Image stitching using gradual image warping in autonomous driving

Christian Kinzig, Jiang Yifan, Martin Lauer, and Christoph Stiller

Karlsruhe Institute of Technology (KIT),
Institute of Measurement and Control Systems,
Engler-Bunte-Ring 21, 76131 Karlsruhe, Germany

Abstract To improve object recognition and tracking in autonomous driving, we first create a seamless panorama. Object recognition can benefit from image stitching, especially at the borders of individual images when an object is only partially visible. This also prevents duplicate detection of the same objects in overlapping image areas that are to be filtered for tracking. In this process, a homography is determined for the overlapping image area, whereby the entire image is transformed using classical image stitching methods. As a result, the deformations propagate to further images that are to be added to the panorama. To avoid this problem, we integrated a step-by-step image warping approach into our existing stitching pipeline. This ensures that after attaching one image to another, the outermost right and left borders of the panorama are no longer deformed. Furthermore, the panorama width remains constant regardless of the calculated homography. We have evaluated our approach on the nuScenes dataset and the Waymo Open Dataset for perception. In addition to a qualitative assessment, we evaluate the resulting panoramas in terms of the deformation of the individual images as well as the deformation of labeled object instances.

Keywords Autonomous driving, panorama, image stitching, homography, warping, deformation

Acknowledgements This research is accomplished within the project UNICAR*agil* (FKZ 16EMO0287). We acknowledge the financial support for the project by the Federal Ministry of Education and Research of Germany (BMBF).

DOI: 10.58895/ksp/1000174496-19 erschienen in:
Forum Bildverarbeitung 2024
DOI: 10.58895/ksp/1000174496 | https://www.ksp.kit.edu/site/books/m/10.58895/ksp/1000174496/

1 Introduction

The UNICAR*agil* [1,2] project, in which four autonomous vehicles were built entirely from scratch, investigated how and whether camera images should be stitched together to form a panorama before object recognition. One of the resulting articles [3] shows that object recognition performs just as well on panoramic images as on individual images without the need for retraining. In addition, in another article [4] we demonstrate that object detection on panoramic images improves compared to single images after retraining in this domain.

To stitch two images together, in a simple procedure, a homography is determined between pairs of feature matches in the overlapping image area to transform one of the images. However, this procedure for stitching images has the disadvantage that the resulting deformations increase with each additional images added to the panorama. For this reason, we have implemented a gradual image warping method based on the approach in [5]. Our main contribution is the elimination of deformations at the outermost right and left borders of the panorama, allowing any number of images to be stitched together horizontally. In this way, the transformations of all individual images can be calculated independently of each other. At the same time, the resulting panorama has a constant image width, which makes it more suitable as training data, as less zero padding needs to be applied. Furthermore, we decided to realize the local alignment not as a grid but as vertical image slices in order to reduce the computational effort.

2 Related Work

In the work by Zaragoza et al. in [6], a global homography between two images is first estimated, then equally sized grid cells in the image are transformed by local homographies to improve the alignment of the images to each other. In contrast, Chang et al. introduced a three-step process to preserve perspective by combining transformations from homography and similarity transformation in [7]. Based on this, Xiang et al. achieve smoother transitions by using weighted combinations of homography and similarity transformation in [5]. Chen and Chuang specifically aim for natural image stitching in [8] by using APAP [6] in

combination with a global similarity transformation to adjust scale and rotation for each image to be stitched. In [9], Zhang et al. developed a method specifically designed to return a rectangular panoramic image to reduce deformations in the images.

3 Implementation

The presented approach extends our image stitching method presented in [3] and [4]. Our image stitching pipeline is shown in Fig 1, with the modifications highlighted in blue. Thus, the deformation of the panorama towards the outermost right and left borders is gradually eliminated. The core components for gradual image warping can be divided into three consecutive steps, where first a homography in overlapping image areas is determined. In the second step, we divide the camera image into vertical sections and determine a transformation for each part of the image from the resulting homography. In the last step, we apply the resulting transformations to each image section and combine them to create a panorama.

3.1 Homography Estimation

The homography between two individual images p and q is determined by features in the overlapping image area. Consequently, the transformation of a feature point in a camera image p into another image q is given by (1).

$$\mathbf{H} \begin{pmatrix} u_p \\ v_p \\ 1 \end{pmatrix} = s \begin{pmatrix} u_q \\ v_q \\ 1 \end{pmatrix}, s \in \mathbb{R} \tag{1}$$

$\mathbf{H}$ stands for the homography and s for the scaling factor. As in [3], we do not perform feature extraction as well as subsequent feature matching. Instead, we use depth information as in a LiDAR point cloud, which we project into the overlapping image areas. Compared to image features, however, we have the disadvantage that an error occurs when projecting into cameras due to the parallax and the rotation of the LiDAR. This error can be reduced if the cameras are triggered as

soon as the LiDAR points in their direction. The Waymo Open Dataset for perception [10] also provides synchronized LiDAR data regarding the movement of the ego vehicle and the movement of other traffic participants. To calculate the homography, we use a method based on the RANSAC algorithm. In addition, learned methods as in [11] to determine a homography between two images would also be possible.

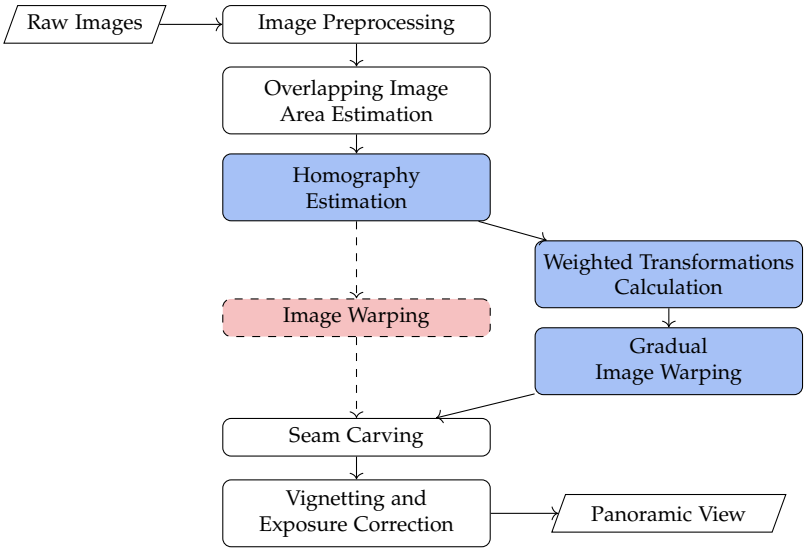

Figure 1: Workflow of our modifications shown in blue to the image stitching pipeline in [3] and [4] by integrating gradual image warping. Consequently, the image warping module shown in red is replaced.

3.2 Weighted Transformations Calculation

Using the homography determined in 3.1, the respective overlapping area is transformed. The opposite overlapping image area adjacent to the next camera image is not transformed. If there is no further camera image to be stitched, we assume a quarter of the image width that is not transformed. The remaining image area in between is warped gradually. First, a configurable parameter k, is used to define how many vertical image sections the center image area is divided into. This

determines how well the individual sections merge into one another. Similar to the approach in [5], we determine weighted transformations for each image section from two individual transformations, as shown in (2).

$$\mathbf{T} = \alpha\mathbf{H} + \beta\mathbf{I}_3 \tag{2}$$

However, we use the homography $\mathbf{H}$ determined in 3.1 and the identity matrix $\mathbf{I}_3$. For the first vertical image section adjacent to the overlapping image area associated with the homography, $\alpha = 1 - \frac{1}{k+1}$ and $\beta = \frac{1}{k+1}$. As the horizontal distance to the overlapping area increases, β gradually increases and α gradually decreases, so that the last vertical image section matches the overlapping area at the far end of the image. This allows any number of camera images to be stitched together horizontally without the deformations in the panorama becoming progressively larger towards the outside. In addition, the identity means that smoother video sequences can be created from the panoramas with a constant image width.

3.3 Gradual Image Warping

Once the transformation matrices have been determined for each individual image section, the image can be warped gradually. However, the transformations are still in image coordinates (u, v). In order to process smaller amounts of data and thus improve the runtime, we first transform each matrix $\mathbf{T}$ into the coordinate system of the respective image section (u_i, v_i) and denote the resulting transformation matrix as $\mathbf{T}_i$. The transformation into the vertical image sections can be described by a translation Δu. Finally, each transformation $\mathbf{T}_i$ is applied to the corresponding image section. These are subsequently projected onto the overall panoramic image.

4 Evaluation

First, we qualitatively compare our approach of gradual image warping with our previous approach as baseline described in [3] and [4].

Furthermore, we evaluate our approach also in comparison to our previous method using a quantitative measure of image deformation. In addition, we separately compare the deformations of labeled object instances. The two publicly available datasets nuScenes [12] and the Waymo Open Dataset for perception [10] are used in our evaluation.

(a) Individual images from which the panorama is composed using a spherical camera model.

(b) Image stitching using the method in [3] and [4].

(c) Image stitching with gradual image warping.

Figure 2: Comparison on image stitching using data from the nuScenes dataset [12].

4.1 Qualitative Comparison

To give an first impression of how our method performs, we use the two panoramas in Fig. 2 and 3 to show the comparison with the use of a homography per overlapping area. Fig. 2 compares both methods using an example from the nuScenes dataset [12] whereas Fig. 3 uses data from the Waymo Open Dataset for perception [10]. Both figures show that gradual image warping can better compensate for strong deformations. This applies in particular to the images from the outermost cameras. The curvature at the top and bottom of the images is due to the use of a spherical camera model, which can be seen in Fig. 2(a)

and 3(a). Particularly with video sequences, strong deformations are noticeable as jumps in the panoramas, as these do not remain constant. Gradual image warping ensures that the deformations are substantially smaller and more consistent. This could improve object tracking, especially if it is assumed that a detected object is in a similar position in the subsequent panoramic image.

(a) Individual images from which the panorama is composed using a spherical camera model.

(b) Image stitching using the method in [3] and [4].

(c) Image stitching with gradual image warping.

Figure 3: Comparison on image stitching using data from the Waymo Open Dataset [10].

4.2 Image Deformation Evaluation

To quantitatively evaluate gradual image warping, we determine the deformations in the warped images compared to the original images. In this case, the term original image refers to images that have already been processed but not warped for image stitching. Pre-processing consists of compensating for lens distortion and converting the image from a pinhole camera model to a spherical camera model.

To measure the deformation, we analyze points $\mathbf{p}_i$ evenly distributed over the images with a distance of 20 pixels. Then these points are

deformed to $\mathbf{p}_{warped,i}$ by image warping either with a single homography or with gradual image warping. First, we determine the average displacement $\overline{\mathbf{d}}$ between all N points in the deformed image and those in the original image, since a constant translation has no influence on the deformation. Accordingly, the average displacement $\overline{\mathbf{d}}$ is calculated separately for the directions u and v in (3).

$$\overline{\mathbf{d}} = \begin{pmatrix} \overline{d}_u \\ \overline{d}_v \end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{p}_i - \mathbf{p}_{warped,i} \tag{3}$$

We then calculate the displacement between the points in the original images and in the warped images, taking into account the average displacement. This results in our error metric $\mathbf{E}_i$ in (4).

$$\mathbf{E}_i = \begin{pmatrix} E_{u,i} \\ E_{v,i} \end{pmatrix} = \left| \mathbf{p}_i - \mathbf{p}_{warped,i} - \overline{\mathbf{d}} \right| \tag{4}$$

The evaluation of the image deformation is performed on 10 sequences of the nuScenes dataset [12] and on 6 sequences of the Waymo Open Dataset for perception [10]. This results in an evaluation of 404 panorama images for nuScenes and 551 for Waymo. The results are displayed as two-dimensional box plots in Fig. 4 for the nuScenes dataset [12] and in Fig. 5 for Waymo Open Dataset for perception [10]. Both graphs clearly show that the deformations for gradual image warping are much smaller compared to the use of a single homography. The difference in deformation is most noticeable in the u direction. The smaller parallax in the Waymo Open Dataset results in significantly reduced warping on average. However, the outliers to the maximum are also higher in this case. The reason for this are the motion-compensated lidar point clouds. With high ego velocity or fast moving objects in the overlapping image area, significantly fewer point correspondences are available to calculate a homography.

4.3 Object Instances Deformation Evaluation

Especially in object recognition with machine learning, it is crucial that the results obtained on datasets can also be reproduced in the real

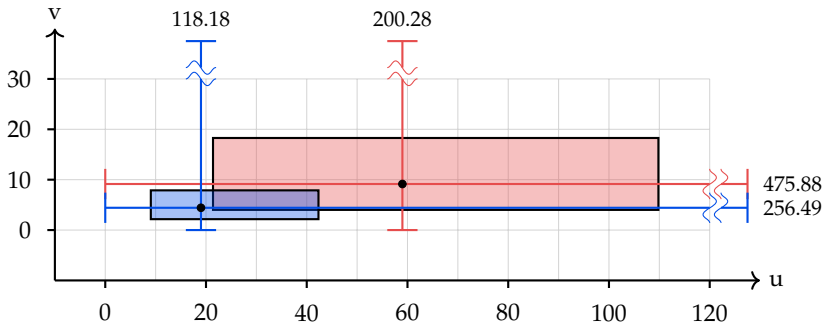

Figure 4: Image deformation analysis over 10 sequences of the nuScenes dataset [12]. 2D box plot of the deformations of the individual images in u- and v-direction in pixels with the method in [3] and [4] (red) compared to gradual image warping in (blue).

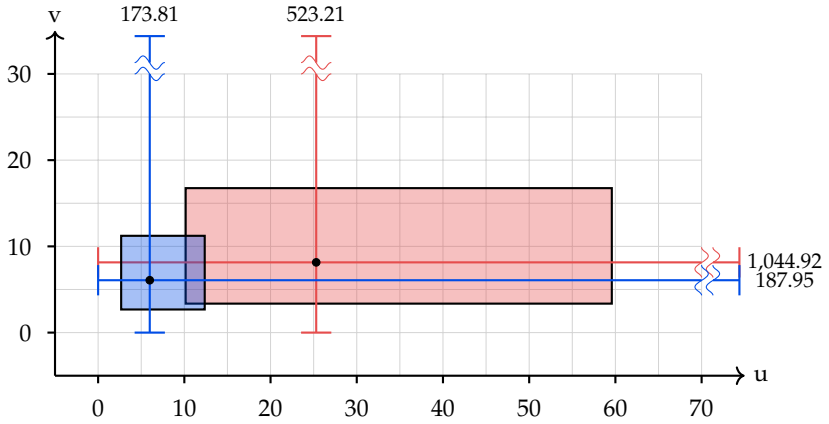

Figure 5: Image deformation analysis over 20 sequences of the Waymo Open Dataset for perception [10]. 2D box plot of the deformations of the individual images in u- and v-direction in pixels with the method in [3] and [4] (red) compared to gradual image warping in (blue).

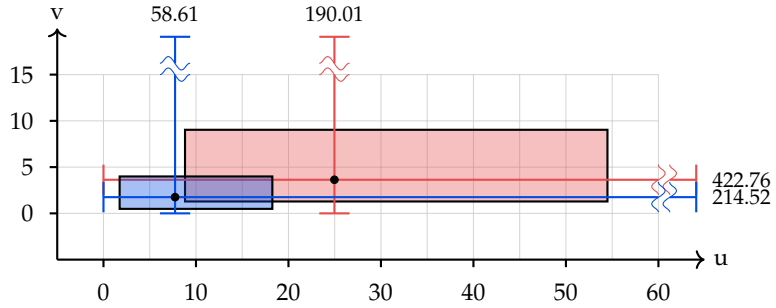

Figure 6: Deformation analysis of the object instances over 10 sequences of the nuScenes dataset [12]. 2D box plot of the deformations of the object bounding boxes in u- and v-direction in pixels with the method in [3] and [4] (red) compared to gradual image warping in (blue).

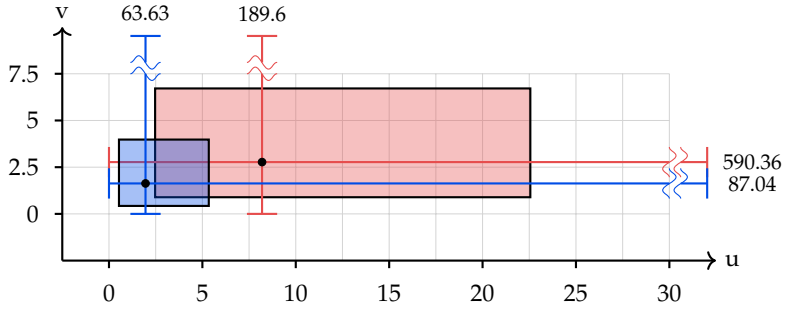

Figure 7: Deformation analysis of the object instances over 20 sequences of the Waymo Open Dataset for perception [10]. 2D box plot of the deformations of the object bounding boxes in u- and v-direction in pixels with the method in [3] and [4] (red) compared to gradual image warping in (blue).

world. Object recognition based on panoramic images has already been investigated in [4], where the network used was pre-trained on raw camera images. Consequently, it is not desirable for the objects in the panoramas to be deformed. For this reason, we run the same evaluation as in section 4.2 for the deformation of all object instances separately. In this case, an average displacement is determined for each object instance and not for each individual image. In the nuScenes dataset [12], the 2D bounding boxes are evaluated with the object classes *car*, *truck*, *bus*, *construction*, *cycle*, *trailer*, *pedestrian* and *cyclist*. In the Waymo Open Dataset for perception [10], we evaluate the panoptic labels with the classes *car*, *truck*, *bus*, *other large object*, *trailer*, *pedestrian*, *pedestrian object*, *bicycle*, *motorcycle*, *cyclist*, *motorcyclist*. The results are shown analogously as two-dimensional box plots for both evaluated datasets in Fig. 6 and 7. As in 4.2, a comparable reduction in image deformations due to gradual image warping can be recognized for the object instances.

5 Conclusion

In this article, we presented a method for improved image stitching using gradual image warping in autonomous driving. To achieve this, the images are warped in vertical sections to gradually compensate for the initial deformation caused by the estimated homography. In the evaluation, we were able to show successfully that the deformations in the panoramic images are significantly compensated for with our approach. We demonstrated this effect not only qualitatively but also quantitatively by evaluating deformations in 955 images from the nuScenes dataset [12] and the Waymo Open Dataset for perception [10]. Since our approach is primarily designed for improving object detection, we specifically measured deformations of object instances labeled in the data. Also in this case, gradual image warping shows clearly reduced image deformations. As a positive side effect, the image width of the resulting panoramas now remains constant. In upcoming research, we plan to investigate object detection capabilities on panoramic images created with gradual image warping.

References

1. T. Woopen *et al.*, "UNICARagil - Disruptive Modular Architectures for Agile, Automated Vehicle Concepts," in *27. Aachen Colloquium Automobile and Engine Technology*, 2018, pp. 663–694.
2. M. Buchholz *et al.*, "Automation of the UNICARagil vehicles," in *29th Aachen Colloquium Sustainable Mobility*, 2020, pp. 1531–1560.
3. C. Kinzig *et al.*, "Real-time seamless image stitching in autonomous driving," in *2022 25th International Conference on Information Fusion (FUSION)*, 2022, pp. 1–8.
4. ——, "Panoptic segmentation from stitched panoramic view for automated driving," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, 2024, pp. 3342–3347.
5. T.-Z. Xiang, G.-S. Xia, and L. Zhang, "Image stitching with perspective-preserving warping," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-3, pp. 287–294, 2016.
6. J. Zaragoza *et al.*, "As-projective-as-possible image stitching with moving dlt," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2339–2346.
7. C.-H. Chang, Y. Sato, and Y.-Y. Chuang, "Shape-preserving half-projective warps for image stitching," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3254–3261.
8. Y.-S. Chen and Y.-Y. Chuang, "Natural image stitching with the global similarity prior," in *Computer Vision – ECCV 2016*. Cham: Springer International Publishing, 2016, pp. 186–201.
9. Y. Zhang, Y.-K. Lai, and F.-L. Zhang, "Content-preserving image stitching with piecewise rectangular boundary constraints," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 7, pp. 3198–3212, 2021.
10. P. Sun *et al.*, "Scalability in perception for autonomous driving: Waymo open dataset," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 2443–2451.
11. H. Le *et al.*, "Deep homography estimation for dynamic scenes," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 7649–7658.
12. H. Caesar *et al.*, "nuscenes: A multimodal dataset for autonomous driving," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11 618–11 628.