IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 29, 2020, accepted December 15, 2020, date of publication December 18, 2020,
date of current version December 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3045764

Large Displacement Optical Flow Estimation Based on Robust Interpolation of Sparse Correspondences

SHIDONG SHI[1], DAOWEN ZHANG[1], CONGXUAN ZHANG[1,2], (Member, IEEE), ZHEN CHEN[1], CHENG FENG[1], AND BINGBING FAN[1]

[1]Key Laboratory of Nondestructive Testing, Ministry of Education, Nanchang Hangkong University, Nanchang 330063, China
[2]Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Corresponding author: Congxuan Zhang (zcxdsg@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2020YFC2003800, in part by the National Natural Science Foundation of China under Grant 61772255, Grant 61866026, and Grant 61866025; in part by the Advantage Subject Team Project of Jiangxi Province under Grant 20165BCB19007, in part by the Outstanding Young Talents Program of Jiangxi Province under Grant 20192BCB23011, in part by the National Natural Science Foundation of Jiangxi Province under Grant 20202ACB214007, in part by the Aeronautical Science Foundation of China under Grant 2018ZC56008, in part by the China Postdoctoral Science Foundation under Grant 2019M650894, and in part by the Innovation Fund Designated for Graduate Students of Nanchang Hangkong University under Grant YC2019038.

ABSTRACT Recently, the interpolation of correspondences method has been widely used in optical flow estimation, because it produces an accurate flow field and costs little runtimes. However, most of the existing matching-based optical flow methods are usually susceptible to non-rigid motion and large displacements. We propose in this article a large displacement optical flow estimation method based on robust interpolation of sparse correspondences, named Riscflow. First, we utilize the deep matching model to achieve an initial matching result of two consecutive frames, and then we exploit a grid-based motion statistics optimization scheme to remove the outliers from the initial matching field. Second, we propose a random forest-based motion boundary extraction model and construct a sparse-to-dense interpolation method by using the boundary information to prevent the dense matching field from edge-blurring. Third, we design a global optical flow estimation method by using an energy function to optimize the dense matching field. Finally, we respectively run the proposed method on the MPI-Sintel and UCF101 databases to conduct a comprehensive comparison with some state-of-the-art optical flow approaches including the variational methods, the matching-based methods, and the deep learning-based methods. The comparison results demonstrate that the proposed method has high accuracy and good robustness of optical flow estimation, and especially gains the benefit of edge-preserving under non-rigid motion and large displacements.

INDEX TERMS Optical flow, edge-preserving interpolation, sparse correspondences, global optimization, large displacements.

I. INTRODUCTION

Estimating optical flow from consecutive frames is a research core of image processing and computer vision, because optical flow includes the image motion and structural information of the observed objects and scenes. Nowadays, optical flow computation technology has been widely used in robot navigation [1], video events detection and analysis [2], unmanned aerial vehicle [3], human action recognition [4], and many other areas [5]–[7].

After the pioneering work of Horn and Schunck [8], a large number of studies have been presented to improve the accuracy and robustness of optical flow computation. These existing methods can be roughly divided into three categories: (1) the variational optical flow estimation approach, (2) the matching-based optical flow estimation approach, and (3) the deep learning-based optical flow estimation approach.

In the early research, the variational method was the most popular approach in optical flow estimation, because it produces an accurate and dense flow field. However, the existing variational optical flow methods are incapable of dealing with non-rigid motion and large displacements. Recently, with the significant success of convolutional neural networks in many vision-related tasks, the deep learning-based method becomes increasingly popular in optical flow estimation. Although the accuracy an robustness of deep learning-based

The associate editor coordinating the review of this manuscript and approving it for publication was Sudipta Roy.

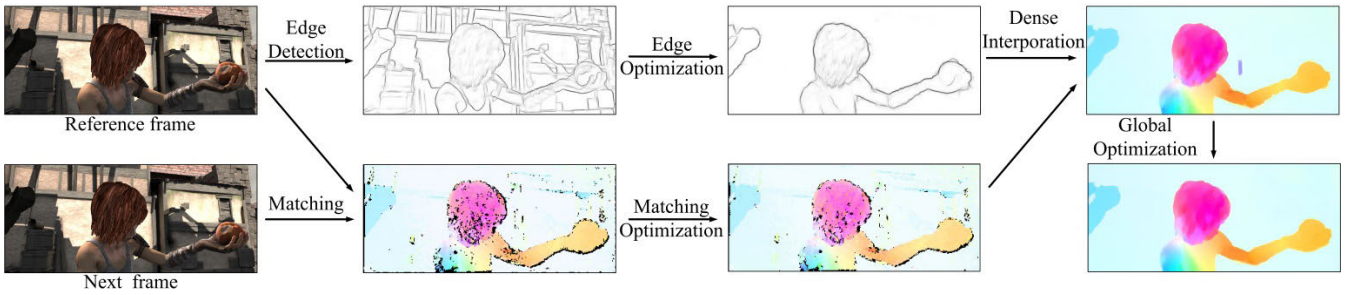


FIGURE 1. Overall flowchart of the proposed method. Given two input frames, we compute matches and boundaries between the reference and next frames, and then optimize the matches and interpolate the spares matching field by using the boundaries information. Finally, we utilize a global energy function to achieve an accurate and robustness optical flow field.

optical flow methods have been greatly improved, most of these methods require supervision and may have difficulty to be directly applied to real-world data.

Being different with the variational and deep learning-based methods, the matching-based method determines optical flow through the correspondences of the pixels between the consecutive frames, which prompts the matching-based optical flow methods to be robust under illumination changes. In spite of the matching-based optical flow methods are reliable in real-world scene, these matching-based methods are usually susceptible to non-rigid motion and large displacements because either the non-rigid motion or large displacements may give rise to the outliers in the matching result. Furthermore, the existing methods are prone to cause the issue of edge-blurring, due to the interpolation schemes of the traditional matching- based methods only depend on the distances of the pixels.

To improve the accuracy and robustness of optical flow estimation under large displacements and address the issue of edge-blurring, we propose a large displacement optical flow estimation method based on robust interpolation of sparse correspondences, named Riscflow. The experimental results demonstrate that the proposed method has high accuracy and good robustness for optical flow computation, especially gains the capacity of edge-preserving. Fig. 1 illustrates the overall flowchart of the proposed Riscflow method. Our contributions are concluded as follows:

- We exploit a robust matching framework by using grid-based motion statistics to remove the outliers from the initial deep matching field. The presented optimization matching scheme is able to improve the robustness of the matching results.
- We propose a sparse-to-dense interpolation method by using boundary information to restrain the interpolating near the image edge and motion boundaries. This interpolation method prevents the dense matching field from edge-blurring.
- We design a global energy function to gain the optical flow by using the dense matching field as the initialization. The proposed global optimizing scheme prompts to an accurate dense optical flow field.

The remainder of this paper is organized as follows: In section II, we briefly review the past works of optical flow computation. Section III describes the robust sparse matching field computing scheme based on deep matching. Section IV presents the large displacement optical flow estimation method with sparse-to-dense interpolation. The experimental results and discussions are presented in Section V. Finally, Section VI concludes the project.

II. RELATED WORKS

A. VARIATIONAL OPTICAL FLOW METHODS

Tracing back to the early research, Horn and Schunck [8] proposed the first variational optical flow estimation model by combining a data term with a smoothing term. After that, most of the studies of flow field estimation focused on how to design an energy function [9]. For the data term, some studies [10], [11] recommended to incorporate a gradient constancy assumption into the data term due to the classical brightness constancy assumption is incapable under illumination changes. To address the issue of image noises, the combination of global and local optimizations [12], [13] was presented to improve the robustness of optical flow computation. To remove the outliers of the flow field, many studies [14], [15] advised to replace the classic L^2 norm of the data term by the L^1 norm.

For the smoothing term, because the homogeneous diffusion strategy used in the original HS model [8] tends to produce the issue of edge-blurring, several researches [16], [17] proposed the image-driven smoothing strategies by using the image gradient to regularize the flow diffusion. However, these image-driven regularization models may cause the problem of over-segmentation in the textured image areas. To preserve the motion boundaries, some publications [18], [19] presented the flow-driven smoothing strategies by utilizing the motion information to regularize the flow diffusion. Because not every image edge coincides with a motion boundary, the combination of flow- and image-driven diffusing strategies was recommended to preserve both image edges and motion boundaries [20].

With the increasement of the model complexity, many studies focused on how to optimize the minimization of the energy function. For instance, some researches [11], [21]

exploited the coarse-to-fine computation scheme to cope with large displacements. Moreover, the non-local constraint [22] was employed to remove the outliers during the optical flow computation. Because motion occlusions may suppress the accuracy of optical flow, several studies [23], [24] constructed the occlusion detection method to modify the optical flow model. In recent years, many spatial filtering strategies [25]–[27] were recommended to deal with image noise and edge-blurring. Some studies proposed a local illumination change model [28] to deal with the weakly textured scenes. Furthermore, an adaptive dual fractional-order variational optical flow method [29] was presented to solve the issues of insufficient illumination and illumination changes. Despite that the accuracy and robustness of variational optical flow estimation have been significantly improved, these variational methods usually require a mass of iterations to minimize the objective function. This computation process dramatically increases the time consumption, which may limits the application of variational optical flow methods.

B. CNN-BASED OPTICAL FLOW METHODS

Recently, inspired by the success of convolutional neural networks (CNNs) in many tasks of computer vision and image processing, the CNN-based optical flow estimation method has become a research hotspot. Dosovitskiy *et al.* [30] constructed the first CNN-based optical flow model named FlowNet. Their study verifies the feasibility of directly estimating the optical flow through a generally convolutional architecture. This significant achievement encourages the following CNN-based optical flow methods, including unsupervised, supervised and semi-supervised models [9].

To improve the performance of the FlowNet method in estimating accuracy, a larger model named FlowNet2.0 [31] was presented by stacking several FlowNetC and FlowNetS [30] networks. Although the FlowNet2.0 method achieved a good result on accuracy, it is more prone to overfitting due to its large network. To improve the computation efficiency, some publications [32]–[34] constructed the lightweight cascaded networks to reduce the number of network parameters. To ensure the robustness of optical flow, a CNN-based patch matching approach using a novel threshold loss [35] was presented to cope with motion occlusions. Moreover, some studies [36]–[40] exploited the feature pyramid-based networks to improve the performance of optical flow estimation under large displacements. Recently, a recurrent all-pairs field transformation [41] has been proven to be an efficient way to improve the accuracy of supervised optical flow methods.

Because the supervised methods usually require a mass of labeled datasets to train the networks, this may limit these methods to be directly applied to the real-world scene. To overcome the abovementioned limitation of the supervised methods, the unsupervised learning methods [42] were exploited to estimate optical flow without the ground truth. To improve the performance of the unsupervised optical flow estimation networks, a typical study [43] firstly modeled the occlusions, and then proposed a new warping way to facilitate the learning of large motion. Another method [44] incorporated the occlusion information into the training loss function and compensated the occluded regions by using multi-frames. To balance the estimation accuracy and training datasets, the semi-supervised learning method [45] was recommended to predict the flow field by using a small labeled dataset. For instance, Lai *et al*. [46] presented a semi-supervised learning model to estimate optical flow by utilizing a generative adversarial network, which is able to learn optical flow with both labeled and unlabeled datasets. Although the unsupervised and semi-supervised learning methods are able to train the networks with unlabeled datasets, their computation accuracy is still far behind the supervised optical flow methods.

C. MATCHING-BASED OPTICAL FLOW METHODS

Being different with the classical variational methods, the matching-based methods usually estimate flow filed by using the feature similarity between the consecutive frames. This encourages the matching-based methods achieve the better robustness under illumination changes and motion occlusions. To handle the limitations of variational optical flow methods in estimating large displacements, Brox *et al.* [47] incorporated a descriptor-based matching term into the traditional variational energy function. The proposed hybrid model penalizes the differences between the variational flow field and the matching field, which effectively improves the accuracy of optical flow in regions of large displacements. Afterwards, Xu *et al*. [48] proposed an extended coarse-to-fine computation framework to fuse the sparse matching field and the estimated flow at each pyramid layer. Their method performs a good performance on motion boundaries. Because the mismatched pixels will decrease the accuracy of optical flow estimation, Stoll *et al.* [49] proposed a computation strategy to reduce the interference of the mismatched pixels based on the self- adaptive additional constraints. To improve the matching accuracy in untextured regions, Weinzaepfel *et al*. [50] replaced the classical rigid matching scheme by a non-rigid patch matching framework, which significantly enhances the matching reliability and estimation accuracy.

To improve both of the accuracy and robustness of optical flow estimation under large displacements and occlusions, Lempitsky *et al.* [51] proposed a discrete- continuous optimization for optical flow, which the proposed method fuses multiple optical flows to obtain a robust dense flow field. Inspired by the layered pyramid computation framework, Hu *et al.* [52] exploited a coarse-to-fine patch matching strategy to cope with the large displacements. Their method performs a competitive result on some public databases. Moreover, Zu *et al*. [53] adopted a context-adaptive matching scheme to improve the accuracy of the matching result under illumination changes, deformations and occlusions. Furthermore, Zhang *et al.* [54] constructed a large displacement flow

field estimation approach by using similarity transformation based dense correspondence. Their method improves the accuracy and robustness of optical flow computation under large displacements and motion occlusions. Afterwards, Revaud *et al.* [55] recommended a pixel-distance-based sparse-to-dense interpolation scheme to address the issue of edge-blurring. Moreover, Hu *et al.* [56] presented a robust interpolation method to deal with the input matching noises by using a superpixel segmentation optimization scheme. To improve the accuracy of dense flow field estimation, Li *et al.* [57] proposed a pyramidal gradient matching approach to achieve the highly accurate and efficient optical flow estimation. Furthermore, Chen *et al.* [58] presented a segmentation-based patch matching framework to cope with the issue of over-segmentation. Yang *et al.* [59] proposed a maximum likelihood function calculation method which increases the robustness of the matching- based optical flow estimation methods.

Matching-based optical flow estimation methods have shown great advantages over the traditional variational methods. However, most of the existing matching-based methods are usually susceptible to non-rigid motion and large displacements. To address the abovementioned issues, we exploited in this paper a large displacement optical flow estimation method based on robust interpolation of sparse correspondences. The proposed method will be described in detail in the following sections.

III. ROBUST SPARSE MATCHING FIELD BASED ON DEEP MATCHING

A. DEEP MATCHING

As shown in Fig. 2, in order to improve the accuracy and robustness of patch matching in regions of non-rigid motion and large displacements, the deep matching method [60] firstly divides the traditional sampling window into N subregions, and then optimizes the position of each subregion according to the similarities between the various subregions. Thus, the correspondences of the pixels between the consecutive frames are determined through the subregions positions.

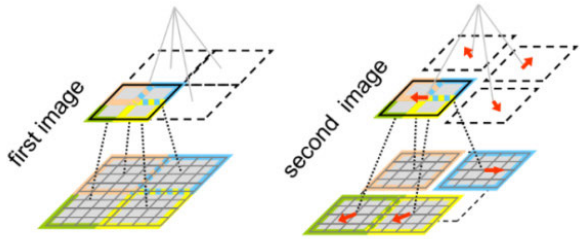
(a) (b) (c)

FIGURE 2. Illustration of the deep matching sampling windows based on region division. (a) Sample window of the reference frame, (b) Sample window of the traditional matching method, (c) Sample window of the deep matching method.

Given the consecutive two frames I_0 and I_1, we decompose the frames I_0 and I_1 into N subregions. Each subregion is composed of four neighboring pixels. Thus, the matching relationship of a subregion R in frame I_0 with its corresponding subregion R' in frame I_1 is defined as following:

$$\mathrm{Sim}(R, R') = \frac{1}{16}\sum_{i=0}^{3}\sum_{j=0}^{3} R_{i,j}R'_{i,j} \tag{1}$$

where $\mathrm{Sim}(R, R')$ indicates the matching relationship of the subregions R and R'. The notations $R_{i,j}$ and $R'_{i,j}$ represent the central pixels of the subregions R and R', respectively.



(a) (b)

FIGURE 3. Illustration of the pyramid sampling based deep matching scheme. (a) Subregion aggregation of the first frame, (b) Corresponding pixel searching of the second frame.

With the matching relationships of the subregions between the consecutive two frames, the implementation process of the deep matching method is concluded as the following two steps: first, as shown in Fig. 3(a), every four neighboring subregions are aggregated at the upper layer during the image pyramid to determine the matching relationships of the larger subregions. Second, by defining the central pixels of the matching subregions as the corresponding pixels, the sparse matching field is achieved by searching the corresponding pixels from the top layer to the bottom layer during the image pyramid, as shown in Fig. 3(b). Because the deep matching model performs well in regions of non-rigid deformation and large displacements, we utilize the deep matching scheme to achieve the initial matching field of the consecutive two frames.

B. NEIGHBORING SUPPORT-BASED MATCHING OPTIMIZATION

Although the deep matching method improves the accuracy of pixel correspondences in areas of non-rigid motion and large displacements, its matching result may include incorrect matching pixels due to the image noises and illumination changes. To remove the incorrect matching pixels, we exploit a neighboring support based matching optimization scheme to optimize the initial matching result [61]. The proposed optimization method is able to effectively improve the accuracy and reliability of the pixel correspondences.

Assume that the motion between the consecutive two frames I_0 and I_1 is continuous and smooth, the motion of the center pixel x_i in a local region should be consistent with that of its neighboring support pixels. Therefore, the number of the neighboring support pixels which are consistent with the

matching center pixel is defined as follows:

$$S_i = \sum_{i=1}^{k} \left| \chi_{a^k b^k} \right| \tag{2}$$

where K indicates the number of the neighboring regions which have the same movement with the matching center pixel. The notations a^k and b^k denote the k^{th} matching regions which have the same geometric relationship with the matching center pixel between the consecutive frames. The symbol $\left| \chi_{a^k b^k} \right|$ represents the number of the matching pixels between the matching regions a^k and b^k. Because the correspondences of various pixels are independent, the number S_i of the neighboring support pixels should be approximately binomial distribution as bellows:

$$S_i \sim \begin{cases} B(Kn, p_t) & if\ x_i\ is\ true \\ B(Kn, p_f) & if\ x_i\ is\ false \end{cases} \tag{3}$$

where n represents the average number of the matching pixels in a local region. The natation $p_t = D_t + \beta(1 - D_t)m/M$ indicates the probability that a pair of pixels is corresponding between consecutive frames when their located regions are matched, and the notation $p_f = \beta(1 - D_t)m/M$ denotes the probability that a pair of pixels is corresponding between consecutive frames when their located regions are non-matched. In the above definition, the symbol D_t indicates the matching accuracy of the deep matching result, the symbol β is a probability factor, the notation m denotes the number of matching pixels in a local region, and the notation M represents the total number of the matching pixels in the sparse matching field estimated by the deep matching method.

As shown in Eq. (3), the difference of the numbers of the neighboring support pixels between the correct and incorrect matching pixels is usually large. We distinguish whether a pixel is a correct matching pixel by counting its neighboring support pixels, thus the identification indicator of a pixel based on its neighboring support pixels is defined as follows:

$$p = \frac{Knp_t - Knp_f}{\sqrt{Knp_t(1 - p_t)} + \sqrt{Knp_t(1 - p_f)}} \tag{4}$$

where P denotes the identification of a pixel based on its neighboring support pixels. According to the Eq. (4), the relationship between the value of the identification indicator P and the other variables can be expressed as follows:

$$\begin{cases} P \propto \sqrt{Kn} \\ \lim\limits_{t \to 1} P \to \infty \end{cases} \tag{5}$$

As shown in Eq. (5), the identification indicator P trends to be infinity with the increasing of the number of the neighboring support pixels. In addition, the performance of the neighboring support based matching optimization scheme is correlative with the accuracy of the deep matching result, because a better matching result produces the more accurate neighboring support pixels.

C. GRID-BASED NEIGHBORING SUPPORT OPTIMIZATION

Despite that the neighboring support optimization scheme is able to improve the accuracy and robustness of the sparse matching field, the pixel-level optimizing process may significantly increase the time consumption. To balance the accuracy and runtime of the optimization scheme, we construct a grid-based framework to implement the neighboring support optimization, which leads to a high-efficiency optimizing process.

First, we divide the input consecutive two frames into a set of $n \times n$ non-overlapping grids. Then, we define the grid that includes the most matching pixels between the consecutive frames as the candidate matching grid. Afterwards, we calculate the matching confidence coefficient of the candidate matching grid as follows:

$$S_{i,j} = \sum_{k=1}^{K} \left| \chi_{i^k, j^k} \right| \tag{6}$$

where $S_{i,j}$ denotes the matching confidence coefficient of the grid i in the reference frame and its corresponding grid j in the next frame. The symbol K indicates the number of the neighboring grids of the grid i. The notation i^k and j^k indicate the k^{th} matching grids between the consecutive frames, and the symbol represents the number of matching pixels in the matching grids i^k and j^k.

With the calculated matching confidence coefficients of the candidate matching grids, we construct a self-adaptive threshold $\tau_{i,j}$ to distinguish whether a candidate matching grid is a correct matching grid, as shown in the follows:

$$\{i, j\} = \begin{cases} Correct, & if\ S_{i,j} \geq \tau_{i,j} \\ Incorrect, & other, \end{cases} \quad 1 \leq i, j \leq N \tag{7}$$

where N denotes the number of the non-overlapping grids. $\tau_{i,j} = \alpha\sqrt{n_{i,j}}$ is the constructed self-adaptive threshold for distinguishing the candidate matching grids, in which the notation $n_{i,j}$ represents the number of the matching pixels in grids i and j, and the symbol α is a weight factor.

With the achieved matching grids, we aggregate the matching pixels located in the correct matching grids to be the valid corresponding pixels, and remove the remainder matching pixels from the initial matching result to gain the final robust sparse matching field. For an illustration, Fig. 4 displays the comparison result between the initial matching field of the deep matching method and the optimized matching field using our method. As shown in Fig. 4, the proposed grid-based neighboring support optimization scheme eliminates most of the incorrect matching pixels and significantly improves the accuracy and robustness of the matching field.

IV. DENSE FLOW FIELD ESTIMATION WITH EDGE-PRESERVING INTERPOLATION

A. EDGE-PRESERVING INTERPOLATION WITH MOTION BOUNDARY

Given the computed sparse matching field, most of the existing matching-based methods estimate the dense flow

FIGURE 4. Illustration of the proposed grid-based neighboring support optimization in improving the accuracy of matching field, where the blue mark indicates the correct matching pixel and the red mark denotes the incorrect matching pixel.

field through the sparse-to-dense interpolation [55], [56]. However, this oversimplified interpolation scheme may blur the motion boundaries. To deal with the abovementioned issue, we exploit an edge-preserving spare-to-dense interpolation framework with motion boundary predicted by the random forest scheme [62]. The proposed method is able to improve the accuracy of the dense flow field and preserve the motion boundary.

FIGURE 5. Estimated motion boundary of the MPI-Sintel datasets. Top is the reference frame and bottom is the estimated motion boundary.
FIGURE 6. Flow fields of the various ablation models on some MPI-Sintel training sets. From top to bottom: sequences of alley_2, cave_4 and market_6.

To acquire the motion boundary, we firstly utilize a structured random forest model to extract the motion boundary from the consecutive frames [63]. Fig. 5 shows the estimated motion boundary of the MPI-Sintel datasets. Assume p_m is a matching pixel in a local region centered at pixel p, we define an edge-preserving distance $D(p_{\mathrm{m}}, p)$ between the central pixel p and its neighboring pixel p_m as follows [64]:

$$D(p_{\mathrm{m}}, p) = \inf_{\Gamma \in \rho_{p_m,p}} \int_{\Gamma} C\,(p_s)\, dp_s \tag{8}$$

where $\rho_{p_m,p}$ denotes the set of all possible paths between the pixels p_m and p. The symbol $C(p_m)$ indicates the cost of one possible path crossing pixel p_s. If the pixel $C(p_m) \to 0$. Hence, the distance of two pixels which belong to the same motion region will be low, and the distance of two pixels which belong to the different motion region will be large. Because each pixel is interpolated by using its neighbors, the proposed interpolation scheme is able to preserve the motion boundaries.

Given the edge-preserving distance between the center and neighboring pixels, the sparse correspondence field is interpolated by using a locally weighted affine estimator, as shown in the follows:

$$W_{LA}(p) = A_p p + t_p \tag{9}$$

where p represents any pixel in the first frame. The symbols A_p and t_p are the affine transformation parameters of the pixel p, which the parameters A_p and t_p are computed by using an overdetermined equations as follows:

$$k_D(p_m, p)(A_p p_m + t_p - p'_m) = 0 \tag{10}$$

where p_m and p'_m denote a set of the corresponding pixels between the consecutive two frames. The notation $k_D(p_m, p) = \exp(-\alpha D(p_m, p))$ represents a Gaussian kernel function for the edge-preserving distance D with a parameter α.

Because each pixel is interpolated based on its neighboring pixels, we limit the matching pixels used in the interpolation at a pixel p to its K nearest neighbors according to the edge-preserving distance $D(p_m, p)$. Thus, the pixel belonging to the same motion layer is close to the other pixels at the same motion layer, and it is far away from the pixels outside the same motion layer. This encourages the interpolation scheme to preserve the motion boundaries.

B. GLOBAL OPTIMIZATION

After the edge-preserving sparse-to-dense interpolation, the dense correspondence field is achieved. However, it may include a few of local optimums. To achieve the global solution, we utilize an energy function to optimize the dense correspondence field, as shown in the follows:

$$E(u, v) = \iint_{\Omega} \varphi\left((I_x u + I_y v + I_t)^2 + \left(\nabla I_x u + \nabla I_y v + \nabla I_t\right)^2\right) + J(|\nabla I|) \cdot \int\int_{\Omega} \varphi(|\nabla u|^2 + |\nabla v|^2) dx dy \tag{11}$$

where $I_x u + I_y v + I_t = 0$ and $\nabla I_x u + \nabla I_y v + \nabla I_t = 0$ are the brightness and gradient constancy assumptions, respectively. The notation $\varphi(x) = \sqrt{x^2 + \varepsilon^2}$ represents the non-squared penalty function, in which $\varepsilon = 0.001$ is a constant. The symbol $J(|\nabla I|) = \gamma \cdot \exp(-\alpha\, |\nabla I|^{\beta})$ is a self-adaptive weight related to the image gradient. This weight decreases the flow diffusion near image edges to preserve the image and motion boundaries and increase the flow diffusion in smooth areas to

produce the global dense flows. We set the factors $\gamma = \alpha = \beta = 1$ by referring to [24].

In the global optimization process, we set the dense correspondence field as the initialize solution, and then adopt an inner and outer iteration scheme to minimize the energy function due to the implicit and nonlinear components contained in Eq. (11). To ensure both computational accuracy and efficiency, we set the inner iteration is 300 and the outer iteration is 3. After the global optimization, the final flow field is achieved.

V. EXPERIMENTAL RESULTS

A. EVALUATION DATASETS AND ERROR METRICS

For a comprehensive evaluation, we respectively run our method on MPI-Sintel [65] and UCF101 [66] datasets to test the performance of optical flow estimation.

The MPI-Sintel data includes the training and test datasets, which offer the non-rigid deformation, large displacements, occlusions and motion blurring for testing the performance of various optical flow methods. For a quantitative evaluation, we use the metrics of average angle error (AAE) and average endpoint error (AEPE) to indicate the performance of optical flow on training sets, and use the metrics of AEPE all, AEPE matched and AEPE unmatched from the MPI-Sintel online benchmark to evaluate the performance of optical flow estimation on test datasets. The metrics of AEPE all, AEPE matched and AEPE unmatched display the results of AEPE over complete frames, the regions that remain visible in adjacent frames and regions that are visible only in one of two adjacent frames, respectively. In addition, we use the metrics of ''s0-10'', ''s10-40'', ''s40+'', ''d0-10'', ''d10-60'', and ''d60+'' to make the specific evaluation of optical flow on large displacements and occlusions, in which the metrics of ''s0-10'', ''s10-40'' and ''s40+'' denote the results of AEPE over the regions with different velocities per frame, and the metrics of ''d0-10'', ''d10-60'' and ''d60+'' indicate the results of AEPE over regions close to occlusion boundaries with different distances, respectively.

Being very different with the MPI-Sintel data, the UCF101 data is composed of 101 action classes, which offers more than 13000 videos for evaluating various vision tasks. It is too large to run our method on all UCF101 datasets, we utilize eight datasets in the Sports classification to make a straightforward evaluation. The experimental datasets offer over 1100 frames for testing the performance of optical flow estimation under large or small displacements, motion occlusions and non-rigid motions. Due to the UCF101 datasets do not provide the ground truth of optical flow, we predict the next frame using the reference and the estimated optical flow, and then use the metrics of the peak signal to noise ratio (PSNR) and Sharpness (SN) to evaluate the performance of next frame prediction. Because the next frame is directly computed by using the estimated optical flow, a better performance on next frame prediction indicates a better result on optical flow estimation.

B. COMPARISON METHODS

To make a convincing comparison, we compare the flow result of our method with that of several state-of-the-art methods including Classic+NL [22], Deepflow [50], LDOF [47], JOF [27], STDC-Flow [54], PWC-Net [36], FlowNet2.0 [31] and IRR-PWC [39], in which the Classic+NL and JOF methods belonging to the variational optical flow approach, the Deepflow, LDOF and STDC- Flow methods belonging to the matching-based optical flow approach, and the PWC-Net, FlowNet2.0 and IRR-PWC methods are CNN-based optical flow approach.

In the comparison methods, the Classic+NL [22] method incorporates a non-local constraint term into a classical energy function, and then minimizes the objective function by applying a weighted median filtering to remove the outliers during the computation process. The JOF [27] method plans a joint filtering framework by combining the median filter and mutual structure guide filter, which performs a good result in term of robustness and edge-preserving. In this paper, the number of pyramid layers is fixed at 6, and the down-sampling factor is set as 0.5 in both of Classic+NL and JOF methods.

The DeepFlow [50] method is a representative matching-based optical flow approach, which utilizes a pyramid layering strategy to improve the accuracy of feature matching. In this paper, the pyramid down-sampling factor is set as 0.95. The LDOF [47] method incorporates a descriptor-based matching term into a variational energy function, which significantly improve the performance of optical flow estimation under large displacements. The STDC-Flow method [54] uses similarity transformation based dense correspondence to improve the accuracy and robustness of optical flow under large displacements and motion occlusions. In this paper, the down-sampling factor is set as 0.5 and the patch size is fixed to be 4×4.

The FlowNet2.0 [31] is constructed by stacking several FlowNetC and FlowNetS networks [30], which significantly improves the accuracy of optical flow estimation. We firstly train the FlowNet2.0 model on FlyingChairs datasets, and fine-tune it on FlyingThings3D, MPI-Sintel and KITTI training sets, respectively. The PWC-Net method [36] uses a feature pyramid-based network to predict flow field, which performs good performance on several public databases. In this paper, we apply a 7-level pyramid to the PWC-Net model and set a search range of 4 pixels to compute the cost volume at each level. We train the PWC-Net model on FlyingChairs datasets by using a long learning rate and use a batch size of 8 to fine-tune it on MPI-Sintel and KITTI training sets. Finally, the IRR-PWC method [39] incorporates an iterative residual refinement scheme into the PWC-Net framework, which improves the robustness of optical flow in regions of occlusions. In this paper, we use the same network parameters and training strategy with that of PWC-Net method to test the performance of IRR-PWC.

C. ABLATION EXPERIMENT

To reveal the impact of the various components of our method in improving the performance of optical flow estimation, we use the final sets of the MPI-Sintel training data to conduct an ablation experiment.

TABLE 1. Ablation experiment results.

Ablation models	AAE	AEPE
Full model	7.08	3.73
No match optimization	8.69	5.17
No edge optimization	8.24	4.41
No global optimization	7.51	3.89

Table 1 lists the results of AAE and AEPE of the proposed method with different modeling choices, in which the Full model represents the proposed Riscflow method, the No match optimization represents that removing the grid-based neighboring support optimization from the Riscflow method, the No edge optimization represents that replacing the edge-preserving interpolation framework by a classical interpolation scheme, the No global optimization represents that removing the global optimization strategy from the presented method. As can be seen in Table 1, the comparison results between different ablation models indicate that removing the grid-based neighboring support optimization, the edge-preserving interpolation or global optimization leads to a significant degradation in estimation accuracy of optical flow. Those comparison results demonstrate the benefit of the matching optimization, edge-preserving interpolation and global optimization for optical flow estimation.

For a visual comparison, Fig. 6 displays the estimated flow fields of the various ablation models on some MPI-Sintel training sets including alley_2, cave_4 and market_6, in which the alley_2 and cave_4 sequences contain the non-rigid deformation, and the market_6 sequence includes the large displacements. As shown in Fig. 6, removing the matching optimization produces the obvious errors in the estimated flow field because the initial correspondence field may include a mass of incorrect matching results. Moreover, removing the edge-preserving interpolation or global optimization leads to the issues of edge-blurring or over-segmentation, because the proposed edge-preserving interpolation scheme and global optimization strategy is able to improve the accuracy of optical flow near motion boundaries. The visual comparison demonstrates that the proposed grid-based neighboring support optimization, edge-preserving interpolation and global optimization are able to cope with the challenge of non-rigid deformation and large displacements.

D. COMPARISON RESULTS FROM MPI-SINTEL TEST DATASETS

In this subsection, we run the proposed Riscflow method on the MPI-Sintel test datasets to conduct a comprehensive comparison with some state-of-the-art methods. Table 2 lists the comparison results of various methods evaluated on the MPI-Sintel test datasets.

TABLE 2. Comparison results of various methods on MPI-Sintel test datasets.

Method	AEPE all	AEPE matched	AEPE unmatched
Classic+NL	7.96	3.77	42.08
JOF	6.92	3.08	38.20
PWC-Net	4.39	1.72	26.17
IRR-PWC	3.84	1.47	23.22
FlowNet2.0	3.96	1.47	24.29
Deepflow	5.38	1.77	34.75
LDOF	7.56	3.43	41.17
STDC-Flow	6.99	3.24	37.56
Riscflow	4.43	1.34	29.64

As can be seen in Table 2, the Classic+NL and JOF methods perform the inferior performance compared with the other methods, because the non-rigid deformation, large displacements and occlusions contained in the MPI-Sintel datasets are the significant challenge for the variational optical flow approaches. The Deepflow, LDOF and STDC-Flow methods achieve the mediocre performance among the evaluation methods, due to the traditional matching-based methods are susceptible to non-rigid motion and large displacements. Although the PWC-Net, FlowNet2.0 and IRR-PWC methods perform the best results on metrics of AEPE all and AEPE unmatched, these CNN-based methods require numerous labeled datasets to train the networks. This specific limitation may prevent the CNN-based optical flow methods from applying in real-world scene. Despite that the proposed Riscflow method results in a slightly backward performance on metrics of AEPE all and AEPE unmatched compared with the CNN-based methods, it achieves the best result on the metric of AEPE matched and outperforms the other variational and matching-based methods on the total metrics. Because the metrics of AEPE all, AEPE matched and AEPE unmatched indicate the performance of optical flow over complete frames, the non-occluded regions and occluded regions, respectively. The comparison results in Table 2 demonstrates that the proposed method performs the better performance in the entire regions compared with the other variational and matching-based methods, and it achieves a better performance in the non-occluded regions compared with the evaluated CNN-based methods.

To make a specific comparison on large displacements and edge-preserving, Table 3 summarizes the comparison results of the various evaluation methods on some specific metrics of MPI-Sintel test datasets. In Table 3, the metrics of ''s0-10'', ''s10-40'', ''s40+'', ''d0-10'', ''d10-60'', and ''d60+'' are employed to make the straightforward evaluation of optical flow on large displacements and motion boundaries, in which the metrics of ''s0-10'', ''s10-40'' and ''s40+'' denote the results of AEPE over the regions with different velocities per frame, and the metrics of ''d0-10'', ''d10-60'' and ''d60+''

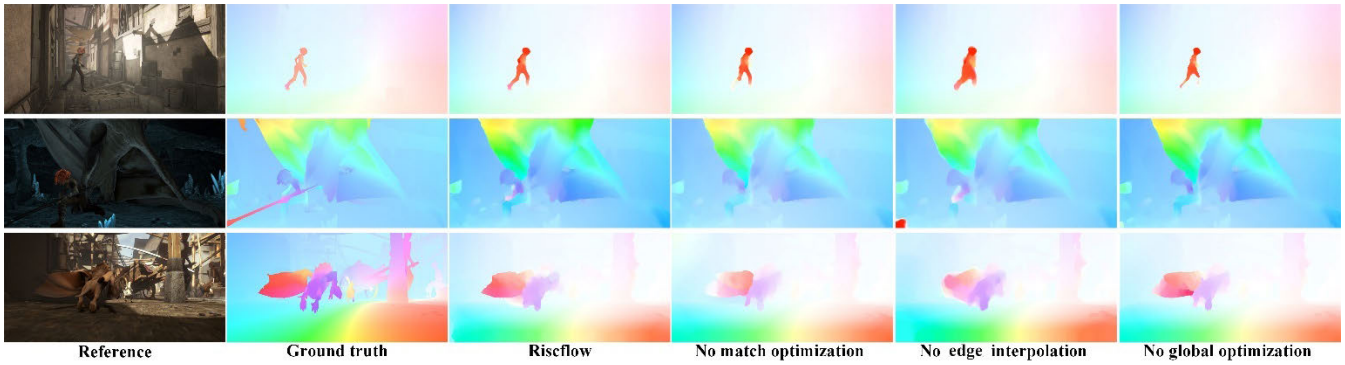


FIGURE 6. Flow fields of the various ablation models on some MPI-Sintel training sets. From top to bottom: sequences of alley_2, cave_4 and market_6.

TABLE 3. Comparison results of the various methods on some specific metrics of MPI-Sintel test datasets.

Method	d0-10	d10-60	d60-140	s0-10	s10-40	s40+
Classic+NL	6.19	3.91	2.51	0.57	2.69	57.37
JOF	5.98	3.42	1.68	0.58	2.47	49.14
PWC-Net	4.28	1.66	0.67	0.61	2.07	28.79
IRR-PWC	3.50	1.30	0.72	0.54	1.72	25.43
FlowNet2.0	3.09	1.32	0.92	0.64	1.90	25.42
Deepflow	4.52	1.53	0.84	0.96	2.73	33.70
LDOF	5.35	3.28	2.45	0.94	2.91	51.67
STDC-Flow	6.23	3.12	1.74	0.67	2.87	48.25
Riscflow	3.44	1.15	0.67	0.58	2.13	29.13

indicate the results of AEPE over regions close to boundaries with different distances, respectively.

As seen in Table 3, the Classic+NL and JOF methods achieve the best and second-best results on metric of s0-10. This indicates that these two variational methods perform a good performance on small displacements. However, they result in the poor performance on the other metrics, which indicate that the Classic+NL and JOF methods are incapable of dealing with large displacements and non-rigid deformation. Despite that the Deepflow, STDC-Flow and LDOF methods perform a slightly better performance than the Classic+NL method on metric of s0-10, they still result in a poor performance on large displacements. The FlowNet2.0 and IRR-PWC methods achieve the best and second-best results on metric of s40+, which demonstrate the CNN-based methods are able to cope with large displacements and non-rigid deformation. The proposed Riscflow method performs the best result on metrics of d10-60 and d60-140, the second-best results on metrics of d0-10 and s0-10, and the competitive results on metrics of s10-40 and s40+, respectively. This indicates that the proposed method performs the good performance in regions of large displacements and motion boundaries, which demonstrates that the proposed method has the significant benefit of edge-preserving under non-rigid motion and large displacements.

For a visual comparison, Fig. 7 displays the estimated flow fields of the proposed Riscflow and other state-of-the-art methods tested on some MPI-Sintel test dataset. As shown in the figure, the Classic+NL and JOF methods produce good results on the datasets including small displacements, such as Market_3 sequence. However, they generate obvious over- segmentation on some datasets containing large displacements and occlusions, such as Cave_3, Market_1 and Market_4 sequences. The Deepflow, LDOF and STDC-flow methods produce edge-blurring on Shaman_1 and Market_4 sequences which including complex edges and motion occlusions. The FlowNet2.0, PWC-Net and IRR-PWC methods result in good performance on most of the test datasets. However, their flow fields include some errors near the image and motion boundaries. The proposed Riscflow method achieves a competitive performance compared with the FlowNet2.0 and PWC-Net methods, and especially gains the better results in regions of large displacements and motion boundaries.

E. COMPARISON RESULTS FROM UCF101 DATASETS

To evaluate the performance of the proposed Riscflow and the other state-of-the-art methods on the real-world data, we utilize the UCF101 datasets to conduct a comprehensive experiment. Fig. 8 illustrates the estimated flow fields of the proposed Riscflow and other state-of-the-art methods tested on some UCF101 datasets.

As shown in Fig. 8, although the FlowNet2.0, PWC-Net and IRR-PWC methods perform the excellent performance on MPI-Sintel datasets, they generate obvious errors in the estimated flow field of the UCF101 datasets because the UCF101 data does not offer the optical flow ground truth for training their networks. The Classic+NL, JOF, LDOF, STDC-Flow and Deepflow methods perform well on the UCF101 datasets. However their flow fields appear the issues of over-segmentation or edge-blurring. The proposed Riscflow method achieves the good results on the UCF101 datasets, because the motion boundaries in the estimated flow field are well preserved.

Due to the UCF101 datasets do not provide the ground truth of optical flow, we predict the next frame using the reference and the estimated optical flow, and then use the metrics of the peak signal to noise ratio (PSNR) and Sharpness (SN) to evaluate the performance of next frame

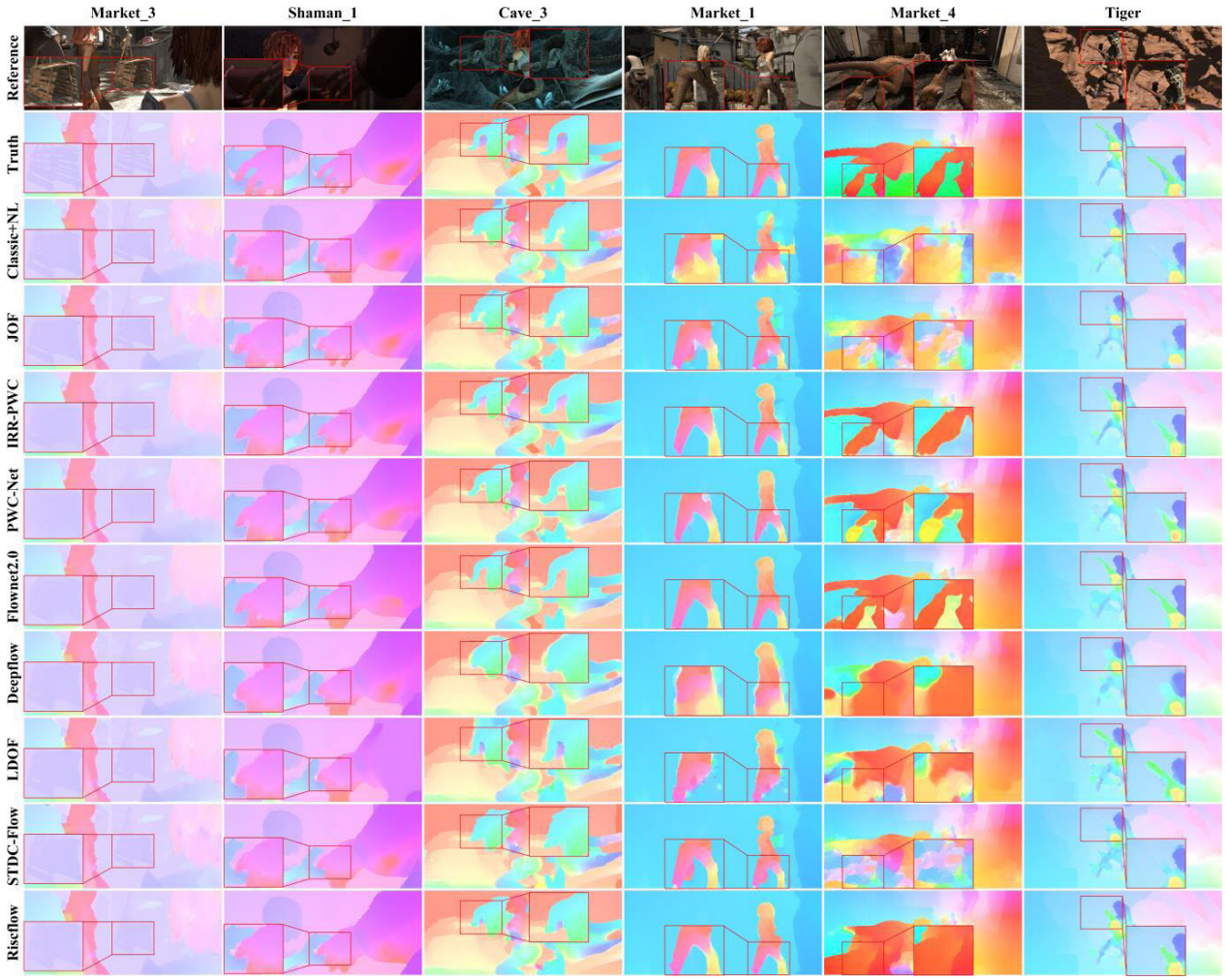


FIGURE 7. Estimated flow fields of the proposed Riscflow and other state-of-the-art methods tested on some MPI-Sintel test dataset.

prediction. Specifically, the PSNR evaluates the level of similarity, and the SN measures the loss of sharpness between the truth frame and the predicted frame. Because the next frame is directly computed by using the estimated optical flow, a better performance on next frame prediction indicates a better result on optical flow estimation. Table 4 summarizes the comparison results of the next frame prediction between the Riscflow and other evaluation methods. As shown in Table 4, the proposed Riscflow method performs the best and second-best results on metrics of SN and PSNR, respectively. These comparison results indicate that the proposed method achieves the good performance in terms of accuracy and robustness, especially owns the benefit of edge-preserving.

F. RUNTIMES

To make a comprehensive comparison between the proposed Riscflow and the other state-of-the-art methods, Table 5 lists the comparison results of the average runtimes of various methods tested on MPI-Sintel and UCF101 datasets, respectively. Specifically, the proposed method and the other variational and matching-based approaches are implemented by MATLAB2010 using a Lenovo computer equipped with an Intel Core I7-6700K CPU. The CNN-based approaches are implemented by a Lenovo computer equipped with an Intel Core I7-6700K CPU and an NVIDIA GTX 1080Ti GPU.

As seen in Table 5, the Classic+NL, STDC-Flow and JOF methods cost the largest time consumption. This is because these methods utilize a coarse-to-fine iteration scheme to improve the accuracy of optical flow, which may significantly increase the time cost. Despite that the Deepflow and LDOF methods cost less runtimes compared with the variational methods, these two matching-based methods perform the poor performance on estimation accuracy. The PWC-Net, FlowNet2.0 and IRR-PWC methods achieve the best results in runtimes because the CNN-based methods have the significant benefit of real-time computation. However, these CNN-based methods require supervised training process and cannot be applied to real-world data where the ground truth is not easily accessible. The proposed Riscflow method costs more time than the other matching-based methods because of the additional optimization schemes including the grid-based neighboring support optimization, edge-preserving interpolation and global optimization.

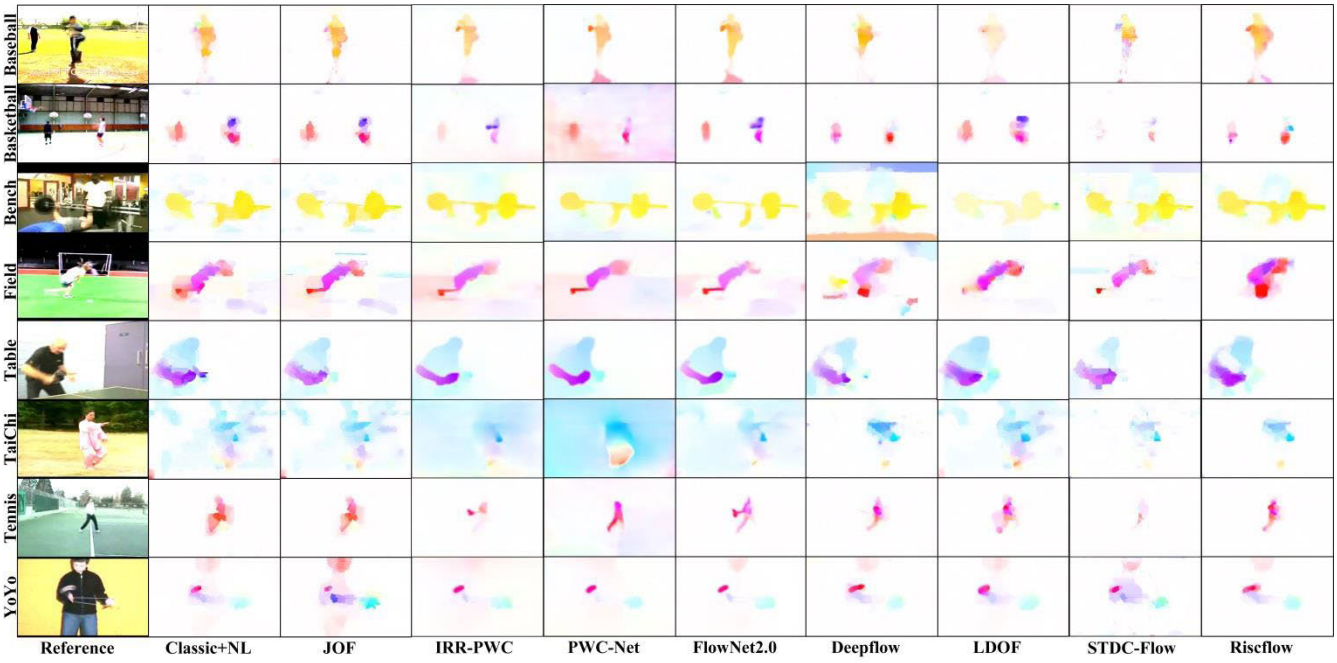


FIGURE 8. The estimated flow fields of the proposed Riscflow and other state-of-the-art methods tested on some UCF101 datasets.

TABLE 4. Comparison results of various methods tested on UCF101 datasets.

Datasets	Classic+NL	JOF	PWC-Net	IRR-PWC	FlowNet2.0	Deepflow	LDOF	STDC-flow	Riscflow
	PSNR/SN	PSNR/SN	PSNR/SN	PSNR/SN	PSNR/SN	PSNR/SN	PSNR/SN	PSNR/SN	PSNR/SN
Baseball	35.05/22.94	35.11/22.96	29.11/18.45	30.96/18.98	31.94/20.48	30.51/22.92	37.41/22.89	38.84 / 27.35	39.16/28.00
Basketball	42.33/27.77	42.49/27.85	36.95/22.96	38.18/23.66	40.58/26.40	40.76/27.79	42.17/27.80	42.64 / 28.62	43.21/28.89
Bench	21.29/17.03	21.34/17.06	21.10/16.70	21.08/16.74	20.82/16.69	22.12/17.26	20.77/16.99	21.67 / 17.11	21.07/17.14
Field	29.86/19.77	29.92/19.81	28.35/19.04	27.42/18.82	28.93/19.29	27.22/18.89	30.08/19.90	29.51 / 19.83	29.65/19.91
Table	23.80/20.48	23.86/20.51	23.68/19.39	23.76/19.51	22.86/19.88	25.02/20.61	23.39/20.41	24.6 / 20.6	24.37/20.68
TaiChi	31.62/20.83	31.68/20.85	31.10/20.37	31.10/20.39	30.46/20.30	33.06/21.60	32.06/20.94	32.96 / 21.46	33.03/21.58
Tennis	35.65/25.97	35.72/26.01	33.09/23.10	32.43/22.71	33.89/24.79	33.45/25.29	35.04/25.89	36.01 / 26.25	36.13/26.53
YoYo	25.29/20.49	25.42/20.56	25.02/19.76	24.32/19.17	24.58/20.07	26.78/20.93	24.89/20.53	25.9 / 20.74	25.43/20.76
Average	30.61/21.91	30.69/21.95	28.55/19.97	28.65/20.00	29.26/20.99	29.87/21.91	30.39/21.92	31.51/22.75	31.50/22.93

TABLE 5. Comparison results of time consumption (Unit: Second).

Methods	MPI-Sintel	UCF101
Classic+NL	565	72
JOF	1402	227
PWC-Net	0.08	0.05
IRR-PWC	0.18	0.06
FlowNet2.0	0.1	0.1
Deepflow	19.0	2.4
LDOF	39.0	3.0
STDC-Flow	813	131
Riscflow	88.2	23.5

Despite that these additional optimizations slightly increase the time consumption, they prompt the proposed method to achieve the good performance on accuracy and robustness of optical flow estimation.

VI. CONCLUSION

In this paper, we proposed a robust-interpolation-based optical flow estimation method to cope with the issue of large displacements and non-rigid deformation. First, we utilized the deep matching model to gain an initial sparse correspondence field between the consecutive two frames, and then exploited a grid-based neighboring support optimization scheme to optimize the sparse correspondence field. Second, we constructed an edge-preserving spare-to- dense interpolation framework to preserve the motion boundaries. Third, we adopted a global energy function to optimize the dense correspondence field to achieve the dense flow field. Fianlly, we respectively ran our method on MPI-Sintel and UCF101 datasets to conduct a comprehensive comparison with several state-of-the-art methods. The experimental results demonstrated that the proposed method has high accuracy and good robustness of optical flow estimation, and especially owns the benefit

of edge-preserving under large displacements and non-rigid deformation.

The proposed Riscflow method costs more time than some comparison methods, especially the deep-learning approaches. One reason is that we utilize a post-processing global optimization scheme to improve the accuracy of optical flow estimation, which may significantly increase the operation time. Another reason is that we simply use a serial framework to implement the proposed method on a CPU processor. In the future work, we will modify the method by using a parallel framework and try to use a GPU to accelerate the computation process. This may promote the proposed method to achieve the real-time implementation.

REFERENCES

[1] Q. Xie, O. Remil, Y. Guo, M. Wang, M. Wei, and J. Wang, "Object detection and tracking under occlusion for object-level RGB-D video segmentation," *IEEE Trans. Multimedia*, vol. 20, no. 3, pp. 580–592, Mar. 2018.

[2] R. V. H. M. Colque, C. Caetano, M. T. L. de Andrade, and W. R. Schwartz, "Histograms of optical flow orientation and magnitude and entropy to detect anomalous events in videos," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 673–682, Mar. 2017.

[3] R. Ke, Z. Li, J. Tang, Z. Pan, and Y. Wang, "Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 1, pp. 54–64, Jan. 2019.

[4] W. Cao, Y. Li, and Z. He, "Weighted optical flow prediction and attention model for object tracking," *IEEE Access*, vol. 7, pp. 144885–144894, 2019.

[5] X. Wang, Z. He, R. Sun, L. You, J. Hu, and J. Zhang, "A crowd behavior identification method combining the streakline with the high-accurate variational optical flow model," *IEEE Access*, vol. 7, pp. 114572–114581, 2019.

[6] G. Wu and W. Kang, "Vision-based fingertip tracking utilizing curvature points clustering and hash model representation," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1730–1741, Aug. 2017.

[7] H. Hu and P. Chen, "Direct optical-flow-aware computational framework for 3D reconstruction," *IEEE Access*, vol. 7, pp. 169518–169527, 2019.

[8] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[9] Z. Tu, W. Xie, D. Zhang, R. Poppe, R. C. Veltkamp, B. Li, and J. Yuan, "A survey of variational and CNN-based optical flow techniques," *Signal Process., Image Commun.*, vol. 72, pp. 9–24, Mar. 2019.

[10] T. Brox, N. Papenberg, and J. Weickert, "High accuracy optic flow estimation based on a theory for warping," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 25–36.

[11] N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert, "Highly accurate optic flow computation with theoretically justified warping," *Int. J. Comput. Vis.*, vol. 67, no. 2, pp. 141–158, 2006.

[12] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade meets horn/schunck: Combining local and global optic flow methods," *Int. J. Comput. Vis.*, vol. 61, no. 3, pp. 211–231, 2005.

[13] M. A. Mohamed, H. A. Rashwan, B. Mertsching, M. A. Garcia, and D. Puig, "Illumination-robust optical flow using a local directional pattern," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 9, pp. 1499–1508, Sep. 2014.

[14] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Comput. Vis. Image Understand.*, vol. 63, no. 1, pp. 75–104, Jan. 1996.

[15] J. Hur and S. Roth, "Joint optical flow and temporally consistent semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 163–177.

[16] A. Wedel, T. Pock, C. Zach, D. Cremers, and H. Bischof, "An improved algorithm for TV-L^1 optical flow," in *Proc. Dagstuhl Motion Workshop*, 2008, pp. 23–45.

[17] N. Monzon, A. Salgado, and J. Sanchez, "Regularization strategies for discontinuity-preserving optical flow methods," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1580–1591, Apr. 2016.

[18] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2432–2439.

[19] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic huber-L1 optical flow," in *Proc. Brit. Mach. Vis. Conf.*, 2009, pp. 108.1–108.11.

[20] H. Zimmer, A. Bruhn, and J. Weickert, "Optic flow in harmony," *Int. J. Comput. Vis.*, vol. 93, no. 3, pp. 368–388, Jul. 2011.

[21] T. Amiaz, E. Lubetzky, and N. Kiryati, "Coarse to over-fine optical flow estimation," *Pattern Recognit.*, vol. 40, no. 9, pp. 2496–2503, Sep. 2007.

[22] D. Sun, S. Roth, and M. J. Black, "A quantitative analysis of current practices in optical flow estimation and the principles behind them," *Int. J. Comput. Vis.*, vol. 106, no. 2, pp. 115–137, Jan. 2014.

[23] A. Ayvaci, M. Raptis, and S. Soatto, "Sparse occlusion detection with optical flow," *Int. J. Comput. Vis.*, vol. 97, no. 3, pp. 322–338, May 2012.

[24] C. X. Zhang, Z. Chen, M. R. Wang, M. Li, and S. F. Jiang, "Robust non-local TV-L^1 optical flow estimation with occlusion detection," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4055–4067, Aug. 2017.

[25] Z. Tu, R. Poppe, and R. C. Veltkamp, "Adaptive guided image filter for warping in variational optical flow computation," *Signal Process.*, vol. 127, pp. 253–265, Oct. 2016.

[26] C. Zhang, L. Ge, Z. Chen, R. Qin, M. Li, and W. Liu, "Guided filtering: Toward edge-preserving for optical flow," *IEEE Access*, vol. 6, pp. 26958–26970, 2018.

[27] C. Zhang, L. Ge, Z. Chen, M. Li, W. Liu, and H. Chen, "Refined TV-L^1 optical flow estimation using joint filtering," *IEEE Trans. Multimedia*, vol. 22, no. 2, pp. 349–364, Feb. 2020.

[28] D.-H. Trinh and C. Daul, "On illumination-invariant variational optical flow for weakly textured scenes," *Comput. Vis. Image Understand.*, vol. 179, pp. 1–18, Feb. 2019.

[29] B. Zhu, L.-F. Tian, Q.-L. Du, Q.-X. Wu, F. Z. Sahl, and Y. Yeboah, "Adaptive dual fractional-order variational optical flow model for motion estimation," *IET Comput. Vis.*, vol. 13, no. 3, pp. 277–284, Apr. 2019.

[30] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2758–2766.

[31] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "FlowNet 2.0: Evolution of optical flow estimation with deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1647–1655.

[32] T.-W. Hui, X. Tang, and C. C. Loy, "LiteFlowNet: A lightweight convolutional neural network for optical flow estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8981–8989.

[33] T.-W. Hui, X. Tang, and C. C. Loy, "A lightweight optical flow CNN—Revisiting data fidelity and regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Feb. 28, 2020, doi: 10.1109/TPAMI.2020.2976928.

[34] T. W. Hui, X. Tang, and C. C. Loy, "LiteFlowNet3: Resolving correspondence ambiguity for more accurate optical flow estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 169–184.

[35] C. Bailer, K. Varanasi, and D. Stricker, "CNN-based patch matching for optical flow with thresholded hinge embedding loss," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2710–2719.

[36] D. Sun, X. Yang, M.-Y. Liu, and J. Kautz, "PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8934–8943.

[37] P. Hu, G. Wang, and Y. P. Tan, "Recurrent spatial pyramid CNN for optical flow estimation," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2814–2823, Mar. 2018.

[38] A. Ranjan and M. J. Black, "Optical flow estimation using a spatial pyramid network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2720–2729.

[39] J. Hur and S. Roth, "Iterative residual refinement for joint optical flow and occlusion estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5747–5756.

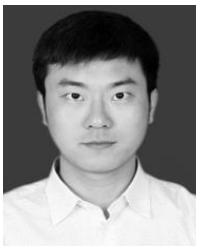
[40] S. Zhao, Y. Sheng, Y. Dong, E. I.-C. Chang, and Y. Xu, "MaskFlownet: Asymmetric feature matching with learnable occlusion mask," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6277–6286.

[41] Z. Teed and J. Deng, "RAFT: Recurrent all-pairs field transforms for optical flow," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 402–419.

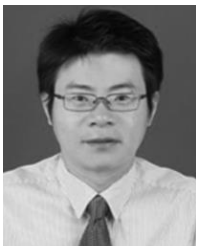
[42] A. Ahmadi and I. Patras, "Unsupervised convolutional neural networks for motion estimation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 1629–1633.
[43] Y. Wang, Y. Yang, Z. Yang, L. Zhao, P. Wang, and W. Xu, "Occlusion aware unsupervised learning of optical flow," in *Proc. CVPR*, Jun. 2018, pp. 4884–4893.
[44] J. Li, J. Zhao, T. Feng, C. Ye, and L. Xiong, "Occlusion aware unsupervised learning of optical flow from video," Mar. 2020, *arXiv:2003.01960*. [Online]. Available: https://arxiv.org/abs/2003.01960
[45] J. Cheng, Y.-H. Tsai, S. Wang, and M.-H. Yang, "SegFlow: Joint learning for video object segmentation and optical flow," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 686–695.
[46] W. S. Lai, J. B. Huang, and M. H. Yang, "Semi-supervised learning for optical flow with generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 354–364.
[47] T. Brox and J. Malik, "Large displacement optical flow: Descriptor matching in variational motion estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 500–513, Mar. 2011.
[48] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1744–1757, Sep. 2012.
[49] C. Barnes, E. Shechtman, D. Goldman, and A. Finkelstein, "The generalized patchmatch correspondence algorithm," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 29–43.
[50] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1385–1392.
[51] V. Lempitsky, S. Roth, and C. Rother, "FusionFlow: Discrete-continuous optimization for optical flow estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
[52] Y. Hu, R. Song, and Y. Li, "Efficient coarse-to-fine patch match for large displacement optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5704–5712.
[53] Y. Zu, W. Tang, X. Bao, Y. Wang, and K. Gao, "Context-adaptive matching for optical flow," *Multimedia Tools Appl.*, vol. 78, no. 1, pp. 641–659, Jan. 2019.
[54] C. Zhang, Z. Chen, F. Xiong, W. Liu, M. Li, and L. Ge, "STDC-flow: Large displacement flow field estimation using similarity transformation-based dense correspondence," *IET Comput. Vis.*, vol. 14, no. 5, pp. 248–258, Aug. 2020.
[55] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "EpicFlow: Edge-preserving interpolation of correspondences for optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1164–1172.
[56] Y. Hu, Y. Li, and R. Song, "Robust interpolation of correspondences for large displacement optical flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 481–489.
[57] Y. Li, "Pyramidal gradient matching for optical flow estimation," 2017, *arXiv:1704.03217*. [Online]. Available: http://arxiv.org/abs/1704.03217
[58] J. Chen, Z. Cai, J. Lai, and X. Xie, "Efficient segmentation-based PatchMatch for large displacement optical flow estimation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 12, pp. 3595–3607, Dec. 2019.
[59] F. Yang, Y. Cheng, J. V. D. Weijer, and M. G. Mozerov, "Improved discrete optical flow estimation with triple image matching cost," *IEEE Access*, vol. 8, pp. 17093–17102, 2020.
[60] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "DeepMatching: Hierarchical deformable dense matching," *Int. J. Comput. Vis.*, vol. 120, no. 3, pp. 300–323, Dec. 2016.
[61] J. Bian, W. Lin, Y. Matsushita, S. Yeung, T. Nguyen, and M. Cheng, "GMS: Grid-based motion statistics for fast, ultra-robust feature correspondence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4181–4190.
[62] P. Dollár and C. L. Zitnick, "Structured forests for fast edge detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 1841–1848.
[63] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.
[64] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "Learning to detect motion boundaries," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2578–2586.
[65] J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 611–625.
[66] K. Soomro, R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," Univ. Central Florida, Orlando, FL, USA, Tech. Rep. CRCV-TR-12-01, 2012.

SHIDONG SHI received the bachelor's degree in biomedical engineering from Nanchang Hangkong University, China, in 2018, where he is currently pursuing the master's degree in instrumentation engineering. His current research interests include image processing and computer vision.

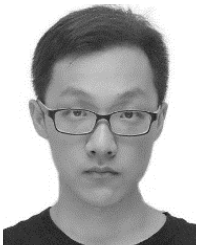
DAOWEN ZHANG received the bachelor's degree in pharmaceutical engineering from Northeast Agricultural University, in 2017, and the master's degree in instrumentation engineering from Nanchang Hangkong University, in 2020. His current research interests include image processing and computer vision.



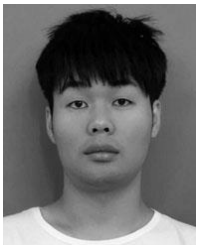
CONGXUAN ZHANG (Member, IEEE) received the Ph.D. degree in measurement technology and instruments from the Nanjing University of Aeronautics and Astronautics, Nanjing, in 2014. From 2018 to 2019, he was a Visiting Scholar with the Department of Biomedical Engineering, University of Kansas. He is currently an Assistant Professor with the School of Measuring and Optical Engineering, Nanchang Hangkong University, China. His current research interests include image processing and computer vision.



ZHEN CHEN received the Ph.D. degree in mechanical design and theory from Northwestern Polytechnical University, Xi'an, in 2003. From 2006 to 2007, he was a Visiting Scholar with the Department of Biomedical Engineering, University of Kansas. He is currently a Professor with the School of Measuring and Optical Engineering, Nanchang Hangkong University, China. His current research interests include image understanding and measurement.



CHENG FENG received the bachelor's degree in automation from the Wuchang University of Technology, in 2016. He is currently pursuing the master's degree in instrumentation engineering with Nanchang Hangkong University, China. His current research interests include image processing and computer vision.



BINGBING FAN received the bachelor's degree in biomedical engineering from Nanchang Hangkong University, China, in 2019, where he is currently pursuing the master's degree in instrumentation engineering. His current research interests include image processing and computer vision.