2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

DSVT: Dynamic 3D Surround View for Tractor-Trailer Vehicles Based on Real-Time Pose Estimation with Drop Model

Zhipeng Dong[1,2], Mengyin Fu[1,2,3], Hao Liang[1,2], Chunhui Zhu[1,2], and Yi Yang[*,1,2]

Abstract— **In recent years, 3D surround view systems have attracted a lot of attention in the field of advanced driver assistance systems (ADAS). However, the foundational assumption of unchanging camera poses in traditional 3D surround view systems, which is designed for single-unit vehicles, results in a failure to manage the non-rigid connections characteristic of tractor-trailer vehicles. Moreover, tractor-trailer vehicles have the feature of long bodies and large wheelbases, leading to severe distortions and abrupt changes in the rendering results of previous 3D texture mapping models. In this paper, we propose DSVT, a dynamic 3D surround view system for tractor-trailer vehicles, designed to address the aforementioned issues. Specifically, we develop a dynamic surround image stitching algorithm based on relative pose estimation, which estimates the relative poses between cameras and stitches all images together to generate a 2D panoramic image. Subsequently, a novel 3D drop model is proposed, mapping the 2D panoramic image onto the 3D model for panoramic viewing. Our system can run in real time on Nvidia AGX Orin. Experimental results in real tractor-trailer scenes show that our system can achieve more accurate and natural visual effects.**

I. Introduction

Advanced driver assistance systems (ADAS) can assist drivers in perceiving their surroundings and controlling the vehicle, thereby significantly enhancing driving safety and comfort [1]. The surround view (SV) system, also known as around view monitoring (AVM), is an important part of ADAS, providing the driver with a 360-degree view of the vehicle's surroundings [2]. This system typically consists of four to six wide-angle or fisheye cameras mounted around the vehicle body, with outputs of 2D panoramic images or 3D panoramic browsing [3].

Previous research on surround view systems typically targets single-unit vehicles, where the relative poses between cameras remain constant during the vehicle's movement [3]. As a result, calibration is only required once at system initialization, allowing for image stitching [4] based on the fixed camera parameters during driving to generate 2D panoramic images.

However, the aforementioned static setup cannot meet the requirements of articulated engineering vehicles (AEV) such as tractor-trailer vehicles [5]. A typical tractor-trailer vehicle consists of a tractor and a trailer, connected non-rigidly by a fifth-wheel coupling. When the vehicle turns, there is a relative deflection between the tractor and the trailer, as shown in Fig. 1.

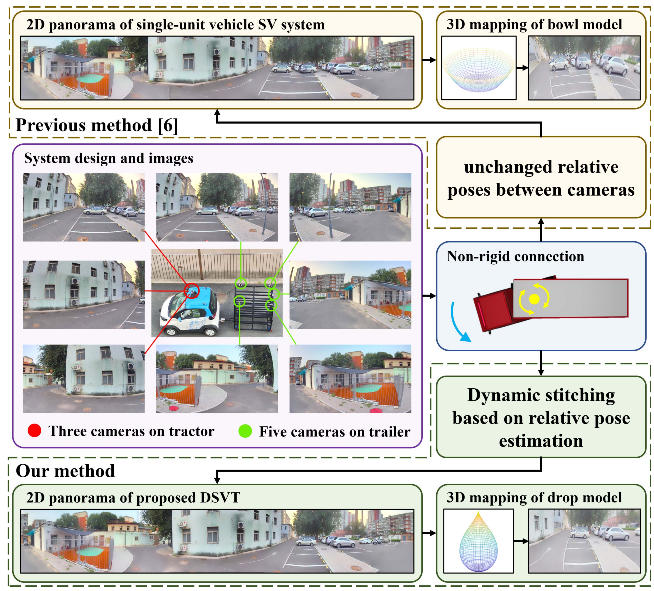


Fig. 1. **Comparison of proposed DSVT with single-unit vehicle surround view system [6].** We develop a dynamic surround image stitching algorithm based on relative pose estimation to generate 2D panoramic images. A novel 3D drop model is proposed for texture mapping.

The non-rigid connections, long bodies, and large wheelbases characteristic of tractor-trailer vehicles present numerous challenges for the design and implementation of SV systems. Firstly, most of the previous stitching methods used in single-unit vehicle SV systems, such as [3], [7], and [8], ignore the changes in relative poses between cameras, making them unsuitable for tractor-trailer scenes. Secondly, previous 3D SV systems often utilize bowl model [6] or burger model [9] for 3D texture mapping. However, these models exhibit abrupt transitions from the bottom flat surface to the side curved surface, which are even more noticeable when applied to tractor-trailer vehicles.

To address the above challenges, we introduce DSVT, which to the best of our knowledge is the first real-time 3D surround view system for tractor-trailer vehicles. The proposed system mainly consists of three parts. In the first part, based on the epipolar geometry, the relative poses between the cameras mounted on the tractor and trailer are estimated in real time. The second part undistorts the captured images

This work was partly supported by the National Natural Science Foundation of China under Grant 62233002, the National Key Research and Development Program of China under Grant 2022YFC2603602, and the Fundamental Research Funds for the Central Universities.

*Corresponding author: Yi Yang (email: yang_yi@bit.edu.cn)
[1]School of Automation, Beijing Institute of Technology, Beijing, China
[2]National Key Lab of Autonomous Intelligent Unmanned Systems, Beijing Institute of Technology, Beijing, China
[3]School of Automation, Nanjing University of Science and Technology, Nanjing, China

979-8-3503-7769-9/24/$31.00 ©2024 IEEE

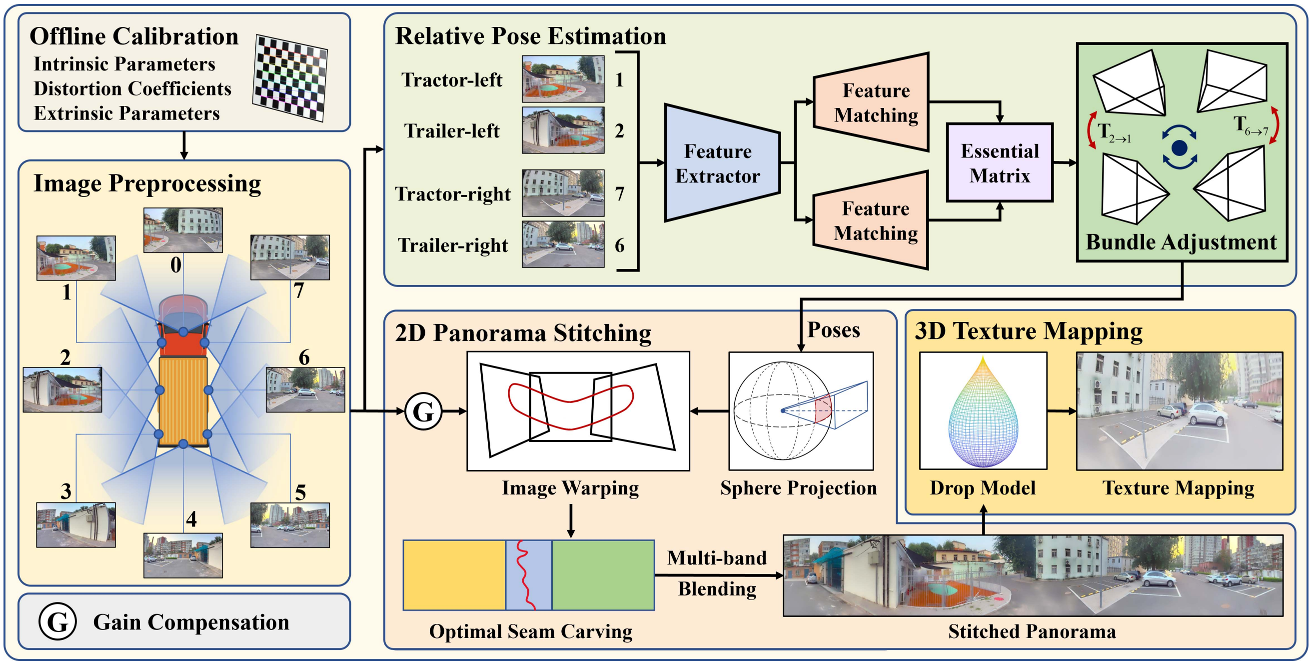


Fig. 2. **Overview of the proposed DSVT.** After undistorting, the captured images are fed into subsequent processing: The pose estimation module is utilized to determine the relative poses between adjacent cameras on the tractor and the trailer. The panorama stitching module then stitches and blends images to generate a 2D panoramic image. The resulting image is then mapped onto our proposed drop-shaped model via the texture mapping module to enable 3D panoramic viewing.

and dynamically stitches them together based on the relative poses between the cameras to create a panoramic image [10]. The third part projects the panoramic image onto our newly proposed water drop model through 3D texture mapping, enabling users to browse the surroundings of the vehicle from various viewpoints and angles. Experimental results demonstrate the effectiveness of our system.

In summary, our main contributions are as follows:

- We propose a dynamic surround image stitching algorithm based on relative pose estimation, designed to address the issue of changes in relative poses caused by the non-rigid connection of tractor-trailer vehicles.
- We propose a new drop model for 3D texture mapping, aimed at resolving the abrupt transition issues present in previous models.
- We propose DSVT, the first vision-based dynamic 3D SV system for tractor-trailer vehicles, achieving real-time 2D panoramic image stitching and 3D browsing.

II. Related Work

A. Overview of Surround View Systems

Most previous SV systems, designed for single-unit vehicles, typically feature 2D and 3D presentation forms. The 2D surround view system projects captured images onto the overhead plane and stitches them together to create a bird's-eye view [11], [12]. The 3D surround view system maps the stitched panorama onto a 3D model and allows the user to switch perspectives for panoramic browsing [6], [9]. Compared to 2D-SV systems, 3D systems provide a much better sense of immersion for the driver.

In recent years, there has been a growing interest in exploring surround view systems for tractor-trailers. Li et al. introduced the variable-angle around view monitor system (VAVM) [5], which delivers more seamless and natural stitching results in comparison to the methods proposed by Feng et al. [13] and others. However, this approach employs an angle sensor to capture the deviation of the tractor, and the resulting representation is a 2D bird's-eye view, limiting its coverage to the immediate surrounding areas.

B. 2D Panorama Stitching

Image stitching technology [4], [10] forms the foundation of surround view systems. Methods based on mesh warping, such as APAP [14], AANAP [15], and LPC [16] along with seam-driven approaches like [17], [18], represent the mainstream solutions for this type of image stitching issue. Nie et al. have leveraged deep learning technologies for image stitching, proposing methods such as UDIS [19] and UDIS2 [20].

For the panoramic stitching module of single-unit vehicle SV systems, fixed parameters are typically used for stitching due to constraints on real-time processing and computational power. However, this approach is not suitable for tractor-trailers. Zhu et al. introduced UVSS [21], a unified video stabilization and stitching framework for surround-view systems of tractor-trailer vehicles. This method pre-calibrates and stitches images from two sets of cameras mounted on the tractor and trailer, generating two semi-surround views. Subsequently, it utilizes a joint video stabilization and stitching strategy to generate a comprehensive panoramic view of the tractor-trailer's surroundings.

C. 3D Texture Mapping

The 3D surround view system will project the panorama onto the 3D model after stitching it together. The main difference between the different 3D-SV systems is their approach for texture mapping. Various 3D models have been proposed, among which the bowl model is the most widely used, and researchers have presented many formulations to describe it [6], [8], [22]–[24]. Gao et al. introduced the ship model [3], which is more consistent with the driver's visual habits compared to the cylindrical model [25]. Zhang et al. proposed the burger model [9] and utilized a two-stage texture mapping process. Beak et al. presented the thin-plate model [7], capable of adaptively adjusting the model mesh based on the surrounding environment.

III. System Framework

The flowchart of the proposed DSVT is shown in Fig. 2. The first step is to calibrate the cameras' intrinsic and extrinsic parameters offline, with the tractor facing forward. All subsequent steps are performed online. Images from 8 wide-angle cameras are captured and undistorted, before moving on to subsequent processing. The pose estimation module is used to estimate the relative poses between adjacent cameras on the tractor and the trailer. The panorama stitching module then reads the relative poses and generates a stitched image. Finally, the 2D panoramic image is projected onto our proposed 3D drop-shaped model by the texture mapping module for virtual observation.

A. Camera Calibration and Image Undistortion

Surround view systems often use wide-angle cameras to increase the overlapping area between cameras and to eliminate visual blind spots. However, the wider field of view means more image distortion. We use a checkerboard pattern and Zhang's algorithm [26] to calibrate wide-angle cameras, aiming to obtain the intrinsic parameters and distortion coefficients. In addition, the extrinsic parameters of cameras also need to be calibrated. The calibration board is placed in the co-viewing area of the adjacent cameras and the extrinsic parameters of the two cameras relative to the calibration board are determined separately. From this, the initial relative poses between neighboring cameras are calculated.

B. Feature Matching and Relative Pose Estimation

As the vehicle moves, the relative poses between the cameras must be calculated in real time due to the non-rigid connection between the tractor and trailer. Built upon the SuperGlue [27] architecture, we implemented the estimation of relative camera poses. As an example, image I_1 is taken by the camera on the tractor pointing left-forward and I_2 is taken by the camera on the left side of the trailer, the corresponding intrinsic parameters are $\mathbf{K}_1$ and $\mathbf{K}_2$. Let $\mathbf{p}_{1,i}$ and $\mathbf{p}_{2,i}$ be matched point pairs $\{(\mathbf{p}_{1,i}, \mathbf{p}_{2,i})\}_{i=1,2,...,N}$ in homogeneous coordinates, where N is the number of matched point pairs.

Our aim is to estimate the relative pose $\mathbf{T}_{2\to1} = [\mathbf{R}|\mathbf{t}] \in \mathcal{SE}(3)$ between the front-left frame I_1 and the left frame I_2. Given the matched point pairs, they should satisfy:

$$\mathbf{p}_{1,i}^T \mathbf{K}_1^{-T} [\mathbf{t}]_\times \mathbf{R} \mathbf{K}_2^{-1} \mathbf{p}_{2,i} = 0 \tag{1}$$

where $[\mathbf{t}]_\times$ is the skew-symmetric matrix representation of $\mathbf{t}$, and $[\mathbf{t}]_\times \mathbf{R}$ is defined as the essential matrix $\mathbf{E}$.

Using RANSAC and five-point-algorithm [28], we can solve $\mathbf{E}$ from (1). Then the relative pose is recovered by SVD. Similarly, the relative pose $\mathbf{T}_{6\to7} = [\mathbf{R}|\mathbf{t}] \in \mathcal{SE}(3)$ between the front-right frame I_7 and the right frame I_6 can also be solved using this method.

Theoretically, with the initial poses of all cameras meticulously pre-calibrated, determining the relative pose between the front-left camera C_1 and the left-side camera C_2 allows for the inference of the relative pose between the front-right camera C_7 and the right-side camera C_6. However, due to significant camera distortions that cannot be entirely rectified, there exists some error in the pose estimation results. Consequently, it's necessary to estimate the poses on both sides and then refine them through bundle adjustment (BA) optimization:

$$\min_{\mathbf{R},\mathbf{t}} E = \sum_{k=1,6} \sum_{i} \left\| \mathbf{p}_{k,i} - \mathbf{K}_k \mathbf{T}_{k+1\to k} \mathbf{K}_{k+1}^{-1} \mathbf{p}_{k+1,i} \right\|_2^2 \tag{2}$$

C. Panorama Stitching and Blending

Giving the cameras' intrinsic parameters and relative poses, it is possible to project all images into a common sphere coordinate system by means of a projection transformation. As exposure variations may exist among images taken by distinct cameras, it is required to perform exposure compensation [10] on all images prior to the stitching procedure. Due to the presence of factors such as parallax, it is therefore necessary to find an optimal stitching seam [18] in the overlapping area in order to avoid blurring and artefacts in the stitched image. In our implementation, the dynamic programming method is used to find the optimal stitching seam [29] in order to reduce the computational burden.

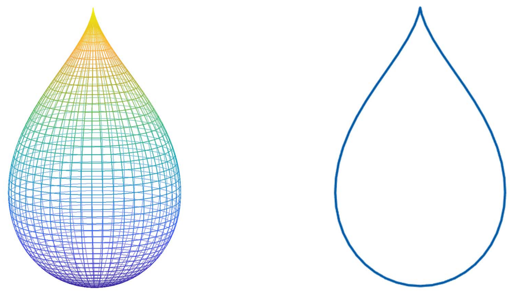
Let $e_{i,j}$ be the pixel error at point $\mathbf{p}(i,j)$ in the overlapping region of images I_f and I_l. Since the width of the overlapping region is smaller than the height, the optimal seam should be vertically oriented. Traverse the point $\mathbf{p}$ horizontally and calculate the cumulative minimum error $E_{i,j}$ for all possible paths to the current point $\mathbf{p}(i,j)$:

$$E_{i,j} = e_{i,j} + min(E_{i-1,j-1}, E_{i-1,j}, E_{i-1,j+1}) \tag{3}$$

Where $E_{i-1,j-1}, E_{i-1,j}, E_{i-1,j+1}$ indicates the minimum cumulative error E of the top left, top, and top right points of the current point, respectively. Dynamic programming algorithm is used to solve (3) to get the optimal seam.

In order to avoid incoherent transitions at the stitching seam, the images on both sides of the seam also need to be blended. We use the multi-band blending algorithm [10] based on Laplacian pyramid to preserve and blend information from each frequency band. The Laplacian pyramid $\mathbf{L}$ can be obtained by extending the Gaussian pyramid $\mathbf{G}$ with the i-th level L_i defined as:

$$\mathbf{L}_i = \mathbf{G}_i - \mathrm{UP}(\mathbf{G}_{i+1}) * \mathcal{G}_{5\times5} \tag{4}$$



(a) The 3D drop model. (b) The sectional view of the model.

Fig. 3. **The proposed drop model and its sectional view.**

where $\mathbf{G}_{i+1}$ is the image of the upper layer of Gaussian pyramid and $\mathbf{G}_0$ is the original image. UP$(\cdot)$ denotes the upsampling operation and $\mathcal{G}_{5\times5}$ is a 5×5 Gaussian convolution kernel. Then the two Laplacian pyramid are fused by feathering blending, and the blended image is reconstructed from the fused pyramids.

D. 3D Model and Texture Mapping

The 3D model is the carrier for the panoramic view of the surroundings. After mapping the stitched panorama onto the 3D model, the user can switch viewpoints and perspectives for virtual panoramic browsing. This paper introduces a novel 3D model of a "water drop" shape, as illustrated in Fig. 3. This model can be generated by revolving a drop-shaped curve around the axis of symmetry. The curve defining the water drop shape can be described by various mathematical expressions, such as the piriform of Longchamps or the Joukowski airfoil. In this paper, the drop-shaped curve is described using the quartic curve as outlined in (5), with its parametric form presented as shown in (6).

$$(x^2+y^2)^2 - 2x(x^2+y^2) + 3y^2 = 0 \tag{5}$$

$$\begin{cases} x = \dfrac{3(1+\cos t)\sin t}{5+4\cos t} \\ y = -\dfrac{3(1+\cos t)(2+\cos t)}{5+4\cos t} \end{cases}, \ t \in [0, 2\pi] \tag{6}$$

Utilizing the parametric equations (6) allows for the rendering of a drop-shaped curve, which can then be revolved to produce a 3D model. Establishing a mapping relationship between the 2D coordinates of panoramic image pixels and the 3D model's mesh vertex coordinates facilitates the texture mapping of the panoramic image. By adjusting the values of multiple coefficients in (6), drop models of various shapes can be obtained.

IV. Experiments

A. System Design and Details

The system configuration is as Fig. 4 shown. The embedded platform on which the proposed DSVT runs is the Nvidia Jetson AGX Orin. The system features an 8-channel camera setup, each with a 120° wide-angle lens, and downscales the image to 1280×720 for real-time processing demands. We utilize OpenCV with CUDA for image processing, employ TensorRT to accelerate feature extraction and matching, and use OpenGL for 3D rendering.

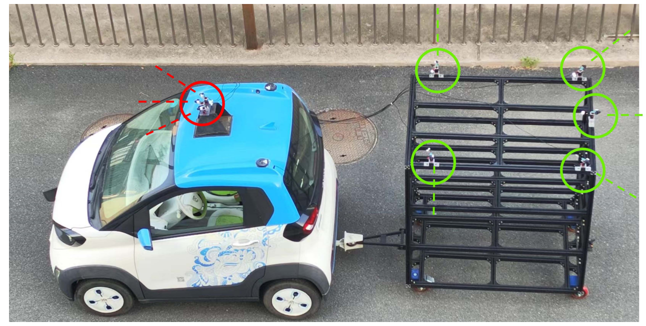


Fig. 4. **A realistic view of the experimental truck equipped with our dynamic surround view system.** There are eight cameras in total, three on the tractor (indicated in red) and five on the trailer (indicated in green).

TABLE I
RUNTIME TESTING

Update mode	Update interval	Cost time	Frame rate
Frame by frame	1	166ms	6.02
Every few frames	2	98ms	10.20
	5	61ms	16.39
	10	46ms	21.74
	30	39ms	25.64
Only first frame	$+\infty$	35ms	28.57

Our dynamic 3D-SV system DSVT has been implemented on an experimental truck, which measures 4.5m in total length, 1.5m in width, and 1.8m in height. The truck is outfitted with a set of 8 cameras, with the arrangement and orientation of the cameras as shown in Fig. 4.

We conducted experiments in multiple different scenes. The 2D panoramic images stitched are shown in Fig. 5 and 8 (b), with a resolution of about 3336×556 pixels. The panoramic images are then mapped onto our proposed drop model, with the results displayed in Fig. 9 (c).

The system's runtime performance is detailed in Table I. Both experimental results and prior work [9] indicate that the entire system does not require frame-by-frame operation, so we update the relative poses and seam lines every few frames. When the system parameters are updated every 10 frames, the average frame rate reaches 20+ fps. With an update interval set to $+\infty$, the system effectively functioning as a static surround view system.

B. Comparison of 2D Panoramic Image Stitching

In this subsection, we conduct comparative experiments in the results of 2D panoramic image stitching in tractor-trailer scenes. Firstly, we compare our stitching results with those of single-unit vehicle SV systems. Secondly, we compare our method with state-of-the-art (SOTA) universal image stitching methods. Finally, compare our method with methods specifically designed for tractor-trailer surround view tasks.

(a) Turn right in the parking lot.

(b) Turn left in the alley.

Fig. 5. **The stitching results of the single-unit vehicle surround view system [6] and our proposed DSVT during vehicle turning.** Red boxes highlight the duplicated and missing views at the junction between the tractor and trailer by the single-unit vehicle SV system, while green boxes highlight our stitching results achieved by our method.

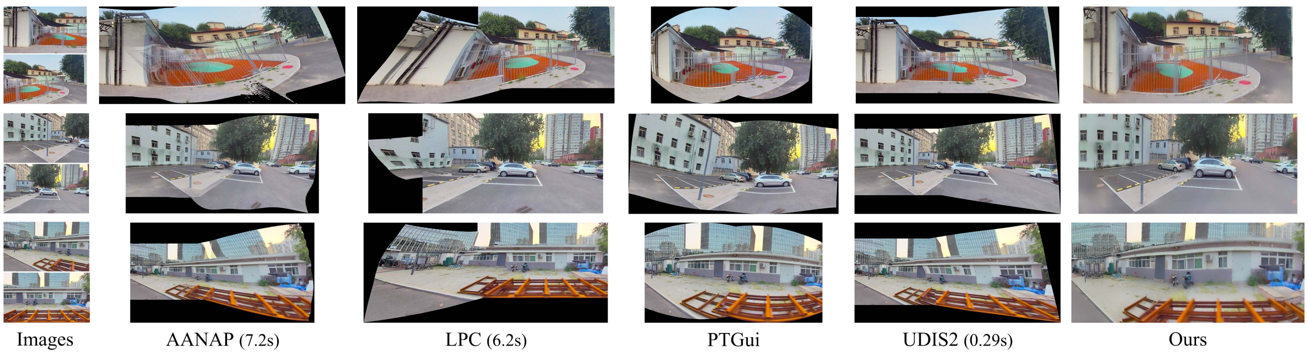


Fig. 6. **The stitching results of universal stitching methods and our proposed DSVT in tractor-trailer scenes.** Including the classic method AANAP [15], the SOTA method LPC [16] and UDIS2 [20], as well as the professional software PTGui.

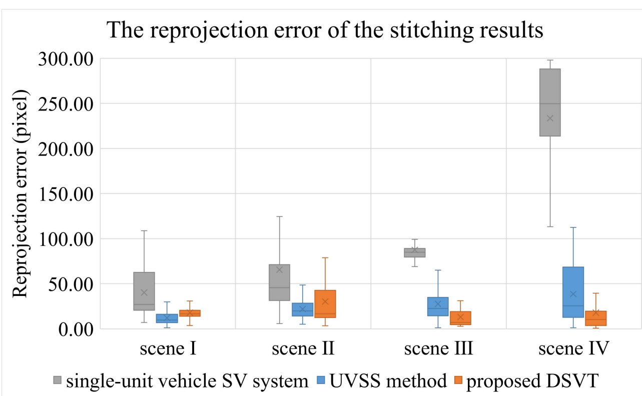


Fig. 7. **Quantitative comparison of single-unit vehicle SV system [6], the UVSS method [21], and our proposed DSVT on the four scenes.**

1) Single-unit vehicle SV system in tractor-trailer scenes: Single-unit vehicle SV systems assume that the relative poses between cameras remain constant. However, tractor-trailer vehicles do not meet this assumption. Fig. 5 shows how the stitching results of our proposed DSVT differ from the single-unit vehicle SV system [6] when the tractor turns. As can be seen, our DSVT can dynamically update the relative poses between cameras to generate a better panorama. In contrast, the single-unit vehicle system cannot deal with the deflection of the tractor, and the stitched images have duplicated and missing views.

To quantitatively compare the performance of single-unit vehicle SV systems and our DSVT in panoramic image stitching, the reprojection error [14] is calculated based on the feature matching and pose estimation results. Take cameras C_1 and C_2 on the left side of the tractor-trailer as an example, the reprojection error [14] E_{12} is defined as:

$$E_{12} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|f(\mathbf{p}_{2,i}) - \mathbf{p}_{1,i}\right\|^2} \tag{7}$$

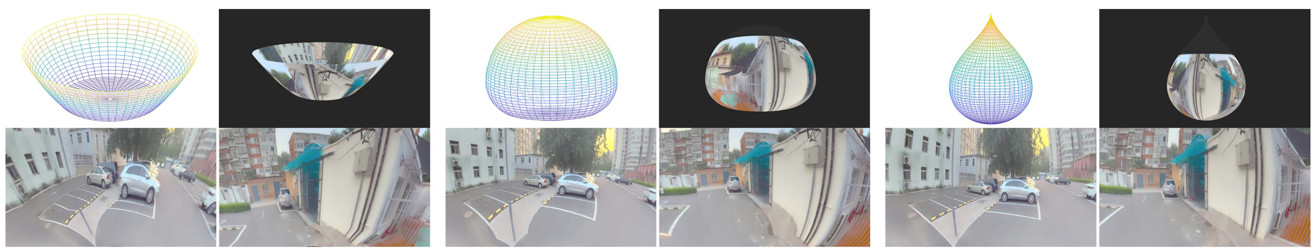
where $f(\cdot)$ denotes the projection of point $\mathbf{p}_{2,i}$ onto image I_1 based on the pose estimation results. The error E_{67} corresponding to cameras C_6 and C_7 on the right side of the tractor-trailer is analogous.

The quantitative comparison results in the four scenes are shown in Fig. 7. It can be observed that our method significantly outperforms the single-unit vehicle SV system across all four scenes, especially in scene 4, where the tractor-trailer is continuously turning.

2) Universal image stitching methods in tractor-trailer scenes: We compare our method with universal image stitching methods, including the SOTA algorithms UDIS2 [20] and LPC [16], the classic algorithm AANAP [15], as well as the professional stitching software PTGui (https://ptgui.com). We utilized these four algorithms to stitch images from two non-rigidly connected cameras located at the junction of the vehicle's tractor and trailer. The experimental platform was equipped with an Intel i7-13700F and RTX 3090Ti, with the results presented in Fig. 6. Our method's results were extracted from the complete surround-view panoramas.

(a) UVSS (b) Proposed DSVT

Fig. 8. **Panoramic images stitched from four different scenes using the UVSS method [21] and our proposed DSVT.** Red boxes highlight structural distortions during mesh warping by UVSS. Green boxes represent the corresponding results of our method. Blue boxes denote the common visual areas at the image's left and right ends that UVSS cannot process.



(a) Bowl model (b) Burger model (c) Proposed drop model

Fig. 9. **Rendering results of the same panorama by bowl model [6], burger model [9] and our drop model.**

From the experimental results shown in Fig. 6, it is evident that AANAP, LPC, and UDIS2 all adopt one image as the reference and then warp and stitch the other image accordingly. Therefore, these algorithms exhibit significant stretching at the edges of the stitched images, making such approaches ill-suited for expansion to 8-channel surround-view image stitching problems. Theoretically, PTGui is capable of jointly stitching 8 images, but the stitching outcome heavily relies on manually annotated and adjusted control points. In contrast, our algorithm's stitching quality in the overlapping areas is comparable to UDIS2 and PTGui, and it facilitates smooth transitions between multiple images. Lastly, the time consumption for these four algorithms to stitch two images on a PC reaches up to the level of seconds, whereas our algorithm can achieve real-time stitching of 8 images on an embedded platform.

3) Tractor-trailer surround image stitching methods: We further compare our method with UVSS [21], which is tailored for surround stitching tasks involving tractor-trailer vehicles. Following the UVSS methodology, we fristly generate two semi-surround views, and the complete panorama is then created through feature matching and mesh warping techniques. This experiment focused solely on the comparison of stitching outcomes, thus omitting the stabilization module and temporal constraints in UVSS. The stitching results are shown in Fig. 8, and the quantitative comparison results of reprojection errors [14] are displayed in Fig. 7.

The results reveal that our method achieves a stitching result at the junction between the tractor and trailer comparable to that of UVSS. However, UVSS essentially stitches two images together, failing to create a seamless loop, resulting in the left and right ends of the stitched image not forming a complete circle. In contrast, our method facilitates a 360° surround view. Additionally, our approach can run in real-time on an embedded platform, whereas UVSS requires up to 6.24 seconds for a single stitching operation.

C. Comparison of 3D Models

We projected the same panoramic image onto the bowl model [6], the burger model [9], and our drop model, with the rendering results shown in Fig. 9. Considering the elongated characteristics of the tractor-trailer's body, we fine-tuned the coefficients of the parametric equation based on the standard drop curve described in (6). The fine-tuned model features a slightly flatter bottom curve and a moderately increased width, thereby achieving a more natural visual effect.

From the comparative images, it is evident that both the bowl model [6] and the burger model [9] exhibit abrupt transitions from the bottom horizontal plane to the side curved surface. When linear structures in the scene pass through these transition areas, very noticeable distortions and unnatural transitions occur. The long body and large wheelbase characteristics of tractor-trailer vehicles result in a significantly elongated 2D panoramic image from the stitching module, further exacerbating the abrupt transitions. In contrast, our proposed drop model avoids such abrupt changes, resulting in a smoother and more natural projection of the panoramic image onto the model.

V. CONCLUSIONS

This paper presents DSVT, a dynamic 3D surround view system for tractor-trailer vehicles. The system is primarily composed of three modules: pose estimation, 2D panorama stitching and 3D texture mapping. To address the non-rigid connection between the tractor and the trailer, the pose estimation module is utilized to estimate the relative poses between adjacent cameras on the tractor and the trailer. The 2D panorama stitching module generates panoramic images based on the relative poses, and the 3D texture mapping module map them onto our proposed drop-shaped model for virtual viewing. Experimental results in various scenes have confirmed that our system is capable of achieving dynamic 3D surround view perception for tractor-trailer vehicles and can run in real-time on embedded platforms.

To the best of our knowledge, this is the first research focused on camera-only real-time dynamic 3D surround view perception specifically for tractor-trailer vehicles. In future work, we plan to explore dynamic 3D models that are more suitable for the surround view of tractor-trailer vehicles.

ACKNOWLEDGMENT

The authors would like to thank Yu Gao, Jiadong Tang, Zhaoxiang Liang, Hao Han, Yaowen Zhang and all other members of ININ Lab of Beijing Institute of Technology for their contribution to this work.

REFERENCES

[1] J. Nidamanuri, C. Nibhanupudi, R. Assfalg, and H. Venkataraman, "A progressive review: Emerging technologies for adas driven solutions," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 2, pp. 326–341, 2022.

[2] V. R. Kumar, C. Eising, C. Witt, and S. Yogamani, "Surround-view fisheye camera perception for automated driving: Overview, survey & challenges," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–22, 2023.

[3] Y. Gao, C. Lin, Y. Zhao, X. Wang, S. Wei, and Q. Huang, "3-d surround view for advanced driver assistance systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 320–328, 2018.

[4] M. Fu, H. Liang, C. Zhu, Z. Dong, R. Sun, Y. Yue, and Y. Yang, "Image stitching techniques applied to plane or 3-d models: A review," *IEEE Sensors Journal*, vol. 23, no. 8, pp. 8060–8079, 2023.

[5] W. Li, L. Cao, P. Yue, S. Wen, J. Liao, J. Xia, and X. Feng, "Vavm: A flexible technique for variable-angle around view monitor system towards articulated engineering vehicle," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 4, pp. 3556–3568, 2022.

[6] S. Gasparyan and G. Vasilianov, "Real-time 3d surround view system for vehicle based on panoramic stitching image," in *Networked Control Systems for Connected and Automated Vehicles*, A. Guda, Ed. Cham: Springer International Publishing, 2023, pp. 85–93.

[7] I. Baek, A. Kanda, T. C. Tai, A. Saxena, and R. Rajkumar, "Thin-plate spline-based adaptive 3d surround view," in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 586–593.

[8] O. Eker, B. Ercan, B. Bayraktar, and M. Bal, "A real-time 3d surround view pipeline for embedded devices," in *VISIGRAPP (4: VISAPP)*, 2022, pp. 257–263.

[9] L. Zhang, J. Chen, D. Liu, Y. Shen, and S. Zhao, "Seamless 3d surround view with a novel burger model," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4150–4154.

[10] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International journal of computer vision*, vol. 74, pp. 59–73, 2007.

[11] S.-H. Cheon, J.-W. Jang, and S.-W. Jang, "Design and implementation of an around view monitoring system of selecting scope of view," *International Journal of Control and Automation*, vol. 7, no. 11, pp. 213–222, 2014.

[12] M. Yu and G. Ma, "360 surround view system with parking guidance," *SAE International Journal of Commercial Vehicles*, vol. 7, no. 2014-01-0157, pp. 19–24, 2014.

[13] X. Feng, W. Li, T. Wei, Y. Zhang, and L. Cao, "Calibration and stitching methods of around view monitor system of articulated multi-carriage road vehicle for intelligent transportation," in *WCX SAE World Congress Experience*, 2019.

[14] J. Zaragoza, T.-J. Chin, M. S. Brown, and D. Suter, "As-projective-as-possible image stitching with moving dlt," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

[15] C.-C. Lin, S. U. Pankanti, K. N. Ramamurthy, and A. Y. Aravkin, "Adaptive as-natural-as-possible image stitching," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1155–1163.

[16] Q. Jia, Z. Li, X. Fan, H. Zhao, S. Teng, X. Ye, and L. J. Latecki, "Leveraging line-point consistence to preserve structures for wide parallax image stitching," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 12 181–12 190.

[17] K. Lin, N. Jiang, L.-F. Cheong, M. Do, and J. Lu, "Seagull: Seam-guided local alignment for parallax-tolerant image stitching," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 370–385.

[18] J. Gao, Y. Li, T.-J. Chin, and M. S. Brown, "Seam-driven image stitching." in *Eurographics (Short Papers)*, 2013, pp. 45–48.

[19] L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Unsupervised deep image stitching: Reconstructing stitched features to images," *IEEE Transactions on Image Processing*, vol. 30, pp. 6184–6197, 2021.

[20] L. Nie, C. Lin, K. Liao, S. Liu, and Y. Zhao, "Parallax-tolerant unsupervised deep image stitching," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, pp. 7399–7408.

[21] C. Zhu, Y. Yang, H. Liang, Z. Dong, and M. Fu, "Uvss: Unified video stabilization and stitching for surround view of tractor-trailer vehicles," in *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023, pp. 9014–9020.

[22] B. Zhang, V. Appia, I. Pekkucuksen, Y. Liu, A. Umit Batur, P. Shastry, S. Liu, S. Sivasankaran, and K. Chitnis, "A surround view camera solution for embedded systems," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2014.

[23] Z. Liang, L. Chen, and F. An, "3d vehicle surround view algorithm for embedded platform," *Journal of Physics: Conference Series*, vol. 2253, no. 1, p. 012026, apr 2022.

[24] J.-H. Kim, S.-K. Kim, T.-M. Lee, Y.-J. Lim, and J. Lim, "Hemispherical 3d around view monitoring algorithm using image synthesis of multi-channel cameras," in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2018, pp. 1466–1471.

[25] M. Lin, G. Xu, X. Ren, and K. Xu, "Cylindrical panoramic image stitching method based on multi-cameras," in *2015 IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2015, pp. 1091–1096.

[26] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, Nov 2000.

[27] P.-E. Sarlin, D. DeTone, T. Malisiewicz, and A. Rabinovich, "Superglue: Learning feature matching with graph neural networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[28] D. Nister, "An efficient solution to the five-point relative pose problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 6, pp. 756–770, 2004.

[29] J. Zhang, Y. Gao, Y. Xu, Y. Huang, Y. Yu, and X. Shu, "A simple yet effective image stitching with computational suture zone," *The Visual Computer*, 2022.