Research Article Vol. 31, No. 23 / 6 Nov 2023 / Optics Express 39007
Optics EXPRESS

Kinematic target surface sensing based on improved deep optical flow tracking

LEI LU,[1] HAO LIU,[1] HONGLIANG FU,[1] ZHILONG SU,[2,3,*] WEI PAN,[4] QINGHUI ZHANG,[1] AND JINHUI WANG[1]

[1] College of Information Science and Engineering, Henan University of Technology, Zhengzhou, 450001, China
[2] Shanghai Institute of Applied Mathematics and Mechanics, School of Mechanics and Engineering Science, Shanghai Key Laboratory of Mechanics in Energy Engineering, Shanghai University, Shanghai 200444, China
[3] Shaoxing Research Institute of Shanghai University, Shaoxing 312074, China
[4] Department of Research and Development, OPT Machine Vision Tech Co., Ltd., Dongguan 523860, China
[*] zhilong8845@shu.edu.cn

Abstract: Reconstruction of moving target surfaces based on active image sensing techniques, such as phase-shifting profilometry, has attracted intensive research in recent years. The measurement error caused by object motion can be addressed successfully by tracking the object movement. However, it either requires high-cost color imaging equipment or is limited by the assumption of 2D translation movement. Therefore, this paper proposes what we believe to be a new method to reconstruct the kinematic object surfaces with any 2D movement sensed by affordable monochrome camera. An improved RAFT optical flow algorithm is proposed to track the object based on the object fringe pattern image directly. The feature points on the object are retrieved immune to the fringe pattern illumination. Then, the RANSAC algorithm and an iteration selection process are employed to select feature points with high quality optical flow. At last, the motion is described mathematically, and the dynamic object is reconstructed successfully. Experiments are presented to verify the effectiveness of the proposed method.

1. Introduction

Three-dimensional (3D) reconstruction, a subfield within computer vision, has gained exceptional popularity as a measurement tool in recent decades, owing to its inherent advantages in capturing real-world objects' visual appearance and geometric shape [1]. It has a wide range of applications in the protection of cultural relics, observation of crustal movements, and forest inventory [2–4].

Phase-shifting profilometry (PSP) is one of the most popular 3D measurement technologies and it has been widely used in industrial inspection, dentistry, robot navigation and other fields [5–9]. The typical PSP reconstruction system includes one camera and one projector. The projector projects at least three phase-shifted sinusoidal fringe patterns onto the object surface, then the camera captures the reflected fringe patterns from another angle. Because of the object height, the captured fringe patterns will have distortions. The phase information contained in the sinusoidal fringe pattern is retrieved and the object is reconstructed based on the phase information and system calibration parameters. As multiple fringe patterns are employed, the noise such as ambient light and object surface reflectivity are removed during the reconstruction, leading to the advantages of high accuracy and robustness. However, PSP requires the object to be kept static during the capture of the multiple fringe patterns, errors will be introduced when the dynamic object is reconstructed [10–11].

Recently, dynamic object reconstruction based on PSP has attracted intensive attention. Zhang et al. proposed a modified two-plus-one phase shift algorithm to alleviate the errors caused by object motion [12]. Two sinusoidal fringe patterns and a uniform plane image are used to

#505993 https://doi.org/10.1364/OE.505993
Journal © 2023 Received 14 Sep 2023; revised 17 Oct 2023; accepted 18 Oct 2023; published 1 Nov 2023

Optics EXPRESS

calculate the phase map. The error introduced by motion is reduced as less sinusoidal fringe patterns are used. Liu et al. estimated the motion of the object based on the difference between two subsequent 3D frames [13]. The phase shift error is determined by leveraging the pinhole model of the projector. The motion is assumed constant among different 3D frames. Wei et al. proposed to address the motion induced error by projecting one static fringe pattern [14]. Instead of projecting multiple phase-shifted fringe patterns, the phase shift generated by the motion is employed to retrieve the phase value. Since only one fringe pattern is used to obtain the phase map, the accuracy is affected by the variations of ambient light and object surface reflection. Wu et al. proposed to reconstruct the isolated moving object by projecting two fringe patterns with different frequencies [15]. The phase shift required by the PSP is generated by the object motion and the model describing the motion-induced phase shift is presented by analyzing the influence introduced by movement. Then, the phase information in different frequencies is retrieved. Finally, the mismatch on the phase information between the two frequencies is compensated and the isolated moving object is reconstructed. Wang et al. proposed to reduce the motion induced error by additional temporal sampling [16]. Each illuminated fringe pattern will be captured twice in one projection cycle, resulting in two sets of phases shifted fringe images being obtained. The motion induced phase error is effectively separated from the fixed phase shift value by evaluating the difference between the two-phase maps. An iterative compensation strategy is further applied to compensate for the phase error until high-quality phase maps are generated. This method requires the object to move at a constant velocity.

The most straight forward approach to reconstruct the dynamic object is tracking the position of the object within the fringe patterns. Lu et al. tracked the object motion by placing markers on the object surface in advance [17]. Then, the object motion is described mathematically by the rotation matrix and translation vector. The influence on the phase value caused by the motion is analyzed and a new reconstruction model describing the fringe patterns with motion information is presented. At last, the correct phase information is retrieved, and the dynamic object is reconstructed successfully. The same research group automated the method by tracking the object based on SIFT (scale-invariant feature transform) algorithm instead of markers [18–21]. The SIFT algorithm cannot directly extract the motion information of the object because of the projected fringe patterns. The method used different color channels in a color camera to separate the pure object feature information and fringe pattern information. The fringe pattern is projected in one of the color channels and this channel is filtered out before applying SIFT. However, the color equipment is required and leading to high cost and low frame rate than the monochrome ones. Duan et al. proposed a motion estimated method to automatically track the motion based on object fringe pattern directly [22]. The fringe-oriented feature extraction technique is proposed, and the object feature is obtained by eliminating the modulation influence of sinusoidal illumination. However, the motion is limited to two-dimensional (2D) translation movement and the method cannot track the object with rotation movement.

This paper presents an automated method to reconstruct the dynamic object with any 2D movement. An improved RAFT (recurrent all-pairs field transforms) optical flow algorithm immune to the fringe pattern illumination is proposed to retrieve the object features from the captured fringe pattern images directly. Then, RANSAC (RANdom SAmpling Consistency) algorithm and an iteration selection process are used to select the correct feature points with high accuracy. Finally, the rotation matrix and translation vector describing the object motion are obtained and the dynamic object is reconstructed to compensate for the errors caused by motion.

Optics EXPRESS

2. Principle

2.1. Principle of tracking-based PSP

For N - *step*PSP, the captured fringe pattern on the reference plane and object surface can be expressed as follows,

$$s_n^p(x,y) = A + B\cos(\phi^p(x,y) + \lambda_n) \tag{1}$$

$$d_n^w(x,y) = A + B\cos(\phi^p(x,y) + \Phi(x,y) + \lambda_n) \tag{2}$$

where $s_n^p(x,y)$ is the n - *th* fringe pattern on the reference plane, $d_n^w(x,y)$ is the n - *th* fringe pattern on the object. $\phi^p(x,y)$ is the phase value on the reference plane and $\Phi(x,y)$ is the phase variation caused by the height of the object; A is the intensity value of the ambient light and B is the amplitude modulation of the intensity of the fringe pattern; $\lambda_n = \frac{2\pi(n-1)}{N}$ is the phase shift and $n = 1, 2, 3, \ldots, N$. The phase maps of the reference plane $\phi^p(x,y)$ and the object $\phi^w(x,y)$ can be calculated by:

$$\phi^p(x,y) = \tan^{-1}\frac{-\sum_{n=1}^{N} s_n^p(x,y)\sin\lambda_n}{\sum_{n=1}^{N} s_n^p(x,y)\cos\lambda_n} \tag{3}$$

$$\phi^w(x,y) = \Phi(x,y) + \phi^p(x,y) = \tan^{-1}\frac{-\sum_{n=1}^{N} d_n^w(x,y)\sin\lambda_n}{\sum_{n=1}^{N} d_n^w(x,y)\cos\lambda_n} \tag{4}$$

The phase value is wrapped into $[0, 2\pi]$ because of the $\tan^{-1}(\cdot)$ function, leading to phase jump and ambiguity. Phase unwrapping is applied to remove the phase discontinuous and monotonous phase value is obtained. At last, the object is reconstructed based on the unwrapped phase and system calibration parameters. From the above derivation the PSP requires the object to keep stationary among different fringe patterns. When the dynamic object is reconstructed, position mismatch will be introduced and the phase shift among the fringe patterns will be violated, leading to error of the phase retrieval.

With the relationship between the object motion and phase information, the fringe pattern for the dynamic object can be expressed as [13]:

$$\widetilde{d}_n^w(x,y) = A + B\cos(\phi^p(\overline{f}(x,y), \overline{g}(x,y)) + \Phi(x,y) + \lambda_n) \tag{5}$$

where $\overline{f}(x,y)$ and $\overline{g}(x,y)$ are the motion information described by the rotation matrix and translation vector. Extending Eq. (5) to n - *th* PSP, the fringe patterns with the motion information can be obtained as follows:

$$\begin{cases} \widetilde{d}_1^w(x,y) = A + B\cos(\phi^p(x,y) + \Phi(x,y)) \\ \widetilde{d}_2^w(x,y) = A + B\cos(\phi^p(\overline{f}_2(x,y), \overline{g}_2(x,y)) + \Phi(x,y) + \lambda_2) \\ \vdots \\ \widetilde{d}_N^w(x,y) = A + B\cos(\phi^p(\overline{f}_N(x,y), \overline{g}_N(x,y)) + \Phi(x,y) + \lambda_n) \end{cases} \tag{6}$$

when the motion information $(\overline{f}(x,y), \overline{g}(x,y))$ is known, the correct phase $\Phi(x,y)$ can be obtained by [13]:

$$\Phi(x,y) = \arctan\frac{D_a - D_b}{D_c - D_d} \tag{7}$$

Optics EXPRESS

where

$$D_a = \sum_{n=1}^{N} \widetilde{d}_n^w(x,y) \cos\lambda_n \sum_{n=1}^{N} \cos(\phi^p(\bar{f}_n(x,y), \bar{g}_n(x,y) + \lambda_n) \sin\lambda_n \tag{8}$$

$$D_b = \sum_{n=1}^{N} \widetilde{d}_n^w(x,y) \sin\lambda_n \sum_{n=1}^{N} \cos(\phi^p(\bar{f}_n(x,y), \bar{g}_n(x,y) + \lambda_n) \cos\lambda_n \tag{9}$$

$$D_c = \sum_{n=1}^{N} \widetilde{d}_n^w(x,y) \cos\lambda_n \sum_{n=1}^{N} \sin(\phi^p(\bar{f}_n(x,y), \bar{g}_n(x,y) + \lambda_n) \sin\lambda_n \tag{10}$$

$$D_d = \sum_{n=1}^{N} \widetilde{d}_n^w(x,y) \sin\lambda_n \sum_{n=1}^{N} \sin(\phi^p(\bar{f}_n(x,y), \bar{g}_n(x,y) + \lambda_n) \cos\lambda_n \tag{11}$$

2.2. Object motion tracking by improved RAFT algorithm

The tracking-based PSP eliminates the motion induced error by object motion tracking. SIFT algorithm is employed in the traditional algorithms and it can achieve high accuracy when the object image without fringe pattern is applied. The periodic sinusoidal fringe pattern on the object surface is "noisy" to the SIFT algorithm, leading to failure of the feature extraction as shown in Fig. 1. This paper proposes to track the object motion by the captured fringe pattern image directly based on an improved RAFT optical flow algorithm. The normalized cross correlation (NCC) module is introduced to the network to immune the influence of the fringe pattern illumination for obtaining reliable optical flow of each point on the kinematic object. Then, correct feature points with high accuracy are selected based on RANSAC algorithm and an iteration selection process. Finally, the rotation matrix and translation vector are calculated to describe the motion mathematically.

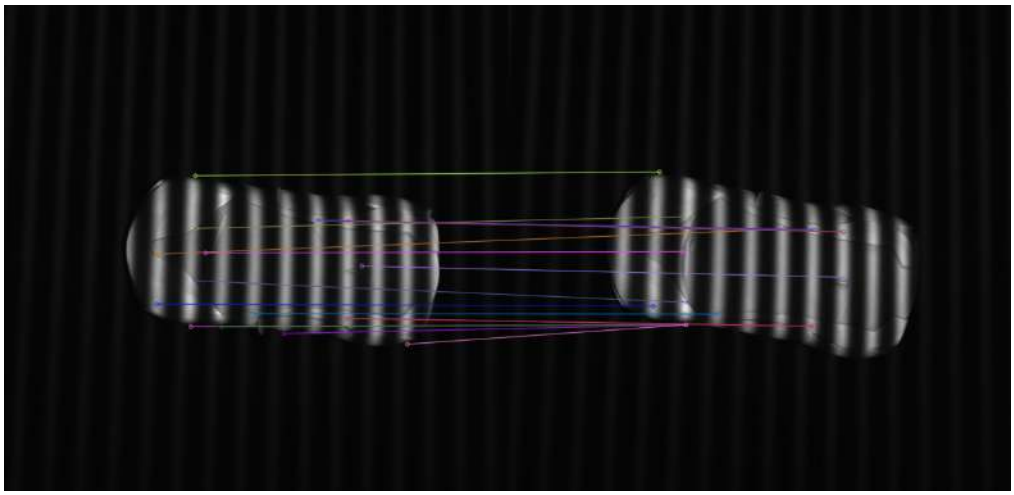

Fig. 1. Mismatches of the SIFT algorithm for fringe pattern images of moving object

2.2.1. Improved RAFT optical flow network

Figure 2 shows the architecture of the proposed improved RAFT optical flow network. Given a pair of consecutive images Frame1 and Frame2 (i.e. $\widetilde{d}_1^w(x,y)$ and $\widetilde{d}_2^w(x,y)$), a dense displacement

Optics EXPRESS

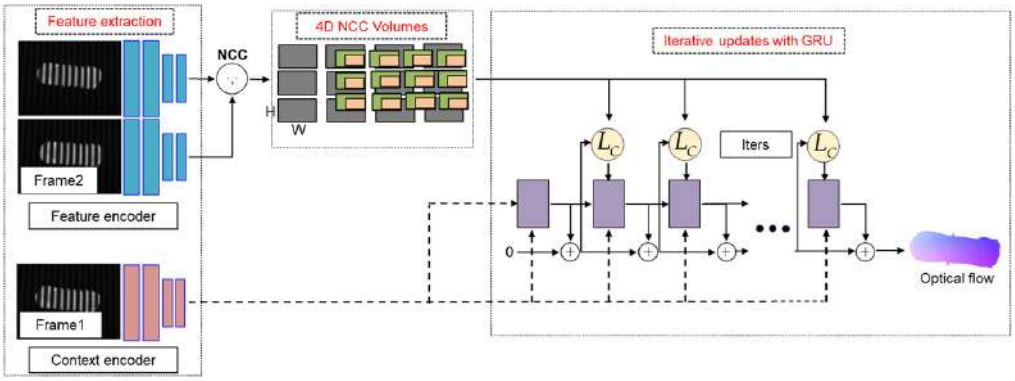

Fig. 2. Architecture of the improved RAFT optical flow network

field $(\widetilde{f}^1, \widetilde{f}^2)$ is estimated and mapping each pixel (u, v) in Frame1 to the corresponding coordinate $(u', v') = (u + \widetilde{f}^1(u), v + \widetilde{f}^2(v))$ in Frame2**.**

By means of supervised learning, the input is the object fringe patterns before and after movement; the labeled data is the motion information for each point. Firstly, we adopt a feature encoder network g_θ to learn the motion information encoded by the inputs Frame1 and Frame2, mapping the input images to a dense feature map at a lower resolution. The feature encoder consists of 6 residual blocks, 2 with 1/2 resolution, 2 with 1/4 resolution and 2 with 1/8 resolution [23]. As a result, the encoder g_θ outputs features g_θ: $\mathbb{R}^{H\times W\times 3} \rightarrow \mathbb{R}^{H/8\times W/8\times D}$ with 1/8 resolution and the depth $D = 256$. In addition, a context encoder network h_θ is introduced to extract the features only from the first input image Frame1, which has the same architecture as feature encoder g_θ but is trained without sharing the network weights. The architecture of the encoder is shown in Fig. 3.

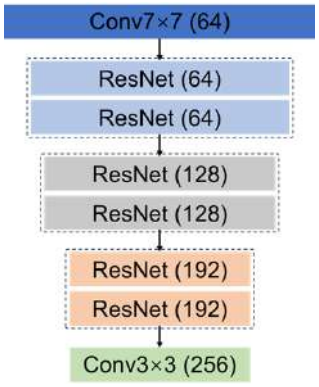

Fig. 3. Architecture of the encoder

The feature encoder network g_θ and the context encoder network h_θ are the first stage of the RAFT network and need to be executed only once.

With the pair of the learned image feature maps $g_\theta(Frame1) \in \mathbb{R}^{H\times W\times D}$ and $g_\theta(Frame2) \in \mathbb{R}^{H\times W\times D}$, we compute the visual similarity by constructing a complete correlation volume between all extracted feature map pairs. In the original RAFT network, only dot product operation is performed on all feature map pairs, which works well for most plain scenes. However, in PSP imaging, the optical flow estimation is affected by the images with periodic sinusoidal fringes.

Optics EXPRESS

The reason is that when the moving object moves in the fringes, its feature points will be affected by the fringe illumination, which will obviously change the light intensity distribution of the feature points. Furthermore, the periodic fringe illumination leads to the original RAFT tracking method being ineffective**.**

To address the problems above, our approach suggests the utilization of the NCC across all pairs of learned feature maps, rather than the dot product, to enhance the network's capacity of learning kinematic information, i.e., optical flow, from two successive fringe patterns under the fringe pattern illumination. Given that the NCC computes the visual similarity by normalizing the effect of varying lighting conditions locally, we found it is more robust to local illumination or intensity distribution variations than the dot-product and can effectively deal with defocusing, scaling, small-motion, and nonlinear illumination variations (which are the main challenge in tracking points in fringe texture images). Moreover, the local NCC operation potentially adapts to the aperture problem in the perspective projection model, helping to appropriately increase the receptive field of subsequent convolution based GRU blocks to significantly improve the motion information representation capability of the whole network. These merits inspire us to build an NCC-based module to encode the kinematic information of the object into a four-dimensional similarity space. As a result, a correlation volume with shape of $\mathrm{W}\times\mathrm{H}\times\mathrm{W}\times\mathrm{H}$ is constructed and then, the last 2 dimensions of the four-dimensional correlation volume are pooled at multiple scales to construct a set of multi-scale volumes. Here we use the kernel sizes 1, 2, 4, and 8 to perform the four times pooling operations with a constant step size, obtaining a four-level correlation pyramid $\{\mathrm{C1}, \mathrm{C2}, \mathrm{C3}, \mathrm{C4}\}$. A lookup operator L_C is defined which generates a feature mapping by indexing the correlation pyramid.

In the next step, a series of optical flows $\{\widetilde{f_1}, \ldots, \widetilde{f_N}\}$ is estimated by updating the operator L_C from an initial starting point $\widetilde{f_0} = 0$, detecting relevant features from the correlation pyramid, as shown in Fig. 4. The detected correlation features are processed by using 2 convolutional layers. In addition, we apply 2 convolutional layers to the optical flow estimation to generate flow features. The update operator takes the optical flow, correlation, and potential hidden states as the input, then outputs the updated $\triangle\widetilde{\mathrm{f}}$ and updated hidden states. The gated activation of the GRU (gated recurrent unit) unit serves as a core part of the update operator by replacing the fully connected layer with a convolution, the hidden state output from the GRU is used to update the optical flow $\triangle\widetilde{\mathrm{f}}$ through the two convolutional layers. In each iteration, an incremental direction $\triangle\widetilde{\mathrm{f}}$ is predicted to the update the current estimate: $\widetilde{f_{k+1}} = \triangle\widetilde{f} + \widetilde{f_{k+1}}$**.**

After several iterations, the resolution of the output optical flow map of the final update operator is 1/8 of the input image. We upsample the optical flow to full resolution by taking the full resolution flow at each pixel to be the convex combination of a 3×3 grid of its coarse resolution neighbors, producing the final optical flow field**.**

We supervise our network throughout the prediction sequence to predict the optical flow $\{\widetilde{f_1}, \ldots, \widetilde{f_N}\}$ and the actual true displacement with exponentially increasing weights. Given the true displacement $\widetilde{f_g}$, the loss for training the proposed network is defined as:

$$\mathrm{Loss} = \sum_{i=1}^{S} \Upsilon^{S-i} \left\| \widetilde{f_g} - \widetilde{f_i} \right\|_1 \tag{12}$$

where Υ is generally set to 0.8.

2.2.2. Post-processing of optical flow data

Based on the improved RAFT optical flow algorithm proposed in this paper, a dense optical flow for all pixels representing the motion information is obtained. Please note that not all the points have the optical flow with high accuracy to meet the measurement accuracy. In the other hand, for rigid object tracking, we do not need all the points on the object to calculate the object

Optics EXPRESS

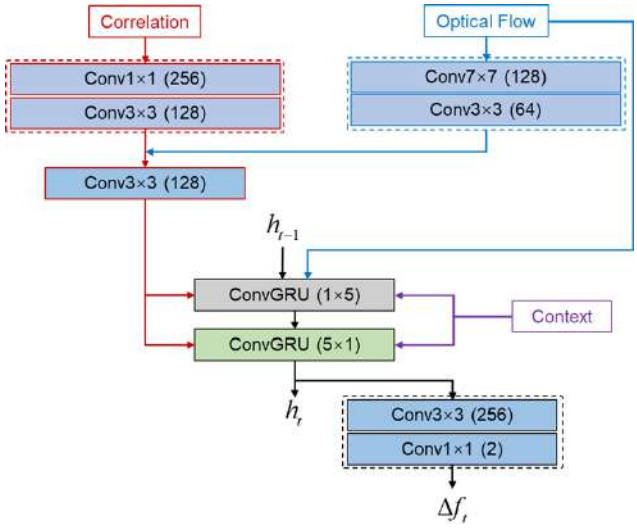

Fig. 4. Architecture of the iterative update block

motion information. To this end, a RANSAC-based iteration selection process is employed to select points with high quality optical flow since the RANSAC algorithm is capable of screening out the valid samples from a set of measured data points [24]. The steps are described as follows.

Step 1: Apply the RANSAC algorithm on the whole dense optical flow data, two sets of points p_1 and q_1 are selected. p_1 is the selected points before motion and q_1 is the corresponding ones after motion. RANSAC can filter out obvious outliers.

Step 2: To further improve the accuracy, an iteration selection process is applied. The rotation matrix R_1 and translation vector T_1 are calculated based on p_1 and q_1'. Then, a new point set q_1 representing the points after motion in theory is obtained based on p_1 and $[R_1, T_1]$.

Step 3: Compare the distance between q_1 and q_1', when the distance is less than a threshold (such as 0.5 pixel), the corresponding point is selected and on the contrary the point is discarded. Two new point sets p_2 and q_2 are obtained with the selected points.

Step 4: A new rotation matrix and translation vector $[R_2, T_2]$ is calculated based on p_2 and q_2;

Step 5: Repeat step 3 to step 5 until all the points meet the requirement of threshold.

With the optical flow with high accuracy, the rotation matrix and translation vector are obtained to describe the object motion and the object is reconstructed based on Eqs. (7)-(11).

3. Experiments

Experiments are implemented to evaluate the performance of the proposed method. The reconstruction system consists of a projector (DLP Tengju TJ20uLB0077G15) with a resolution of 1280 × 720 pixels and a camera (Hikvision MV-CA013-21UC) with the same resolution as shown in Fig. 5. A workstation with a GPU of Nvidia GeForce GTX-3060Ti is used to train the proposed network. 4-step PSP is employed in the experiment. The reconstruction result for the static object (object 1) with traditional PSP is shown in Fig. 6.

Based on the above system, the data for the training of the improved RAFT optical flow network is prepared. 2000 pairs of the object fringe pattern images before movement and after movement are captured for training, validation, and testing in a ratio of 8:1:1. To obtain the motion information used in the label, the images without fringe pattern at the same position are also obtained as shown in Fig. 7. The SIFT algorithm is applied on the images without fringe pattern directly, as shown in Fig. 8, and the rotation matrix and translation vector describing the

Optics EXPRESS

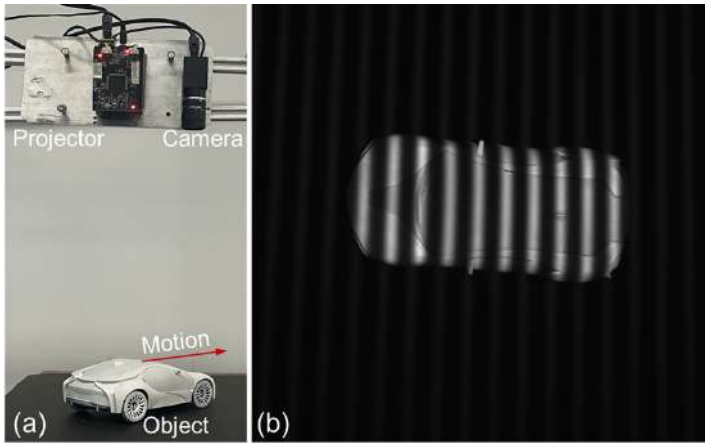

Fig. 5. (a) PSP imaging system and (b) the captured fringe patterns

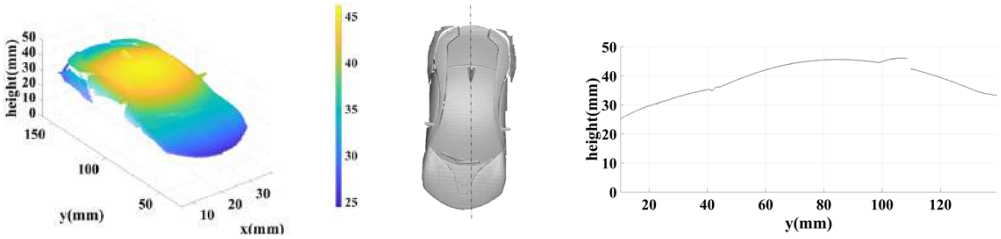

Fig. 6. Reconstruction results of the conventional PSP with the static object. (a) Reconstruction results in Mesh display; (b) front view of the object; (c) The cross section of the dashed line in Fig. 6(b).

movement is calculated. Then, the movement information for each pixel is added for the object images with fringe pattern as the label**.**

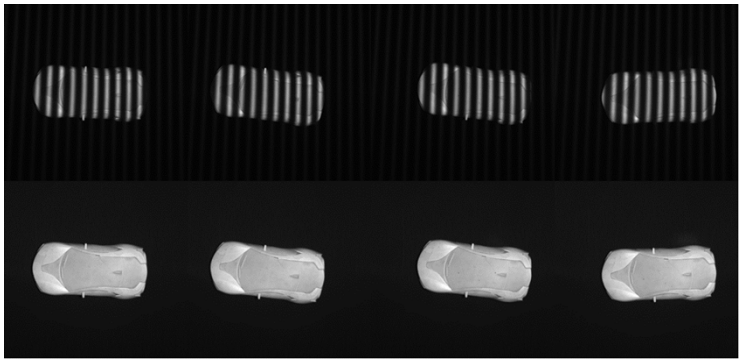

Fig. 7. Captured images of the kinematic object

The fringe patterns before movement and after movement and their corresponding motion information for each pixel are employed to train the improved RAFT optical flow network. In the loss function, the estimated motion information and the correct ones are compared, the motion

Optics EXPRESS

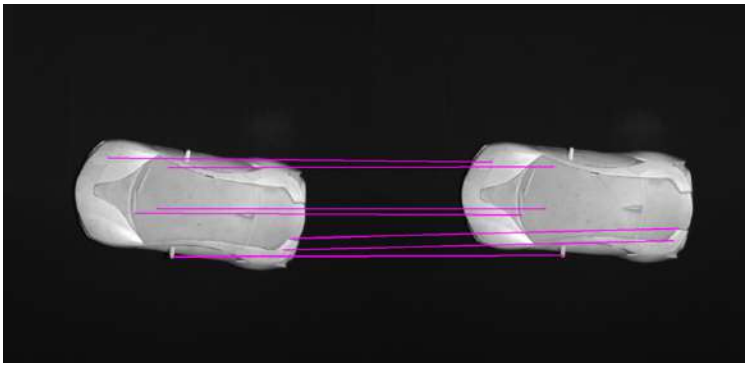

Fig. 8. Matched feature points by SIFT

information is learned by the network and the effect introduced by the fringe pattern illumination is removed. As shown in Fig. 9, the training loss of the network tends to be stable and converges to 0.25 pixels after 900 epochs. We stopped the training process at 1000 epochs, and the elapsed time of the entire training process is about 31 hours.

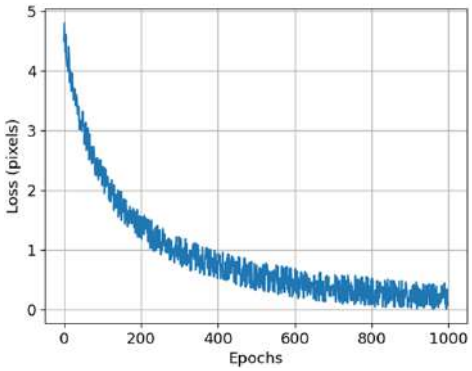

Fig. 9. Training loss curve of the improved RAFT optical flow network

The improved RAFT optical flow network estimates the motion information for every pixel on the object. To demonstrate the performance of the improved RAFT optical flow algorithm, we compare the motion estimates obtained by the original RAFT algorithm with the counterparts obtained by the improved version under the same condition. Because of the real captured fringe patterns, we adopt the SIFT-based method [18] to build the ground truth for the comparison. When the difference between the ground truth and estimated displacement amplitude is less than a half pixel, we count it as the correct point. The correct points for the original RAFT and our improved version are shown in Fig. 10. It is apparent that our improvement on the RAFT optical flow algorithm effectively reduces the effect of fringe patterns on the object surface, obtaining more correct point correspondences. As there are some inaccurate estimates in practice without any prior ground truth, the next step requires selecting the accurate point correspondences from the predicted optical flow data under a given error threshold.

As the number of the correct point is much bigger than the error ones, the RANSAC algorithm and an iteration selection process are applied to select the correct point from all points on the object surface. The optical flow with high accuracy (12 thousand in this paper) are selected to calculate the rotation matrix and translation vector. For the convenience of display, the partial

Optics EXPRESS

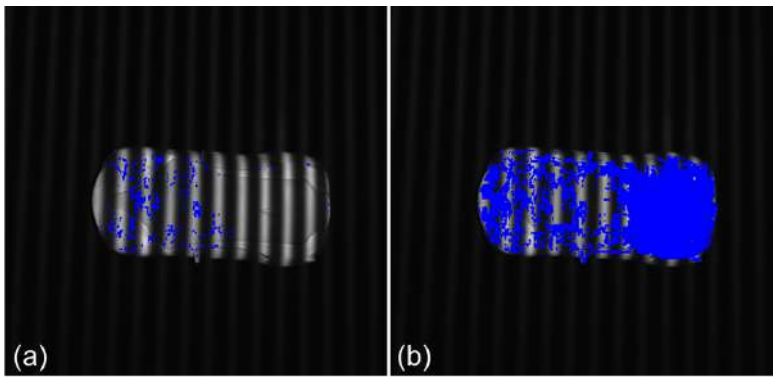

Fig. 10. Correct points estimated by (a) the original RAFT network and (b) our NCC-enhanced version.

correct points obtained with the fringe pattern image are shown in Fig. 11. It can be seen that the object motion is tracked successfully.

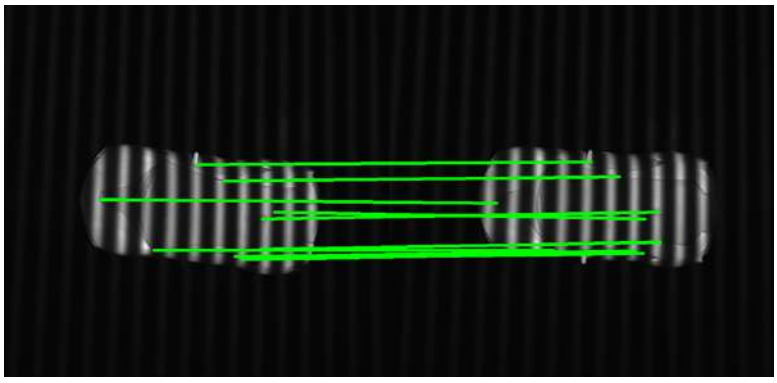

Fig. 11. Partial correct points obtained by the fringe pattern image.

With the obtained motion information, the surface of the kinematic object is reconstructed, and the result is shown in Fig. 12.

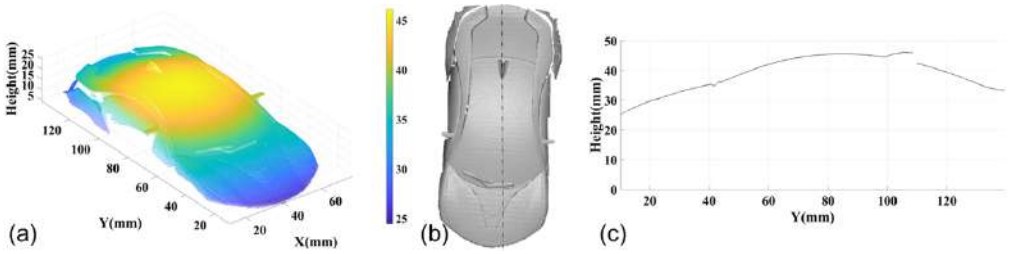

Fig. 12. (a) Surface reconstruction of the object 1, (b) its top view, and (c) the cross section of the dashed line in (b).

Different objects (object 2, object 3, object 4 and object 5) are reconstructed to further verify the performance of the proposed method. The tracked point correspondences and reconstruction results are shown in Fig. 13.

Optics EXPRESS

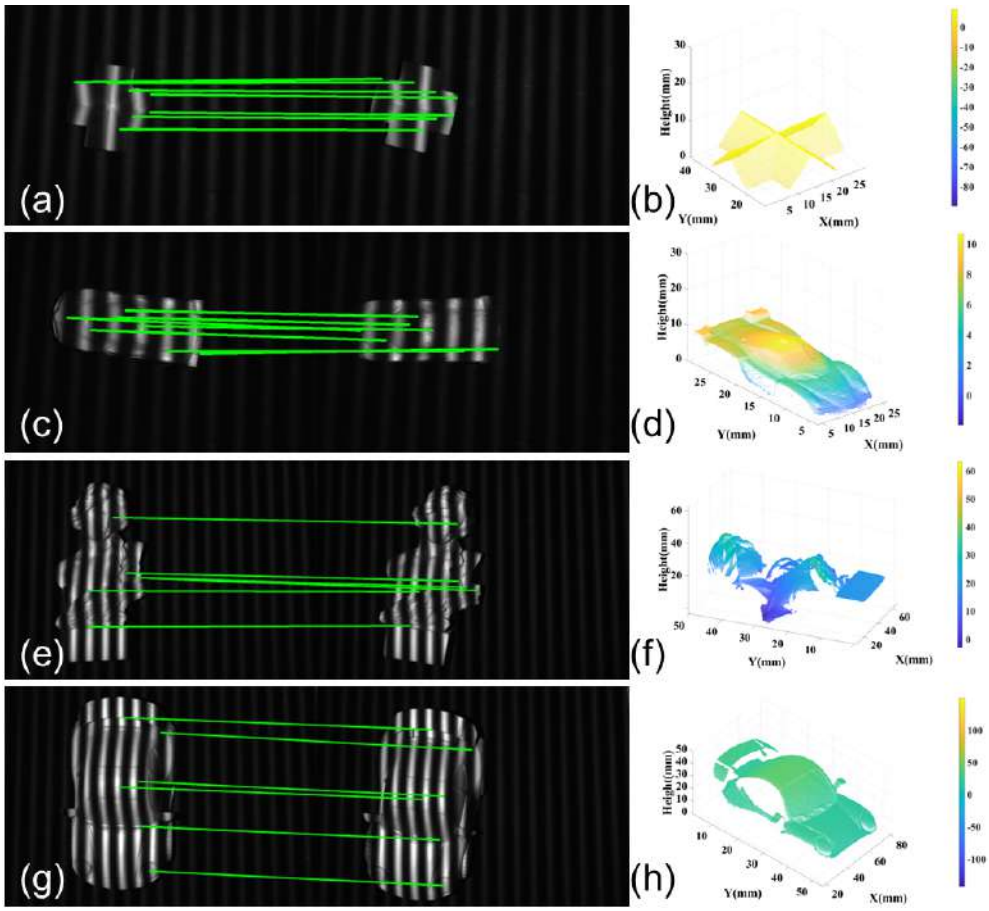

Fig. 13. Reconstructed surface for different objects: (a), (c), (e), and (g) are the partial correct point correspondences obtained from a pairs of fringe texture image for the moving objects 2-5; (b), (d), (f), and (h) are the reconstructed surfaces of the four objects correspondingly.

To quantitatively evaluate the effectiveness of the proposed method, the RMS (root mean square) error is calculated for each object, where the ground truth is the reconstructed 3D shape of the static object counterpart. We also compared the performance of the proposed method with the traditional SIFT-based method. The results, listed in Table 1, show that the proposed method achieves comparable performance to the traditional SIFT-based method.

Table 1. RMS errors of our and SIFT-based methods

Methods	Object 1	Object 2	Object 3	Object 4	Object 5
Our method	0.091 mm	0.098 mm	0.089 mm	0.094mm	0.082mm
SIFT-based method	0.080 mm	0.089 mm	0.081 mm	0.085mm	0.078mm

From the experiment results in Fig. 13 and Table 1, one can see that the trained RAFT network is capable of generalizing to sensing the surfaces of different moving objects within the framework of PSP 3D imaging. This has been made possible by adopting the following strategies in addition to the model selection. First, we employ the real PSP imaging system in Fig. 5 to collect the fringe texture images from real scenes, and the established dataset covers different viewing angles, spatial positions, lighting, and noise conditions, etc. Second, The L1 distance is adopted

Optics EXPRESS

to measure the training loss in Eq. (12), providing an underlying regularization which helps to restrain the overfitting. In addition, because of the sparsity of L1 loss, the complexity of the network can be reduced. Both help to improve the generalization of the proposed RAFT model. Finally, the early stopping strategy is used in the training phase by monitoring the model's performance on validation set, preventing the model from overfitting to the training data.

4. Conclusions

In this paper, a new deep learning-based fringe texture image tracking method is proposed for PSP technique to reconstruct the surface point clouds of kinematic targets. An improved RAFT optical flow algorithm that is immune to the influence of the fringe pattern illumination is proposed, allowing for retrieving the moving feature points from the fringe pattern image directly. The RANSAC algorithm and an iteration selection process are employed to select the inferenced point correspondences with high quality optical flow. Compared with the existing method, the proposed method is capable of reconstructing the surfaces of kinematic objects undergoing any 2D movement with affordable monochrome equipment. In future work, we will try to reconstruct the surface of objects with complex 3D movement by developing some methods for estimating the 3D kinematic parameters.

Funding. National Natural Science Foundation of China (12002197, 62375078); General science foundation of Henan Province (222300420427); Cultivation Program for Young Backbone Teachers in Henan University of Technology.

Disclosures. The authors declare no conflict of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. A. H. Nguyen and Z. Wang, "Time-Distributed Framework for 3D Reconstruction Integrating Fringe Projection with Deep Learning," Sensors **23**(16), 7284 (2023).
2. V. Rajan, M. Koeva, M. Kuffer, *et al.*, "Three-Dimensional Modelling of Past and Present Shahjahanabad through Multi-Temporal Remotely Sensed Data," Remote Sens. **15**(11), 2924 (2023).
3. C. Liu, L. Ji, L. Zhu, *et al.*, "Present-Day Three-Dimensional Deformation across the Ordos Block, China, Derived from InSAR, GPS, and Leveling Observations," Remote Sens. **15**(11), 2890 (2023).
4. Y. Shen, R. Huang, B. Hua, *et al.*, "Automatic Tree Height Measurement Based on Three-Dimensional Reconstruction Using Smartphone," Sensors. **23**(16), 7248 (2023).
5. C. Zuo, S. Feng, L. Huang, *et al.*, "Phase shifting algorithms for fringe projection profilometry: A review," Opt Lasers Eng . **109**, 23–59 (2018).
6. S. Zhang, "High-speed 3D shape measurement with structured light methods: A review," Opt Lasers Eng . **106**, 119–131 (2018).
7. E. Shoji, A. Komiya, J. Okajima, *et al.*, "Three-step phase-shifting imaging ellipsometry to measure nanofilm thickness profiles," Opt Lasers Eng . **112**, 145–150 (2019).
8. L. Felipe-Sesé and F. A. Díaz, "Damage methodology approach on a composite panel based on a combination of Fringe Projection and 2D Digital Image Correlation," Mechanical Systems and Signal Processing **101**, 467–479 (2018).
9. B. Li and S. Zhang, "Superfast high-resolution absolute 3D recovery of a stabilized flapping flight process," Opt. Express **25**(22), 27270–27282 (2017).
10. Y. Wang, Z. Liu, C. Jiang, *et al.*, "Motion induced phase error reduction using a Hilbert transform," Opt. Express **26**(26), 34224–34235 (2018).
11. S. Feng, C. Zuo, T. Tao, *et al.*, "Robust dynamic 3-D measurements with motion-compensated phase-shifting profilometry," Opt Lasers Eng . **103**, 127–138 (2018).
12. S. Zhang and S. T. Yau, "High-speed three-dimensional shape measurement system using a modified two-plus-one phase-shifting algorithm," Opt. Eng. **46**(11), 113603 (2007).
13. Z. Liu, P. C. Zibley, and S. Zhang, "Motion-induced error compensation for phase shifting profilometry," Opt. Express **26**(10), 12632–12637 (2018).
14. Y. Wei, L. Lu, and J. Xi, "Reconstruction of moving object with single fringe pattern based on phase shifting profilometry," Opt. Eng. **60**(08), 084106 (2021).
15. K. Wu, M. Li, L. Lu, *et al.*, "Reconstruction of isolated moving objects by motion-induced phase shift based on PSP," Appl. Sci. **12**(1), 252 (2021).
16. Y. Wang, V. Suresh, and B. Li, "Motion-induced error reduction for binary defocusing profilometry via additional temporal sampling," Opt. Express **27**(17), 23948–23958 (2019).

Optics EXPRESS

17. L. Lu, J. Xi, Y. Yu, *et al.*, "New approach to improve the accuracy of 3-D shape measurement of moving object using phase shifting profilometry," Opt. Express **21**(25), 30610–30622 (2013).
18. L. Lu, Z. Jia, W. Pan, *et al.*, "Automated reconstruction of multiple objects with individual movement based on PSP," Opt. Express **28**(19), 28600–28611 (2020).
19. L. Lu, Y. Ding, Y. Luan, *et al.*, "Automated approach for the surface profile measurement of moving objects based on PSP," Opt. Express **25**(25), 32120–32131 (2017).
20. L. Lu, Z. Jia, Y. Luan, *et al.*, "Reconstruction of isolated moving objects with high 3D frame rate based on phase shifting profilometry," Opt. Commun. **438**, 61–66 (2019).
21. L. Lu, V. Suresh, Y. Zheng, *et al.*, "Motion induced error reduction methods for phase shifting profilometry: A review," Opt Lasers Eng . **141**, 106573 (2021).
22. M. Duan, Y. Jin, H. Chen, *et al.*, "Automatic 3-D measurement method for nonuniform moving objects," IEEE Trans. Instrum. Meas. **70**, 1–11 (2021).
23. Z. Teed and J. Deng, "Raft: Recurrent all-pairs field transforms for optical flow," In *2020 European Conference on Computer Vision*, (Springer International Publishing, 2020), pp. 402–419.
24. H. Zhang, G. Zheng, and H. Fu, "Research on image feature point matching based on ORB and RANSAC algorithm," In *Journal of Physics: Conference Series*, (IOP Publishing, 2020), pp. 012187.