacts and curve artifacts can be observed at the outermost boundaries and deformed boundaries, respectively. In particular, the line artifacts are essentially generated by zero padding [1]. When a convolutional kernel extracts the feature at the boundary with zero padding, the sharp transitions between the zero values and the original content are wrongly identified as an edge. Similarly, this edge effect can be induced by the deformed boundary of rectified images. As the convolutional layer increases, such an effect is gradually enlarged and constructs new features on the feature maps.

For the wrong perception case, the network mistakenly recognizes the boundary and its surrounding region as a surfboard. Since the edge effect introduces new features on the shallow feature maps and new semantics is formed...
Table 2. Ablation study on the size of the control points in TPS.

<table>
<thead>
<tr>
<th>Size</th>
<th>PSNR ↑</th>
<th>SSIM ↑</th>
<th>FID ↓</th>
<th>LPIPS ↓</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 × 3</td>
<td>17.96</td>
<td>0.5134</td>
<td>25.00</td>
<td>0.1347</td>
</tr>
<tr>
<td>5 × 5</td>
<td>18.41</td>
<td>0.5288</td>
<td>21.49</td>
<td>0.1221</td>
</tr>
<tr>
<td>7 × 7</td>
<td>18.59</td>
<td>0.5385</td>
<td>19.62</td>
<td>0.1153</td>
</tr>
<tr>
<td>9 × 9</td>
<td>18.68</td>
<td>0.5450</td>
<td>19.01</td>
<td>0.1136</td>
</tr>
<tr>
<td>11 × 11</td>
<td>18.64</td>
<td>0.5392</td>
<td>19.16</td>
<td>0.1139</td>
</tr>
</tbody>
</table>

Figure 9. Ablation study on the curriculum learning for rectangling task. C1, C2, and C3 denote different curriculums.

in the deep feature maps. More results are reported in the supplementary material. Furthermore, the deformed boundaries not only exist in the rectified image, but also in other research regions such as image warping, image stitching, and roll shutter correction. Thus, the edge extension effect can introduce new features onto feature maps and further influence the downstream perception performance.

4.4. Ablation Study

**Control Points:** As shown in Table 2 and Figure 8, the experiments demonstrate more control points can facilitate the structure approximation and localization ability of RecRecNet for the deformed boundaries. And the size of 9 × 9 achieves the best performance on all evaluation metrics.

**Curriculum Learning:** We define RecRecNet without a curriculum as Vanilla and define RecRecNet with different curriculums: 2-DoF → 4-DoF as Vanilla + C1, 2-DoF → 8-DoF as Vanilla + C2, and 4-DoF → 8-DoF as Vanilla + C3, respectively. In Figure 9 (left), the training loss curves show that curriculum learning can facilitate a better initialization point for the training and faster convergence on the rectangling task. Vanilla training took around 40 more epochs to converge to a similar training performance as curriculum settings, but it is prone to overfitting on validation data. Moreover, a reasonable design for the curriculum (such as Vanilla + C3) helps to improve the boundary-aware ability of RecRecNet, enhancing the structure recovery performance as shown in Figure 9 (right).

4.5. Cross-domain Evaluation

To evaluate the generalization, we collect 300 rectified wide-angle image results from the state-of-the-art rectification methods [25, 45]. Their results are derived from various types of datasets and real-world wide-angle lenses such as the Rokinon 8mm Cine Lens, Opteka 6.5mm Lens, and GoPro. These lenses are usually used to obtain the panoramic image. Figure 10 shows RecRecNet can well generalize to other domains and different rectified structures while it trained using only a synthesized image dataset with one type of camera model. We also show some comparison results as shown in Figure 11, where fractured semantics and distorted object can be observed in previous methods.

5. Conclusion

In this work, we consider a new rectangling rectification task for the wide-angle image. While it has never been studied in previous literature, we demonstrate that the deformed boundary of the prevalent rectified wide-angle image can significantly influence the vision perception models. To eliminate this issue, we present to construct a win-win representation for the rectified wide-angle image and design a novel RecRecNet. Equipped with a flexible TPS transformation motion model, RecRecNet can formulate the local deformation from the deformed boundary to the straight one in an unsupervised end-to-end manner. Besides, we inspire RecRecNet to learn the gradual deformation rules with a DoF-based curriculum learning, which can relieve
the complexity of non-linear and non-rigid transformation. Furthermore, a detailed analysis is provided to explain why the deformed image boundary makes the current vision perception deformed. In future work, we plan to extend to a general paradigm for rectangling any deformed images and further study the relationship between the image boundary and vision perception performance. Moreover, embedding our rectangling algorithm into an online data augmentation for training vision models would be also interesting.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (No.62172032, No.62120106009).

References


[24] Kang Liao, Chunyu Lin, Yunchao Wei, Feng Li, Shangrong Yang, and Yao Zhao. Towards complete scene and regular shape for distortion rectification by curve-aware extrapolation. In Proceedings of the IEEE/CVF International Confer-


Rainer Stiefelhagen. Capturing omni-range context for omnidirectional segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (ECCV)*, September 2018. 1, 2, 6