

Received 17 April 2024, accepted 10 May 2024, date of publication 14 May 2024, date of current version 6 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3400884

RESEARCH ARTICLE

A Non-Rigid Three-Dimensional Image Reconstruction Algorithm Based on Deformable Shape Reliability

HAIYING CHEN¹ AND SYED ATIF MOQURRAB²

¹Department of Public Courses, Tianmen Vocational College, Tianmen 431700, China

²School of Computing, Gachon University, Sujeong-gu, Seongnam-si 13120, Republic of Korea

Corresponding author: Syed Atif Moqurrab (atif@gachon.ac.kr)

This work was supported by the New Generation Information Technology Innovation Project 2021 of the Science Research and Development Center for Higher Education Institutions of the Ministry of Education, China, under Grant 2021ITA02003.

ABSTRACT Most reconstruction algorithms for non-rigid three-dimensional (3D) images assume that non-rigidity can be represented as a linear combination of a fixed number of rigid bases. However, this assumption struggles to establish reliable shape functions and initial values for nonlinear and non-rigid motions, decreasing reconstruction reliability. This paper introduces an enhanced-reliability reconstruction algorithm for non-rigid 3D images. Our algorithm models the dynamic non-rigid shape basis as a low-rank matrix composed of image points and depth factors, improving the restoration of non-rigid shape base changes and providing accurate parameters for constructing objective functions. By leveraging manifold alignment and physical continuity constraints, our method optimizes the function structures. Assuming minimal reconstruction error and shape change, we solve for the motion structure parameters and select the key initial shape basis value by minimizing the objective function with the L-M nonlinear optimization method. Our experimental results on 3D image sequence reconstructions demonstrate significant error reduction, underscoring our model's credibility, robust reliability, and minimal re-projection error.

INDEX TERMS Non-rigid, 3D image reconstruction, shape reliability, low-rank matrix, improved objective function.

I. INTRODUCTION

The analysis and understanding of visual information by computers facilitate a more accurate simulation of the real world, giving rise to the field of computer vision [1]. This discipline primarily focuses on acquiring various real-world details through the capture of 2D images, encompassing shape and motion recognition of 3D scenes [2]. A prominent research area within this field is the recovery of motion scenes and their corresponding parameters from continuous image sequences, commonly referred to as 3D motion vision. While significant advances have been made in the reconstruction of rigid bodies, exploring non-rigid body reconstructions continues to present considerable challenges [3]. Non-rigid 3D image reconstruction seeks to recover a 3D model of

non-rigid objects from 2D images captured from multiple viewpoints, employing image processing and computer vision techniques [4]. This technology has wide-ranging applications in movie production, game development, and industrial design [5], [6]. However, the complexity and diversity of non-rigid motion render non-rigid 3D image reconstruction a challenging task, with ensuring the reliability of shape bases computation standing out as a particular concern.

In the realm of 3D image reconstruction, Greffier et al. employed deep learning algorithms to reconstruct original images using methods such as Filtered Back Projection (FBP), enhanced AIDR 3D (AIDR 3De), and AiCE at three levels (mild, standard, and strong) [7]. This approach, which relies on linear assumptions for reconstruction, falls short in adequately addressing the impacts of nonlinear deformations, thus proving ineffective in establishing stable

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Steven Li¹.

and reliable shape objective functions for nonlinear, non-rigid motion. On the other hand, Matthieu et al. suggested training deep neural networks (DNNs) as denoisers to learn a priori image models, which would replace hand-crafted proximal regularization operators in optimization algorithms. Their AIRI framework, aimed at imaging complex intensity structures from visibility data, merges the robustness and interpretability of optimization with the efficiency and learning capability of neural networks [8]. Nevertheless, this method encounters difficulties in practical applications, particularly in selecting suitable initial values for nonlinear non-rigid motion, resulting in low reliability of reconstruction results.

Lin et al. approached the challenge by converting optimal spatial deformation into a nonlinear regularized variational optimization problem, incorporating local smoothing and input constraints. They leveraged data parallelism and flash memory optimization strategies for online tracking and reconstruction of non-rigid scenes [9]. Despite these efforts, the method struggled to effectively manage nonlinear non-rigid motion during the research phase, impacting the establishment of reliable shape objective functions and leading to suboptimal reconstruction outcomes. Murase's study utilized a helical digital body phantom to generate degraded projection data, employing system function graphs and Gaussian noise. The entire system matrix (SM) was calculated by linking each projection data set with slices for 3D image reconstruction [10]. However, this approach did not adequately capture the characteristics and variations of non-rigid motion, resulting in unreliable initial values for non-linear non-rigid motion and diminished reconstruction reliability.

In Jo's research, attention modules were introduced, and simple algebraic pre-smoothing techniques like Gaussian filtering were applied to data. This pre-smoothed data served to derive an operator for image reconstruction through dynamic mode decomposition [11]. Although this method facilitated camera calibration with rigid parts and subsequent application to non-rigid reconstruction, it fell short in establishing stable and reliable shape objective functions and initial values for all non-rigid motions, posing challenges for generalization in practical applications.

II. DESIGN OF NON-RIGID 3D IMAGE RECONSTRUCTION ALGORITHM

A. THE CALCULATION OF LOW-RANK MATRICES FOR DYNAMIC SHAPE BASES IN NON-RIGID 3D IMAGE RECONSTRUCTION

Non-rigid body shapes undergo specific transformations at different time points. To effectively complete non-rigid 3D image reconstruction, it is essential to accurately describe these dynamic shape changes. In this context, we utilize a low-rank matrix, combined with image points and depth factors, to represent the data structure of basis parameters in dynamic shape for non-rigid 3D image reconstruction. This approach enables the capture of both local and global features of the shape, thereby better restoring the original non-rigid

shape basis's change process and providing precise manifold parameter values for the structure of the objective function.

The process for calculating the low-rank matrix for the dynamic shape basis in non-rigid 3D image reconstruction involves several steps:

Digitalization of Non-Rigid 3D Images: Utilize the coordinates and depth factors of all image points to digitize the non-rigid 3D image.

Construction of the Point Cloud Matrix: Based on the 3D image, construct a point cloud matrix of the shape for a specific frame.

Singular Value Decomposition (SVD): Use singular value decomposition to arrange the constructed matrix in descending order of importance. Then, select the first k th singular values to retain the low-rank matrix for describing the dynamic shape basis in non-rigid 3D image reconstruction.

It is assumed that a classic pinhole camera model facilitates the non-rigid 3D image acquisition process [12], [13], allowing the non-rigid 3D image to be described as follows:

$$\lambda \mathbf{U} = \mathbf{K}(\mathbf{R}\mathbf{t}) \mathbf{w} = \mathbf{P}\mathbf{w} \quad (1)$$

where λ denotes the depth factor, $\mathbf{U} = (u \ v \ 1)^T$ is the homogeneous coordinates of the image point, \mathbf{K} , \mathbf{R} and \mathbf{t} denote the camera's intrinsic matrix, rotation matrix and translation vector corresponding to the camera's shooting position, respectively, and \mathbf{R} and \mathbf{t} also perform the camera's extrinsic matrix, $\mathbf{w} = (x \ y \ z \ 1)^T$ and $\mathbf{P} = \mathbf{K}(\mathbf{R}\mathbf{t})$ denote homogeneous coordinates of 3D points and camera projection matrix.

With total F frames and N 3D spatial points, the equation is transformed as follows for the f^{th} frame:

$$\mathbf{U}_f \lambda_f = \mathbf{K}(\mathbf{R}_f \mathbf{t}_f) \mathbf{Y}_f \quad (2)$$

where $\mathbf{U}_f = (u_{f,1}, u_{f,2}, \dots, u_{f,N})$ is the image matrix consists of the points from the f -th image, λ_f represents the diagonal matrix composed of all depth factors $\lambda_{f,i}$, $\mathbf{Y}_f = (w_{f,1}, w_{f,2}, \dots, w_{f,N})$ denotes a low-rank matrix composed of all 3D spatial points at time f .

Finally, singular value decomposition is performed on the shape point cloud matrix of the constructed f^{th} frame image. The singular values are arranged in descending order, and the first k^{th} singular values are retained. The low rank matrix of the dynamic shape basis for non-rigid 3D image reconstruction can be obtained as:

$$\mathbf{Y}_f = \frac{\mathbf{I}_k \mathbf{u}_{f,k} \lambda_{f,k}}{\mathbf{K}(\mathbf{R}_{f,k} \mathbf{t}_{f,k})} \quad (3)$$

where, \mathbf{I}_k represents an orthogonal matrix.

Thus, the low rank matrix calculation of dynamic shape basis for non-rigid three-dimensional image imaging is completed. When an object undergoes rigid motion, the 3D spatial point \mathbf{Y}_f remains unchanged throughout the motion. Otherwise, \mathbf{Y}_f changes at each time instance f , representing a dynamic process of shape base variation.

B. DESIGN OF A STRONGLY RELIABLE OBJECTIVE FUNCTION FOR NON-RIGID IMAGE RECONSTRUCTION

1) THE PROPOSED MODEL

The enhancement of reliability in our current method fundamentally addresses a nonlinear optimization challenge for dynamic shape bases. This involves identifying the minimum value within a specific parameter structure. During the training of parameters for non-rigid motion structures, the chosen objective function aims to minimize the discrepancy between the 3D coordinates of the manifold and the transformed 2D coordinates of the manifold group. This is based on assumed motion structure parameters within the photography model, ensuring that the derived motion structure parameters align closely with the actual observed image points and their theoretically predicted 3D shapes. By reducing this discrepancy, we enhance both the quality and accuracy of the reconstruction results.

Thus, the developed objective function effectively synchronizes the motion structure parameters with the manifold, ensuring reliable shape reconstruction. In the context of non-rigid motion, the structure parameters across continuous frames exhibit minimal changes. Such physical continuity is leveraged as a constraint in deriving the motion structure parameter matrix. By incorporating physical continuity, the objective function is designed to dampen the amplitude of shape changes, thereby bolstering the reliability of the motion structure parameters. In essence, the objective function employs manifold alignment and physical continuity constraints to reconcile motion structure parameters with observed data, yielding accurate and reliable outcomes. This approach also stably navigates the challenges posed by imperfect data.

Therefore, building on the previously described calculation of the low-rank matrix for the dynamic shape basis in non-rigid 3D image reconstruction, this process delineates the data structure of parameters associated with this dynamic shape basis. Additionally, it facilitates the acquisition of manifold parameter values amid changes in the non-rigid shape basis. During the projection of the manifold motion group, the manifolds c, d, e, f and the centroid manifold C are used to describe the translation of the object. Assuming that the matrix formed by the 2D coordinates of these five points in the i^{th} frame is denoted as Φ_i , and the initial 3D coordinates are denoted as Φ'_i , then

$$\Phi_i = \begin{bmatrix} u'_{1c} & u'_{1d} & u'_{1e} & u'_{1f} & u'_{1C} \\ v'_{1c} & v'_{1d} & v'_{1e} & v'_{1f} & v'_{1C} \\ \dots & \dots & \dots & \dots & \dots \\ u'_{Fc} & u'_{Fd} & u'_{Fe} & u'_{Ff} & u'_{FC} \\ v'_{Fc} & v'_{Fd} & v'_{Fe} & v'_{Ff} & v'_{FC} \end{bmatrix} \quad (4)$$

$$\Phi'_i = \begin{bmatrix} c_{ix} & c_{ix} & c_{ix} & c_{ix} & c_{ix} \\ c_{ix} & c_{ix} & c_{ix} & c_{ix} & c_{ix} \\ c_{ix} & c_{ix} & c_{ix} & c_{ix} & c_{ix} \end{bmatrix} \quad (5)$$

The transformation function resulting from the equation after camera transformation is:

$$\Phi_i = [\text{quater}(\mathbf{R})] \times [\text{quater}(\mathbf{Q}_i) \times \Phi'_i + \mathbf{T}_i] \quad (6)$$

In Equation (6), $\text{quater}(\mathbf{R})$ represents the camera rotation matrix, $\text{quater}(\mathbf{Q}_i)$ represents the rotation vector of the i^{th}

shape basis, \mathbf{T}_i represents the translation vector of the i^{th} shape basis.

By subjecting each assumed manifold for each frame to the aforementioned camera transformation, we can obtain a 2D matrix Φ composed of the five manifolds in the F frames.

$$\Phi'_i = \begin{bmatrix} u'_{1c} & u'_{1d} & u'_{1e} & u'_{1f} & u'_{1C} \\ v'_{1c} & v'_{1d} & v'_{1e} & v'_{1f} & v'_{1C} \\ \dots & \dots & \dots & \dots & \dots \\ u'_{Fc} & u'_{Fd} & u'_{Fe} & u'_{Ff} & u'_{FC} \\ v'_{Fc} & v'_{Fd} & v'_{Fe} & v'_{Ff} & v'_{FC} \end{bmatrix} \quad (7)$$

Equation (7) provides the solution of 2D manifold coordinates through camera transformation with the assumed 3D coordinates of manifolds. After taking the difference between the matrix Φ and the actual measured values of 2D coordinates in the manifold group, shown in Equation (8), we obtain the matrix \mathbf{DE}_1 . Finally, we compute the two-norm of all elements in this matrix to obtain de_1 :

$$de_1 = \sqrt{\sum_{j=1}^5 \sum_{i=1}^{2F} \|\mathbf{DE}_1\|^2} = \sqrt{\sum_{j=1}^5 \sum_{i=1}^{2F} \|\Phi_{ij} - \mathbf{M}_{ij}\|^2} \quad (8)$$

During the optimization and training of the entire motion structure matrix, the displacement and velocity changes of feature points between consecutive frames are slow for a high-speed camera [14], [15]. Reflecting such a property in the rotation matrix \mathbf{Q} and translation matrix \mathbf{T} of non-rigid body motion, the variation in the rotation matrix \mathbf{Q}^T reflects the change in rotation angle between consecutive frames according to the displacement difference formula $\Delta S = S_i - S_{i-1}$ (where S_i represents the displacement of a point in the i^{th} frame). Similarly, the expression of displacement difference S in the translation matrix \mathbf{T} represents the change in the non-rigid body's translational position during motion. Therefore, the parameters in the i^{th} frame should exhibit very small changes compared to the $(i - 1)^{\text{th}}$ frame. Here, these two matrices are combined into one matrix \mathbf{RS} with size $F6$.

$$\mathbf{RS}_{F6} = \begin{bmatrix} \theta_{1X} & \theta_{1Y} & \theta_{1Z} & X_1 & Y_1 & Z_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{iX} & \theta_{iY} & \theta_{iZ} & X_i & Y_i & Z_i \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{FX} & \theta_{FY} & \theta_{FZ} & X_F & Y_F & Z_F \end{bmatrix} \quad (9)$$

Based on the above discussion, the difference between the i^{th} and the $(i - 1)^{\text{th}}$ the frames of \mathbf{RS}_{F6} combines to a new matrix \mathbf{RS}' with size $(F - 1)6$ in Equation (10), as shown at the bottom of the next page. Each element of this matrix represents the variation in parameters between consecutive frames in non-rigid body motion. Such changes are minimal during the capture process of a camera.

Due to the continuity of non-rigid body motion and the motion structure parameters reflected in the image sequence obtained through high-speed capture [16], [17], the variation between consecutive frames is very small. This introduces the first constraint de_2 added during the construction of the objective function, describing the variation in the

displacement of parameters between consecutive frames.

$$de_2 = \sqrt{\sum_{j=1}^6 \sum_{i=1}^{F-1} \|\mathbf{RS}'_{ij}\|^2} \quad (11)$$

Image sequences taken from the high-speed camera have a resolution of 640×480 and a frame rate of at least 1000 frames per second, save rapidly moving objects and subtle changes, and provide more accurate details. Assuming the time interval t between frames of the extracted image sequence, and the velocity interval for each frame's image capture is a small constant. The physical motion laws shows:

$$\bar{v}_i = \frac{\Delta s}{t} = \frac{S_i - S_{i-1}}{t} \text{ and } \Delta v = \bar{v}_i - v_{i-1} \quad (12)$$

The ratio of the displacement difference of an object during this time interval to t is the average velocity of the motion. Similarly, for a high-speed image sequence, such changes in average velocity are also very small. Assuming the average velocity from the i^{th} to the $(i-1)^{\text{th}}$ frame is \bar{v}_i , then the average velocity between the $(i+1)^{\text{th}}$ frame and the i^{th} frame is \bar{v}_{i+1} . This physical law is used into the motion parameter matrix of the non-rigid body, shown as the difference matrix \mathbf{RS}'' between the i^{th} and $(i-1)^{\text{th}}$ frames with size $(F-2) \times 6$. At this point, a constraint de_3 is obtained, describing the velocity changes in non-rigid body motion.

$$de_3 = \sqrt{\sum_{j=1}^6 \sum_{i=1}^{F-2} \|\mathbf{RS}''_{ij}\|^2} = \sqrt{\sum_{j=1}^6 \sum_{i=1}^{F-1} \|\mathbf{RS}'_{ij} - \mathbf{RS}'_{i-1j}\|^2} \quad (13)$$

After this discussion, we have obtained the error de_1 between the points describing the transformation of the manifold motion group and the measured points as the main subject for minimizing. Additionally, we have two constraints de_2 and de_3 describing the displacement and velocity changes of parameters in the non-rigid body motion process. In a set of high-speed captured image sequences, the variations in these two constraints between consecutive frames should be very small. This results in a strong and reliable objective function $f(\mathbf{R}, \mathbf{Q}, \mathbf{T})$ used to determine the motion structure parameter matrix.

$$\begin{aligned} f(\mathbf{R}, \mathbf{Q}, \mathbf{T}) &= \sum_{j=1}^6 \sum_{i=1}^3 w_i de_i \\ &= w_1 \sqrt{\sum_{j=1}^5 \sum_{i=1}^{2F} \|\Phi_{ij} - \mathbf{M}_{ij}\|^2} \\ &\quad + w_2 \sqrt{\sum_{j=1}^6 \sum_{i=1}^{F-1} \|\mathbf{RS}'_{ij}\|^2} \\ &\quad + w_3 \sqrt{\sum_{j=1}^6 \sum_{i=2}^{F-1} \|\mathbf{RS}'_{ij} - \mathbf{RS}'_{i-1j}\|^2} \end{aligned} \quad (14)$$

$\text{Min}_{\mathbf{Q}, \mathbf{R}, \mathbf{T}} \|f(\mathbf{Q}, \mathbf{R}, \mathbf{T})\|$ is the objective function used to train the motion structure parameter matrix, which consists of three motion structure parameter matrices $\mathbf{Q}, \mathbf{R}, \mathbf{T}$ that describe camera transformations. These matrices represent the rotation matrix of the shape basis, the camera rotation matrix, and the translation matrix of the shape basis, respectively; w_1, w_2, w_3 are the weight coefficients of different constraint parts in the entire objective function. By considering the constraints of reconstruction error and minimal frame changes before and after shape deformation, the objective function aims to achieve its minimum value, ensuring that the reconstruction results are closer to the actual observation data and minimizing unreasonable shape changes. Therefore, by minimizing the objective function $f(\mathbf{Q}, \mathbf{R}, \mathbf{T})$, the reliability of the reconstruction results can be improved. Next, based on the above, the Levenberg-Marquardt (L-M) nonlinear optimization method is then employed to determine these motion structure parameters [18], and complete the selection of key initial values for the shape basis.

2) THE SELECTION OF INITIAL VALUES

The essence of solving the problem using nonlinear optimization is to determine the description of the rotation matrix \mathbf{Q} and the translation matrix \mathbf{T} . In the L-M optimization method, these two parameters are treated as unknown. First, introducing the initialization method for the rotation matrix \mathbf{Q} :

In this paper, a method combining factorization techniques is used for the initialization of the rotation matrix \mathbf{Q} [19]. As analyzed in Chapter Three, considering a known measurement matrix \mathbf{W}_{2FP} composed of feature points $\begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} \begin{matrix} i=1, \dots, F \\ j=1, \dots, P \end{matrix}$ from each frame of images (where F is the number of image frames and P is the number of feature points in one frame), the objective is to determine the 3D structure \tilde{S}_{i3P} and the rotation matrix $\mathbf{R}_{i3 \times 3}$ for each frame image.

Assuming that the 3D shape of the non-rigid body is a weighted linear combination of shape bases, we have: $\tilde{S}_i = \sum_{l=1}^K \omega_{il} S_l$, where ω_{il} is the weight coefficients, S_l is the shape bases, and K is the number of shape bases. Therefore, based on the number of shape bases K and the weighting coefficient ω_{il} , when $K=1, \omega_{il} = 1$, corresponding to the case of a rigid object; When $K > 1, \omega_{il} > 1$, it corresponds to the case of non-rigid objects.

Under the weak perspective projection model, there is

$$\begin{bmatrix} u_{i1}, \dots, u_{iP} \\ v_{i1}, \dots, v_{iP} \end{bmatrix} = \bar{R}_i (\sum_{l=1}^K \omega_{il} S_l) + \bar{T}_i \mathbf{e}_n^T \quad (15)$$

where \bar{R}_i represents the first two rows of the rotation matrix $\mathbf{R}_{i3 \times 3}$, \bar{T}_i is the first two elements of the translation vector

$$\mathbf{RS}'_{(F-1)6} = \begin{bmatrix} \theta_{2X} - \theta_{1X} & \theta_{2Y} - \theta_{1Y} & \theta_{2Z} - \theta_{1Z} & X_2 - X_1 & Y_2 - Y_1 & Z_2 - Z_1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \theta_{FX} - \theta_{F-1X} & \theta_{FY} - \theta_{F-1Y} & \theta_{FZ} - \theta_{F-1Z} & X_F - X_{F-1} & Y_F - Y_{F-1} & Z_F - Z_{F-1} \end{bmatrix} \quad (10)$$

$\mathbf{T}_{i3 \times 1}$, and \mathbf{e}_n^T represents the position vector of the object in the camera coordinate system, $\mathbf{e}_n^T = [1, \dots, 1]_{1 \times n}$.

Transforming Equation (15) with the 2D coordinate origin as the centroid, we have $\tilde{\mathbf{W}} = \mathbf{M}_{2F \times 3K} \mathbf{B}_{3K \times P}$, where $\mathbf{M}_{2F \times 3K}$ represents the projection from the 3D coordinates of the shape basis to 2D image plane, $\mathbf{B}_{3K \times P}$ represents the three-dimensional shape coordinates obtained from the shape transformation of P feature points using the shape basis matrix \mathbf{B} .

The matrix $\mathbf{M}_{2F \times 3K}$ contains information about the rotation matrix. To further decompose the motion matrix \mathbf{M} and obtain the rotation matrix $[\bar{\mathbf{R}}_i]$ and the weighting coefficient ω_{il} , we perform singular value decomposition (SVD) [20], [21] on \mathbf{M} by rearranging \mathbf{M} with a matrix block.

$$\mathbf{M} = [\mathbf{M}_1^T, \mathbf{M}_2^T, \dots, \mathbf{M}_F^T]^T \quad (16)$$

Then, each matrix block is given by:

$$\mathbf{M}_i = [\omega_{i1} \bar{\mathbf{R}}_i \dots \omega_{iK} \bar{\mathbf{R}}_i]_{2 \times 3K} \quad (17)$$

where $\bar{\mathbf{R}}_i = \begin{bmatrix} r_{i1} & r_{i2} & r_{i3} \\ r_{i4} & r_{i5} & r_{i6} \end{bmatrix}$. Reordering the matrix blocks, we have:

$$\tilde{\mathbf{M}}_i = \begin{bmatrix} \omega_{i1} \\ \omega_{i2} \\ \vdots \\ \omega_{iK} \end{bmatrix} [r_{i1} \ r_{i2} \ \dots \ r_{i6}] = \tilde{\mathbf{\Omega}}_i \tilde{\mathfrak{R}}_i \quad (18)$$

Clearly, the most rank of $\tilde{\mathbf{M}}_i$ is 1. Therefore, performing SVD on $\tilde{\mathbf{M}}_i$ can decompose it into the deformed rotation matrix $\tilde{\mathfrak{R}}_i$ and the weighted coefficient matrix $\tilde{\mathbf{\Omega}}_i$. By applying SVD to each of the F matrices $\tilde{\mathbf{M}}_i$, we can obtain the deformed rotation matrices $\tilde{\mathfrak{R}}_i$ and the weighted coefficient matrices $\tilde{\mathbf{\Omega}}_i$ for each frame.

The decomposition result of Equation (18) is still not unique. Thinking about $\tilde{\mathbf{M}}_i = \tilde{\mathbf{\Omega}}_i C \frac{1}{C} \tilde{\mathfrak{R}}_i$ for any non-zero constant C , the C can be optimized by minimizing Equation (19):

$$f(C) = \min \|\tilde{\mathbf{\Omega}}_i - \tilde{\mathbf{\Omega}}_{i-1}\|_F \quad (19)$$

Further, we can obtain $\tilde{\mathfrak{R}}_i = \frac{1}{C} \tilde{\mathfrak{R}}_i$ and $\tilde{\mathbf{\Omega}}_i = \tilde{\mathbf{\Omega}}_i C$.

Adjusting the row vector $\tilde{\mathfrak{R}}_i$ yields a 2×3 rotation matrix $\bar{\mathbf{R}}_i$. However, the rotation matrix obtained through the singular value decomposition of $\tilde{\mathbf{M}}_i$ is $\bar{\mathbf{R}}_i = \begin{bmatrix} r_{i1} & r_{i2} & r_{i3} \\ r_{i4} & r_{i5} & r_{i6} \end{bmatrix}$. Therefore, in order to avoid singular value problems, maintain the continuity of rotation, and reduce numerical errors, the quaternion method is employed to address rotation matrices in this study [22]. The rotation matrix part for each frame is expressed using three parameters $\mathbf{R}_i = [a_i, b_i, c_i]$.

According to the quaternion method and the exponential mapping of the rotation matrix, $a_i = r_{i6}; b = -r_{i3}; c = r_{i2}$ is used as the initial value for the rotation matrix \mathbf{Q}_i in the i^{th} frame [23]. Once \mathbf{Q}_i is obtained, it is substituted into the objective function. At this point, \mathbf{T}_i is initialized by a random

initialization method. Then, a nonlinear optimization method is applied to the Equation (15) to obtain a set of values for \mathbf{T}_i , denoted as \mathbf{T}'_i . Then, \mathbf{T}'_i and \mathbf{Q}_i are used as the initial values for the final estimation of the motion structure parameter matrix. This process significantly enhances the reliability of the following reconstruction.

C. EUCLIDEAN RECONSTRUCTION

With the motion structure parameter matrices $\mathbf{Q}, \mathbf{R}, \mathbf{T}$, the known 2D feature points of image sequence are optimized to obtain the 3D coordinates for each point. Let the known 2D feature points $\begin{bmatrix} u_{ij} \\ v_{ij} \end{bmatrix} \ i = 1, \dots, F \ j = 1, \dots, P$ form the measurement matrix

$$\mathbf{W}_{2F \times P} = \begin{bmatrix} u_{11} & \dots & u_{1P} \\ v_{11} & \dots & v_{1P} \\ \dots & \dots & \dots \\ u_{F1} & \dots & u_{FP} \\ v_{F1} & \dots & v_{FP} \end{bmatrix}$$

(where F is the number of frames in the image and P is the number of feature points in the image), the reconstruction of non-rigid motion is as follows: For each point P in the F frames, the known 2D coordinates of this point can be represented as:

$$\mathbf{W}_j = [u_{1j}, v_{1j}, \dots, u_{Fj}, v_{Fj}]^T \quad (20)$$

\mathbf{W}_j represents the j^{th} feature point in the image sequence, and it is a known feature point obtained from the image sequence. The 3D reconstruction based on non-linear optimization uses an assumed 3D feature point

$\mathbf{W}'_j = \begin{bmatrix} X_{1j} & Y_{1j} & Z_{1j} \\ \dots & \dots & \dots \\ X_{Fj} & Y_{Fj} & Z_{Fj} \end{bmatrix}$ as the parameters to be solved. After transforming it with the obtained parameter matrix, the 2D coordinates $\mathbf{T}_j = [u'_{1j}, v'_{1j}, \dots, u'_{Fj}, v'_{Fj}]^T$ of the points are obtained.

Then the matrix de_4 shows the difference between \mathbf{W}_j and \mathbf{T}_j , which is the main part of the objective function. For the j^{th} feature point in the i^{th} frame, the camera transformation is given by:

$$\text{de}_4 = \mathbf{T}_{ij} - \mathbf{W}_{ij} = [\text{quater}(\mathbf{R})] \times [\text{quater}(\mathbf{Q}_i) \mathbf{W}'_{ij} + \mathbf{T}_i] - \mathbf{W}_{ij} \quad (21)$$

In Equation (21), the known parameter is the j^{th} feature point in the measurement matrix, and the unknown parameter of the objective function is the 3D coordinate point \mathbf{W}'_{ij} to be determined. Due to the continuity of the object's motion, the change in the displacement of the matrix de_4 between consecutive frames should be very small according to the displacement difference formula $\Delta S = S_i - S_{i-1}$. Based on this constraint, the strong reliability objective function for the non-linear optimization of the j -th feature point \mathbf{W}'_j is obtained:

$$\text{Min}_{\mathbf{w}'_j} \|f(\mathbf{W}'_j)\| = \sqrt{\sum_{i=3}^{2F} \|\text{DE}\mathbf{A}_{i1} - \text{DE}\mathbf{A}_{i-21}\|^2} \quad (22)$$

Equation (22) is the objective function for the 3D reconstruction of the j^{th} feature point. Using the L – M non-linear optimization method, \mathbf{W}_j^T is calculated as the 3D coordinates of the j^{th} point with size $3F$. Then, applying the same 3D reconstruction to all feature points, the 3D coordinates of P feature points are represented by TS.

$$\mathbf{TS}_{3FP} = \begin{bmatrix} \mathbf{W}_1^T \\ \mathbf{W}_2^T \\ \dots \\ \mathbf{W}_P^T \end{bmatrix} \quad (23)$$

Equation (23) represents a matrix of size $3FP$, which is the result of the 3D reconstruction of all feature points. This matrix represents the 3D coordinates of the feature points, enhancing the reliability of the reconstruction.

In summary, the proposed algorithm for non-rigid 3D image reconstruction addresses the significant limitations identified in existing algorithms, particularly their challenges in establishing stable shape objective functions and determining initial values for nonlinear, non-rigid body movements. The methodology of the algorithm unfolds in a series of meticulously designed steps. Initially, it employs image points and depth factors to construct a low-rank matrix that accurately describes the dynamic shape basis of non-rigid bodies. This matrix not only facilitates a better restoration of the transformations of the non-rigid body shape basis but also provides precise manifold parameters essential for the formulation of objective functions. Subsequently, the algorithm leverages manifold alignment and physical continuity constraints to optimize the construction of the objective function. This optimization step is crucial for aligning the motion structure parameters with the observed data and reducing the amplitude of shape changes, thereby ensuring the accuracy and reliability of the reconstruction results.

Further enhancing the robustness of the algorithm, constraints on reconstruction error and minimal frame-to-frame shape changes are integrated. These constraints ensure that the minimization of the objective function yields reconstruction results that closely align with actual observational data, while also preventing unrealistic alterations in shape, thereby improving the reliability of the reconstruction outcomes. Following this, the L-M (Levenberg-Marquardt) nonlinear optimization method is applied to solve for the motion structure parameters and to select the key initial values of the shape basis, based on the minimized objective function. The culmination of these steps is the Euclidean reconstruction, which, by using the solved motion structure parameters, obtains the 3D coordinates of each point, achieving a reliable reconstruction process. This comprehensive approach not only overcomes the deficiencies of prior models but also sets a new benchmark for accuracy and reliability in non-rigid 3D image reconstruction.

III. EXPERIMENTAL RESULTS

A. EXPERIMENTAL PARAMETERS

Experiment data is the Stanford 40 Actions dataset from the Stanford Image Library, which is a large-scale image

and video database created by the Stanford Vision Lab. This database contains multiple datasets, covering different tasks in fields such as computer vision and machine learning.

The Stanford 40 Actions dataset contains approximately 9500 video clips, of which 6000 are male and 3500 are female. There are 2500, 5000, and 2500 video clips for different age distributions, including adolescents, young people, and elderly people. And this dataset covers 40 different categories of human actions. Therefore, use this dataset to complete the reconstruction performance test of the proposed algorithm. To ensure the effectiveness of the test, it is necessary to set the experimental parameters, as shown in Table 1.

TABLE 1. Parameter settings.

Parameter	Numerical value
Depth Factor	0.04
Weight of point reprojection error	0.5
Constraint threshold for small frame changes before and after shape	0.01
Iteration Times of L-M Nonlinear Optimization Method	20
Initial Step Size of L-M Nonlinear Optimization Method	0.1

B. IMPACT OF THE IMAGE NUMBER ON THE PROPOSED MODEL

Taking dynamic facial images as an example to verify the impact of the image number on the proposed algorithm [24], first a set of changing camera intrinsic parameters $f_i = 800 + i$ (i representing the i -th dynamic image) is generated, with scale factor 1.1, distortion factor 0.5, and intrinsic matrix $(u_0, v_0) = (320, 240)$.

Then, 100 3D points within a unit sphere are randomly generated and divided into three rigid body elements: the first rigid body element consists of the first 50% of the space points; the second consists of the middle 30%, and the third consists of the last 20%. Simultaneously, the external parameter matrix of the camera is varied to generate images of sizes 640×480 from 20 to 200 frames. Adding 1 pixel of Gaussian noise to each image 100 times under each number of images, the average reprojection error e is calculated in Equation (24), which refers to the difference between the reconstructed 3D model and the original image, that is, the difference between the reconstructed 3D model and the original image after being reprojected back to the image plane. It focuses more on the overall consistency between the reconstructed 3D model and the original image.

$$e = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \frac{1}{\lambda_{i,j}} \left\| \mathbf{m}_{i,j} - \left(\mathbf{P}_i \begin{pmatrix} X_{i,j} \\ 1 \end{pmatrix} \right) \right\| \quad (24)$$

In Equation (24), e represents the reprojection error, i and j denote the i^{th} frame and the j^{th} space point, m and n represent the number of images and space points, $\mathbf{m}_{i,j}$ and \mathbf{p}_i represent the j^{th} image point on the i^{th} frame and the projection matrix of the camera, $X_{i,j}$ and $\|\cdot\|$ represent the j^{th} 3D space point on

the i^{th} frame and the 2-norm, respectively, $\lambda_{i,j}$ represents the depth factor of the j^{th} 3D spatial point in the i^{th} frame.

Therefore, using this indicator, the impact of the image number on the proposed model is analyzed to demonstrate the reconstruction effect. The experimental results are shown in Figure 1.

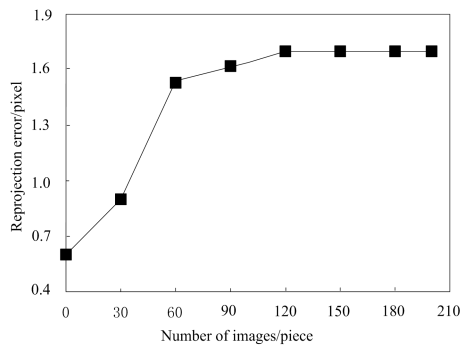


FIGURE 1. Variation of reprojection error with the number of images.

The physical meaning of Equation (24) is that it calculates the average error between the reprojected image points and the actual image points. A smaller reprojection error indicates a higher reconstruction accuracy, while a larger error implies lower reconstruction accuracy. Additionally, to investigate the impact of the number of space points on the algorithm, following the same procedure as above, the camera's internal parameters were varied, while the number of space points ranged from 20 to 150. These space points were divided into three rigid body elements according to the aforementioned ratio. Using these 3D space points, 150 images were generated, and 1 pixel of Gaussian noise was added to each image. The algorithm was run 100 times for each number of space points, and the average reprojection error was calculated, as shown in Figure 2.

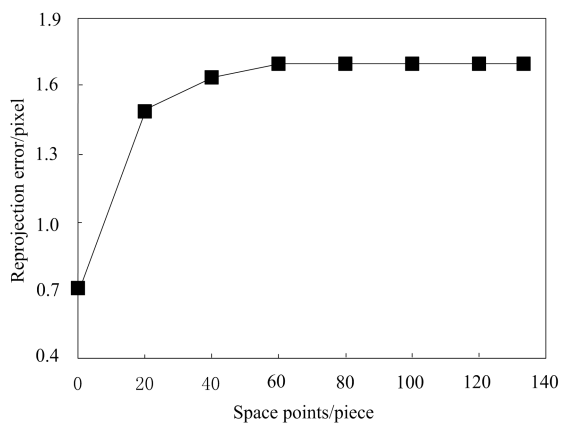


FIGURE 2. Variation of reprojection error with spatial points.

From Figures 1 and 2, it can be seen that as the number of spatial points and images increases, the amplitude of the reprojection error in our algorithm will show a trend of first increasing and then decreasing. The reason is that when the number of spatial points and images is relatively small, the

number of equations and unknowns is close, making the solution relatively unstable and more susceptible to noise, resulting in a larger residual error; On the contrary, when there are more spatial points and images, the number of equations is greater than the unknowns, and the solving process is more overconstrained and stable, so the residual difference is relatively small. So as the number of spatial points and images increases, the residual variation will show a trend of first increasing and then decreasing, and eventually stabilizing. This indicates that the number of images has a certain impact on the model in this article. The more images there are, the smaller the reprojection error, and the better the reconstruction processing effect of the proposed algorithm.

C. ANALYSIS OF THE IMPACT OF DEPTH FACTOR VALUES ON CAPTURING GLOBAL FEATURE QUANTITY

In non-rigid 3D image reconstruction, the value of depth factor determines the field of view of camera imaging. Based on the requirements of the high-speed camera used in the article, the range of depth factor values is set to [0.01,0.04]. One image is randomly selected from the Stanford 40 Actions dataset, and imaging is performed on the image at different depth factor values. The number of global features that can be captured is counted, and the more features can be captured, the more, The better the subsequent reconstruction effect, the capturing global features are shown in Table 2.

TABLE 2. Results of capturing global feature quantity.

λ Depth factor value	Capture the number of global features
0.01	25
0.02	42
0.03	57
0.04	69

According to the results presented in Table 2, as the depth factor increases, there is an upward trend in the number of global features that can be captured. This upward trend is attributable to the fact that as the depth factor rises, distant objects within the camera's imaging range become encompassed within the field of view, resulting in an expanded number of global features that can be identified and captured. Therefore, in the above parameter settings, the depth factor values set are effective and can provide reliable support for subsequent 3D image reconstruction to improve the reconstruction effect.

D. RECONSTRUCTION RESULTS

Selecting any two samples from the moving facial image and three frames each (62nd frame, the 98th frame, the 141st frame, and the 51nd frame, the 52th frame, and the 58st frame, respectively), we have Figure 3.

The selected three frames were reconstructed using the algorithm proposed in this paper, and the results are shown in Figure 4.

Additionally, to further illustrate the superior performance of the proposed algorithm in non-rigid 3D image reconstruction, five comparison algorithms from literature [7]

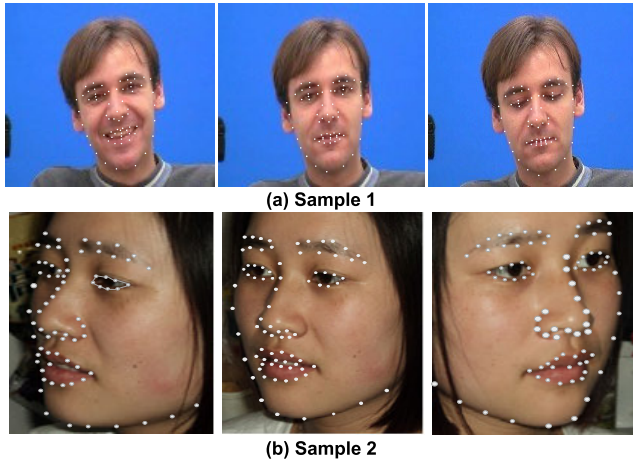


FIGURE 3. Moving facial image sequence.

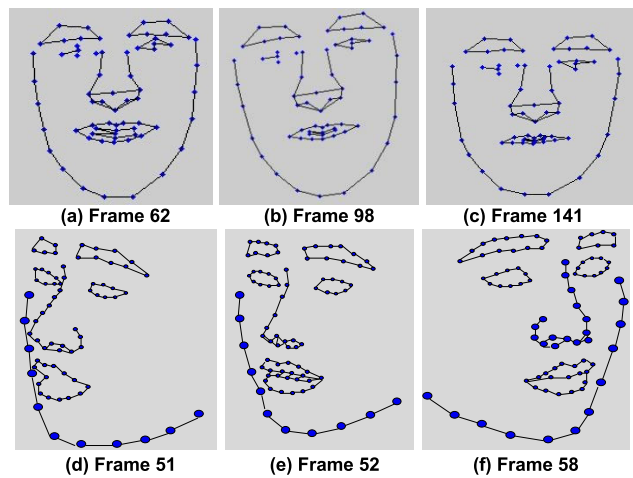


FIGURE 4. Model reconstruction results in this article.

based on deep learning, literature [8] based on regularization operators, literature [9] based on translation kernel factorization, literature [10] based on projection, and literature [11] based on constrained bilateral smoothing and dynamic mode decomposition were selected. These five comparison algorithms were used to reconstruct the selected images in the study, and the results are shown in Figures 5-9.

By analyzing Figures 4-9, it can be concluded that the proposed algorithm can effectively restore the 3D structure and motion of non-rigid bodies. The reconstruction results of the selected image are basically consistent with the initial image, indicating that the proposed algorithm can effectively achieve the goal of 3D non-rigid body reconstruction.

There are certain differences between the reconstruction results of the five comparative algorithms and the initial image. Among them, the reconstruction results based solely on deep learning algorithms have a small difference from the initial image, but there are angle issues and missing details. Therefore, comparing the reconstruction results of the five comparative algorithms with the reconstruction results of the algorithm in this paper, it can be seen that the robustness and reconstruction effect of the algorithm in this paper are

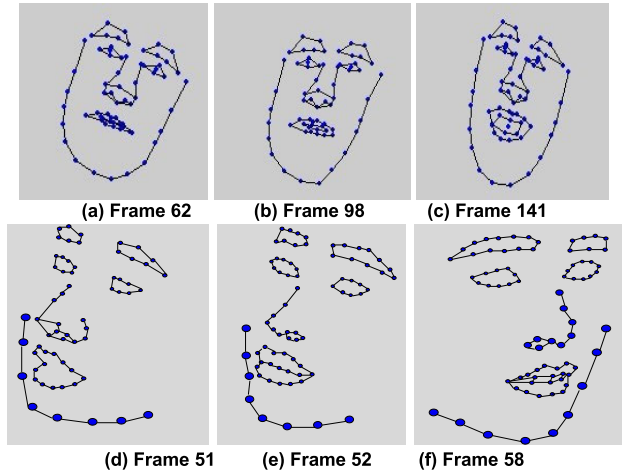


FIGURE 5. Reconstruction model based on deep learning algorithm.

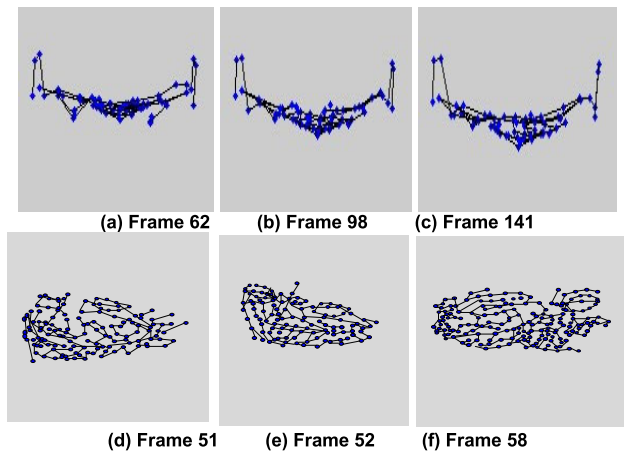


FIGURE 6. Reconstruction model based on regularization operator.

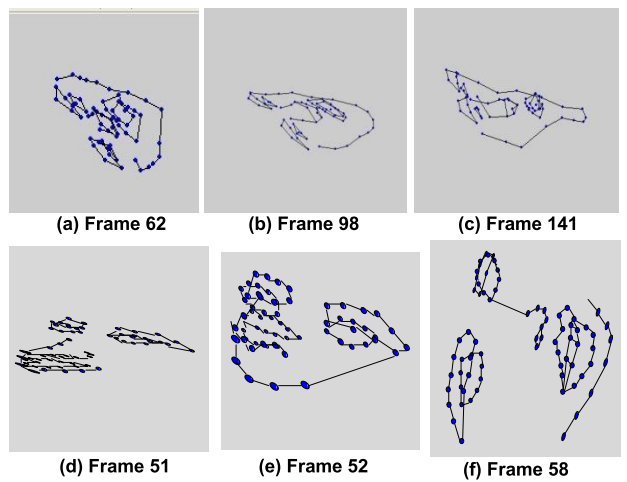


FIGURE 7. Reconstruction based on translation kernel factorization.

better than those of several comparative algorithms. The main reason is that the algorithm in this paper solves the problem of robustness degradation caused by the continuous change of shape basis in non-rigid motion images. Compared

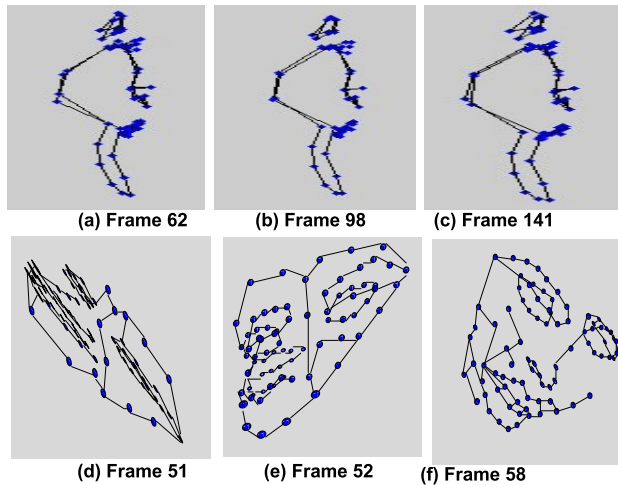


FIGURE 8. Projection based reconstruction model.

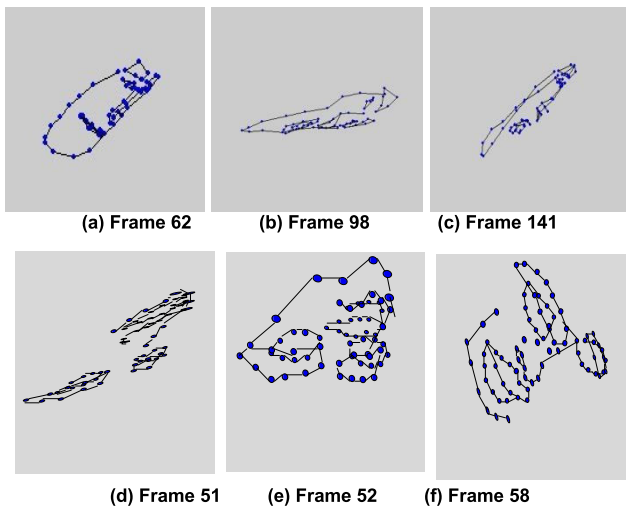


FIGURE 9. Reconstruction model based on constrained bilateral smoothing and dynamic mode decomposition.

with the fixed shape basis calculation applied in other comparative algorithms, The proposed algorithm determines the low rank matrix of the dynamic shape basis for non-rigid 3D image reconstruction, describes the data structure of the dynamic shape basis variables for non-rigid 3D image reconstruction, captures local and global features of the shape, and better represents the motion robustness process. Therefore, compared with several comparative algorithms, it has higher reconstruction robustness and effectiveness.

Further validating the local reconstruction results of the proposed algorithm, AIDR 3D reconstruction technology is an advanced medical image processing technique that uses iterative methods to reconstruct high-quality 3D images. This technology is based on a series of mathematical and physical principles, including filtering, backprojection, reconstruction algorithms, etc. It can shorten imaging time, reduce radiation dose, and lower imaging costs. Therefore, AIDR 3D reconstruction technology was selected for comparison with the proposed algorithm. Taking the first set of male image

sequences as an example, AIDR 3D reconstruction and the proposed algorithm were used to reconstruct the frontal, lateral, and top views of the local nasal tip, and the compared results were shown in Figures 10 and 11.

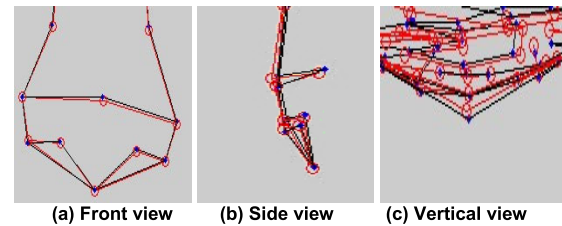


FIGURE 10. Comparison of frontal, lateral, and overhead views of the reconstructed nasal tip using the proposed algorithm.

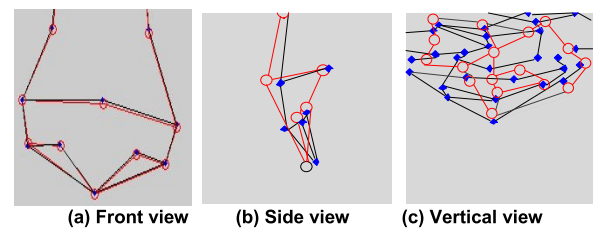


FIGURE 11. Comparison of frontal, lateral, and overhead views of the reconstructed nasal tip using the AIDR 3D algorithm.

In Figures 10 and 11, the red circles and lines represent the locally reconstructed nasal tip image using the proposed algorithm, while the blue dots and lines represent the local feature points of the nasal tip in the dynamic facial image. From Figure 10, it can be observed that the locally reconstructed nasal tip image using the proposed algorithm conforms to the local features of the nasal tip in dynamic facial images. From Figure 11, it can be observed that the frontal image of the nasal tip reconstructed locally using AIDR 3D reconstruction technology is basically consistent with the proposed algorithm, and can well represent the local feature points of the nasal tip. However, when reconstructing the side and top views, the reconstruction results are significantly different from the local feature points of the nasal tip, and the reconstruction effect is not good compared to the proposed algorithm. This indicates that the reconstructed nasal tip of the proposed algorithm can better reflect the characteristics of high nasal bridges in European and American individuals, proving the effectiveness of the proposed algorithm.

In order to visually illustrate the effect of incorporating a constraint term for non-rigid motion and velocity variation into the proposed algorithm, the left eye reconstruction results of the proposed algorithm were compared with those of five contrastive algorithms, as shown in Figure 12.

In Figure 12, the red circles represent the left eye reconstruction results under the constraint of non-rigid motion and velocity variation, while the blue dots indicate the relative positions of the eye and eyebrow. Based on the results in Figure 12, it can be observed that the proposed algorithm clearly reflects the left eye reconstruction results under the

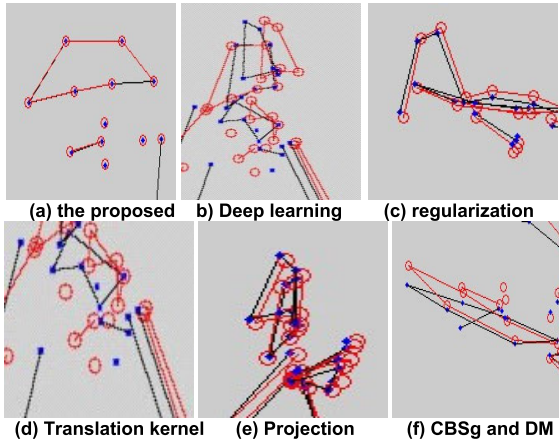


FIGURE 12. Left eye reconstruction results of the proposed algorithm and five comparison algorithms.

constraint of non-rigid motion and velocity variation, which is consistent with the reconstructed model results presented in this paper, and the relative positions of the eye and eyebrow indicated by the blue dots are relatively accurate. This indicates that the incorporation of a constraint for non-rigid motion and velocity variation in the proposed algorithm results in a more continuous deformation of the non-rigid body, leading to a more accurate reconstruction.

E. RECONSTRUCTION PERFORMANCE

Back projection error refers to the error generated during the projection process from a 3D model to a 2D image. It measures the difference between the pixel values projected by the model onto the image plane and the actual pixel values. Compared to reprojection error, it focuses more on the accuracy of individual pixel points and is more detailed. It can be used to evaluate the accuracy and reliability of reconstruction algorithms. Therefore, this article uses the variation of back projection error during the iteration process to prove the accuracy and reliability of this algorithm. The back projection error can be defined as:

$$\sigma = \frac{\|\mathbf{W} - \mathbf{W}_r\|_F}{\|\mathbf{W}\|_F} 100\% \quad (25)$$

In Equation (25), \mathbf{W} is the original measurement matrix, and \mathbf{W}_r is the measurement matrix obtained from the reconstructed backprojection. The backprojection errors obtained under the algorithm of this article and five comparison algorithms are shown in Figure 13.

From Figure 13, it can be seen that when the number of iterations reaches 1000, the average back projection errors of reconstruction algorithms based on deep learning, regularization operator, translation kernel decomposition, projection reconstruction, constrained bilateral smoothing, and dynamic mode decomposition are 0.59%, 0.80%, 0.90%, 1.32%, and 1.51%, respectively. In contrast, the average inversion error of the proposed algorithm is only 0.33%.

This indicates that the proposed algorithm has the smallest back projection error compared to the five comparison

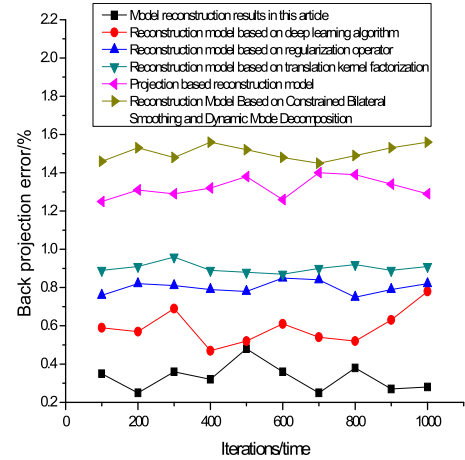


FIGURE 13. Backprojection errors obtained under the proposed algorithm and five comparative algorithms.

algorithms, indicating that the reconstruction results of the proposed algorithm are more accurate and reliable. This is because the proposed algorithm constructs a low rank matrix describing the dynamic non-rigid body shape basis by combining image points and depth factors before reconstruction. The algorithm can effectively restore the process of non-rigid body shape basis changes and optimize the construction objective function using manifold alignment and physical continuity constraints. The L-M nonlinear optimization method is used to solve the problem and obtain the key initial values of the shape basis, This further reduces the back projection error of the reconstruction results, making them closer to the actual shape changes, more accurate and reliable, and having good reconstruction performance, which can effectively achieve accurate reconstruction of non-rigid 3D images. To further verify the reconstruction performance of the proposed algorithm, a 3D ShapeNet dataset was selected for comparative testing of the back projection error index and the above 5 comparison algorithms. Randomly select 5 sets of images from this dataset for 3D reconstruction, each containing 100 images. The back projection error results of each algorithm reconstruction are shown in Table 3.

According to the results obtained in Table 3, the proposed algorithm can also have good reconstruction performance in this dataset, with an average back projection error of 0.30%. The reconstruction algorithms based on deep learning, regularization operator, translation kernel decomposition, projection reconstruction, constrained bilateral smoothing, and dynamic mode decomposition have their respective average back projection errors of 0.60%, 0.88%, 1.01%, 1.32%, and 1.59%, respectively. In contrast, the proposed algorithm can still maintain a low reconstruction backprojection error, has good reconstruction accuracy, performs well in reconstruction, and can ensure the quality of reconstruction results. The reason is that the proposed algorithm utilizes image points and depth factors to form a low rank matrix, which can better capture the changing characteristics of

TABLE 3. Back projection error result.

Number of image groups	Proposed algorithm	Deep learning algorithms	Regularization operator algorithm	Translation Kernel Factorization Algorithm	Projection reconstruction algorithm	Constrained Bilateral Smoothing and Dynamic Mode Decomposition Algorithm
1	0.28%	0.55	0.83	0.92	1.24	1.55
2	0.32%	0.63	0.92	1.10	1.38	1.63
3	0.25%	0.53	0.81	0.90	1.19	1.50
4	0.34%	0.69	0.96	1.14	1.45	1.68
5	0.31%	0.61	0.90	0.99	1.37	1.61
Average backprojection error	0.30%	0.60%	0.88%	1.01%	1.32%	1.59%

dynamic non-rigid body shape bases. By using low rank representation, the dimensionality of the original data can be effectively reduced while retaining important information, which is beneficial for reducing information loss during the reconstruction process and improving the accuracy of the reconstruction. And through manifold alignment and physical continuity constraints, the shape changes during the reconstruction process are limited within a reasonable range to avoid unreasonable distortion or deformation, thereby helping to maintain the authenticity and stability of the reconstruction results and reducing reconstruction backprojection errors.

The Structural Similarity Index (SSIM) is a metric used to compare the degree of similarity between two images. The SSIM can be employed to assess the quality loss in images and is widely utilized in image processing and compression algorithms. It evaluates the similarity between images by comparing three aspects: brightness, contrast, and structure. Brightness compares the average brightness of the images, contrast assesses the differences in contrast of the images, and structure compares the structural information in the images. All three aspects take into consideration the characteristics of human visual perception, thus making the evaluation more in line with human perception of image similarity. The SSIM value typically ranges from -1 to 1 , where 1 indicates complete similarity between two images, 0 indicates no similarity, and -1 indicates complete dissimilarity. In order to verify the degree of detail feature preservation in the 3D motion facial images reconstructed using this algorithm, the restoration effect of the algorithm on image details was compared with five motion facial image comparison algorithms through SSIM analysis. As shown in Figure 3, select the moving face image of Sample 1 as the sample, extract 68 facial feature points, and compare the SSIM index of the images obtained by our algorithm with five comparison algorithms, as shown in Figure 14.

From Figure 14, it can be observed that for the 68 facial feature points, the SSIM indices of the images obtained using reconstruction algorithms based on deep learning, regularization operator, translation kernel factorization, projection-based reconstruction, and constrained bilateral smoothing and dynamic mode decomposition reach 0.962, 0.125, 0.569, 0.124, and 0.397, respectively.

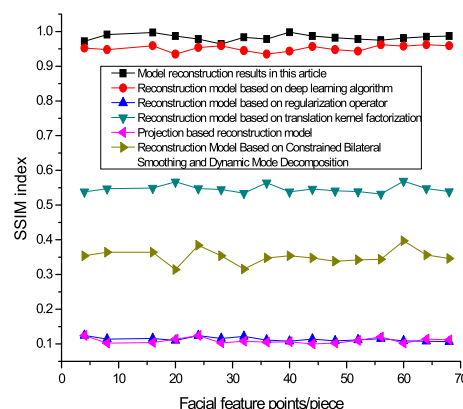


FIGURE 14. SSIM index of images obtained by the proposed algorithm and five comparison algorithms.

When using the algorithm proposed in this article for non-rigid 3D image reconstruction, the SSIM index reaches a maximum of 0.998, close to 1, indicating that the difference between the reconstructed image using the algorithm proposed in this article and the moving facial image is relatively small. This is because the algorithm in this article uses image points and depth factors to form a low rank matrix that describes the dynamic non-rigid body shape basis, better capturing facial motion and deformation features, and using manifold alignment and physical continuity constraints to optimize the objective function, further providing reliable support for subsequent reconstruction, effectively maintaining consistency between the reconstruction results and real facial motion.

The consistency of curvature refers to the degree of similarity between the reconstructed results and the real image in terms of curvature properties in image reconstruction. Curvature values are usually used to measure the similarity. The closer the curvature value of the reconstructed results is to the curvature value of the real image, the more the reconstruction algorithm captures the curvature features of the real image and can accurately restore the details of curvature changes. It has good reconstruction performance and can achieve accurate non-rigid 3D image reconstruction. In order to verify the degree of detail restoration of non-rigid 3D images reconstructed using the algorithm proposed in this paper, 8 images were randomly selected from

the Stanford 40 Actions dataset. The curvature values of the reconstructed images using the algorithm proposed in this paper were compared with those obtained from five comparison algorithms based on the consistency index of curvature, as shown in Figure 15.

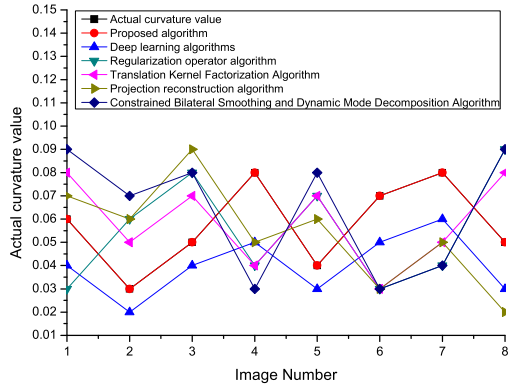


FIGURE 15. Using the algorithm proposed in this article and five comparison algorithms to calculate the curvature value of reconstructed images.

From Figure 15, it can be seen that for any 8 selected images, there is a significant difference between the curvature values of the reconstructed images and the actual images using reconstruction algorithms based on deep learning, regularization operators, translation kernel decomposition, projection reconstruction, constrained bilateral smoothing, and dynamic mode decomposition. The curvature value of the reconstructed image using the proposed algorithm is consistent with the curvature value of the actual image. This demonstrates the effectiveness of the proposed algorithm in improving the degree of detail restoration in 3D image reconstruction, thereby effectively enhancing the quality of image restoration. This is because the proposed algorithm optimizes the construction of the objective function through manifold alignment and physical continuity constraints, making the reconstruction results more in line with the shape and characteristics of real objects. It also uses the L-M nonlinear optimization method to solve the shape based key initial values, better guiding and optimizing the reconstruction process, and improving the degree of detail restoration.

IV. DETAILED DISCUSSION

A. EXPERIMENTAL DESCRIPTION

Throughout the completion of the entire paper, the non-rigid 3D image reconstruction algorithm based on the reliability of the deformable shape priors was subjected to three sets of experiments.

Set 1: This experiment investigates the impact of the number of moving facial images on the algorithm proposed in this paper, using moving facial images as an example. Firstly, generate variable camera internal parameters $f_i = 800 + i$, set distortion factors of 1.1 and 0.5, and calibrate them using the internal parameter matrix $(u_0, v_0) = (320, 240)$; Secondly, randomly generate 100 3D spatial points on a unit

sphere and divide them into 3 rigid body elements; Finally, using different camera extrinsic matrices, generate 20-200 images with size 640×480 , apply 1-pixel Gaussian noise to each image using the IMNOISE function in MATLAB, run 100 times for each number of images, and then calculate the average reprojection error.

Set 2: In the first step, 68 feature points were marked on the moving face using a white marker pen, as shown in Figure 16.



FIGURE 16. Facial feature points marked with a white marker pen in a certain frame of the image sequence.

These marker points will be used as features for extraction, as shown in Figure 16. These points basically cover all areas of the face, including parts such as the eye socket and nose bridge that do not change significantly during facial expression changes, as well as points like the mouth.

A larger number of points were taken at the mouth at the first step, which can more clearly display these points with larger changes during facial expression changes.

The second step is to use the SONY HDR-XR150E-400 megapixels high-speed camera to capture images, which has a resolution of 4.2 million pixels and can accurately extract the coordinates of these feature points. Approximately 10 seconds of facial expression changes were captured, and then 150 frames of images were captured using professional image capture software.

On step 3, MATLAB was used on simulation to extract the 2D coordinates of these feature points for each frame,

$$\text{forming a measurement matrix } \mathbf{W}_{2FP} = \begin{bmatrix} u_{11} & \dots & u_{1p} \\ v_{11} & \dots & v_{1p} \\ \dots & \dots & \dots \\ u_{F1} & \dots & u_{FP} \\ v_{F1} & \dots & v_{FP} \end{bmatrix} \quad (F)$$

is the number of image frames, P is the number of feature points in the image).

On Step 4, a 3D reconstruction process based on variable shape basis reliability was carried out on the measurement matrix. For each feature point, the 3D coordinates were taken in 150 frames of the image. After reconstructing all feature points, a matrix consisting of 68 feature points was formed, each feature point represented by 3D coordinates as a 360×68

$$\text{matrix } \mathbf{TS}_{3FP} = \begin{bmatrix} \mathbf{W}_1^T \\ \mathbf{W}_2^T \\ \dots \\ \mathbf{W}_P^T \end{bmatrix} \quad (\mathbf{W}_i^T \text{ representing 3D coordinates of the point } i \text{ in 150 frames}).$$

Set 3: This experiment describes the performance of non-rigid 3D image reconstruction, with indicators including back projection error, similarity index, and consistency of curvature. Firstly, obtain non-rigid 3D image data and record projection data from different angles. Using the collected projection data, the image is reconstructed using the algorithm proposed in this article and five comparative algorithms. Compare the reconstructed image with the moving facial image. 68 feature points marked on moving faces were loaded into a non-rigid 3D image reconstruction algorithm program based on deformable basis reliability, and experimental research was conducted.

B. RESULTS DISCUSSION

In the first set of experiments, as the number of spatial points and images increases, the amplitude of the reprojection error will show a trend of first increasing and then decreasing. This is mainly because when there are a large number of spatial points and images, the number of equations is more than the unknowns, and the solving process is more overconstrained and stable, so the residual difference is relatively small. So as the number of spatial points and images increases, the residual variation will show a trend of first increasing and then decreasing, and will eventually stabilize [25].

In the second set of experiments, the algorithm proposed in this paper can effectively restore the 3D structure and motion of non-rigid bodies, and the reconstruction results of the selected images are basically consistent with the initial images. This is mainly because when the algorithm in this article trains the motion structure parameters of non-rigid objects, the selected objective function ontology is the image model, and the difference between its 2D coordinates and the image manifold group is the minimum value. When a non-rigid object is in motion, the variation of its motion structure parameters between the front and back frames is very small. This physical continuity can provide constraints for obtaining the parameter matrix of the motion structure. The algorithm proposed in this article constitutes an objective function with high reliability in these two aspects, effectively solving the problem of reliability degradation caused by the continuous change of shape basis in non-rigid motion images, and has higher reconstruction reliability. The algorithm proposed in this article has a good local reconstruction effect. For the local image of the nasal tip, it can obtain the local features of the nasal tip that match the motion of facial images, better reflecting the high nasal tip characteristics of Europeans and Americans. This is mainly because in this article selects the appropriate reconstruction algorithm and the correct parameter selection and optimization, thereby effectively achieving good local reconstruction results. The reconstruction results of the left eye under the constraint of velocity variation in non-rigid body motion in this algorithm are consistent with the reconstruction results of the moving face image, and the relative position of the eye and eyebrow is relatively accurate. This is mainly because the algorithm proposed in this article incorporates physical laws into the motion parameter matrix of non-rigid bodies, reflects them

in the matrix, and obtains a constraint to describe the velocity change in non-rigid body motion. It also obtains the error between the transformed points of the manifold motion group as the minimum value subject and the actual measurement points, as well as two constraints to describe the displacement and velocity change of the parameters before and after the frame during non-rigid body motion, thus, by adding a constraint on velocity variation in non-rigid motion, the deformation of the non-rigid body becomes more continuous, and the position of the reconstructed 3D structure is more accurate [26].

In the third set of experiments, the back projection error of the algorithm proposed in this paper was the smallest, indicating that the low-priced approximation of the measurement matrix obtained by the algorithm proposed in this paper is more accurate. This is mainly because the algorithm proposed in this article effectively controls the complexity of the model and reduces the possibility of overfitting. By retaining the main low rank structures, the essential features of the data can be better described without being affected by excessive noise and details. The SSIM index obtained from non-rigid 3D image reconstruction using this algorithm is close to 1, proving that the difference between the reconstructed image using this algorithm and the moving face image in Figure 3 is smaller. This is mainly because the algorithm used in this article uses nonlinear optimization to solve parameter problems. By selecting appropriate initial values, it can help accelerate the convergence speed of the optimization algorithm and also help avoid getting stuck in local optima. By providing an initial value close to the optimal solution, the similarity between the two images can be improved, thereby reducing the difference in non-rigid 3D images. The curvature value of the reconstructed image in this article is consistent with the actual curvature value of the image, which is effective in improving the degree of detail restoration in 3D image reconstruction and also improving the quality of non-rigid 3D image restoration. This is mainly because the algorithm proposed in this article optimizes the construction of the objective function through manifold alignment and physical continuity constraints, making the reconstruction results more in line with the shape and characteristics of real objects. The L-M nonlinear optimization method is used to solve the shape based key initial values, better guiding and optimizing the reconstruction process, improving the degree of detail restoration, and thus improving image quality [27].

V. CONCLUSION

This paper presents a novel reconstruction algorithm for non-rigid 3D images, leveraging a low-rank matrix derived from image points and depth factors to accurately represent dynamic non-rigid shapes. By incorporating an improved method for defining the nonlinear objective function and selecting initial values, alongside classical nonlinear optimization techniques, the algorithm effectively reconstructs the 3D structure and parameter matrices for non-rigid forms. Its key strengths include optimized objective function and initial value computation, and a precise approach to

transformation matrix calculation, ensuring consistent image treatment. Simulation results confirm the algorithm's high reliability and its success in minimizing back projection errors, highlighting its potential to advance non-rigid 3D image reconstruction.

REFERENCES

- [1] H. Diao, Y. Zhang, L. Ma, and H. Lu, "Similarity reasoning and filtration for image-text matching," in *Proc. AAAI Conf. Artif. Intell.*, 2021, pp. 1218–1226.
- [2] B. Koger, A. Deshpande, J. T. Kerby, J. M. Graving, B. R. Costelloe, and I. D. Couzin, "Quantifying the movement, behaviour and environmental context of group? Living animals using drones and computer vision," *J. Animal Ecol.*, vol. 92, no. 7, pp. 1357–1371, 2023.
- [3] H. Li, Z. Su, N. Li, X. Liu, S. Wang, and Z. Luo, "Non-rigid 3D shape retrieval based on multi-scale graphical image and joint Bayesian," *Comput. Aided Geometric Design*, vol. 81, Aug. 2020, Art. no. 101910.
- [4] Y. Lao, O. Ait-Aider, and A. Bartoli, "Solving rolling shutter 3D vision problems using analogies with non-rigidity," *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 100–122, Jan. 2021.
- [5] H. Zhou, X. Xiang, X. Wang, and W. Ren, "Chained end-to-end pedestrian multi-object tracking network with multi-feature fusion," *Comput. Eng.*, vol. 48, no. 9, pp. 305–313, 2022.
- [6] E. Tretschk, N. Kairanda, B. R. Mallikarjun, R. Dabral, A. Kortylewski, B. Egger, M. Habermann, P. Fua, C. Theobalt, and V. Golyanik, "State of the art in dense monocular non-rigid 3D reconstruction," *Comput. Graph. Forum*, vol. 42, no. 2, pp. 485–520, 2023.
- [7] J. Greffier, D. Dabli, J. Frandon, A. Hamard, A. Belaoui, P. Akesoul, Y. Fuamba, J. Le Roy, B. Guiu, and J. Beregi, "Comparison of two versions of a deep learning image reconstruction algorithm on CT image quality and dose reduction: A phantom study," *Med. Phys.*, vol. 48, no. 10, pp. 5743–5755, Oct. 2021.
- [8] M. Terris, A. Dabbech, C. Tang, and Y. Wiaux, "Image reconstruction algorithms in radio interferometry: From handcrafted to learned regularization denoisers," *Monthly Notices Roy. Astronomical Soc.*, vol. 518, no. 1, pp. 604–622, Jun. 2022.
- [9] J. H. Lin and L. Ma, "An non-rigid online, volume reconstruction method based on energy optimization of dense deformation field," *ACTA Electronica Sinica*, vol. 49, no. 5, pp. 936–943, 2021.
- [10] K. Murase, "Three-dimensional image reconstruction in projection-based magnetic particle imaging," *Jpn. J. Appl. Phys.*, vol. 60, no. 8, Aug. 2021, Art. no. 088001.
- [11] G. Jo, Y. J. Lee, and I. Ojeda-Ruiz, "2D and 3D image reconstruction from slice data based on a constrained bilateral smoothing and dynamic mode decomposition," *Appl. Math. Comput.*, vol. 420, May 2022, Art. no. 126877.
- [12] Z. An, Q. Lin, L. Yang, Y. Guo, and P. Li, "Localizing RFIDs in pixel dimensions," *ACM Trans. Sensor Netw.*, vol. 19, no. 1, pp. 1–24, Feb. 2023.
- [13] S. Yang, H. Chen, F. Xu, S. Zhao, and J. Yuan, "Target tracking algorithm based on Siamese region proposal network for UAV," *Comput. Eng.*, vol. 48, no. 1, pp. 288–304, 2022.
- [14] A. Prades-Valls, J. Corominas, N. Lantada, G. Matas, and M. A. Núñez-Andrés, "Capturing rockfall kinematic and fragmentation parameters using high-speed camera system," *Eng. Geol.*, vol. 302, Jun. 2022, Art. no. 106629.
- [15] S. Liu, S. Huang, X. Xu, J. Lloret, and K. Muhammad, "Efficient visual tracking based on fuzzy inference for intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 15795–15806, Dec. 2023.
- [16] C. Wang, J. Zhang, and M. Li, "Object tracking algorithm combining convolution and transformer," *Comput. Eng.*, vol. 49, no. 4, pp. 281–288, 2023.
- [17] S. Liu, S. Wang, X. Liu, C.-T. Lin, and Z. Lv, "Fuzzy detection aided real-time and robust visual tracking under complex environments," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 1, pp. 90–102, Jan. 2021.
- [18] J. Schino, "Normalized ground states to a cooperative system of Schrodinger equations with generic L^2 -subcritical or L^2 -critical nonlinearity," *Adv. Differ. Equ.*, vol. 27, nos. 7–8, pp. 467–496, 2022.
- [19] I. Harris, "Regularization of the factorization method applied to diffuse optical tomography," *Inverse Problems*, vol. 37, no. 12, Dec. 2021, Art. no. 125010.
- [20] Y. Luo, L. Li, J. Liu, S. Tang, L. Cao, S. Zhang, S. Qiu, and Y. Cao, "A multi-scale image watermarking based on integer wavelet transform and singular value decomposition," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114272.
- [21] M. Zhang, W. Ding, Y. Li, J. Sun, and Z. Liu, "Color image watermarking based on a fast structure-preserving algorithm of quaternion singular value decomposition," *Signal Process.*, vol. 208, Jul. 2023, Art. no. 108971.
- [22] Y. Yu, Y. Zhang, and S. Yuan, "Quaternion-based weighted nuclear norm minimization for color image denoising," *Neurocomputing*, vol. 332, pp. 283–297, Mar. 2019, doi: 10.1016/j.neucom.2018.12.034.
- [23] L. Hruđa, I. Kolingerová, M. Lávička, and M. Manák, "Rotational symmetry detection in 3D using reflectional symmetry candidates and quaternion-based rotation parameterization," *Comput. Aided Geometric Design*, vol. 98, Oct. 2022, Art. no. 102138, doi: 10.1016/j.cagd.2022.102138.
- [24] S. Liu, S. Huang, W. Fu, and J. C.-W. Lin, "A descriptive human visual cognitive strategy using graph neural network for facial expression recognition," *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 1, pp. 19–35, Jan. 2024.
- [25] S. Liu, X. Xu, Y. Zhang, K. Muhammad, and W. Fu, "A reliable sample selection strategy for weakly supervised visual tracking," *IEEE Trans. Rel.*, vol. 72, no. 1, pp. 15–26, Mar. 2023.
- [26] W. Munir, A. M. Siddiqui, M. Imran, I. Tauqir, N. Zulfqar, W. Iqbal, and A. Ahmad, "Background subtraction in videos using LRMF and CWM algorithm," *J. Real-Time Image Process.*, vol. 18, no. 4, pp. 1195–1206, Aug. 2021.
- [27] Z. Nazih, M. S. Adil, A. Awais, I. Waseem, M. Imran, I. Tauqir, and M. Wajiha, "An improved multiexposure image fusion technique," *Big Data*, vol. 11, no. 3, pp. 215–224, Jun. 2023.



HAIYING CHEN was born in Xinyang, Henan, in 1976. She received the B.S. degree from Henan Normal University, in 1999, and the M.S. degree from Central China Normal University, in 2008.

From 1999 to 2005, she was a Mathematics Teacher with Xinyang Vocational and Technical College; From 2008 to 2018, she was a Mathematics Teacher with Wuhan Design and Engineering College; and From 2019 to 2020, she was a Mathematics Teacher with Xianning Vocational and Technical College. Since 2021, she has been a Mathematics Teacher with Tianmen Vocational College. She has authored five books and more than 20 articles. Her research interests include probability theory and mathematical statistics, and applied mathematics.

Prof. Chen won the third prize of the 8th Hubei Provincial Teaching Achievement Award.



SYED ATIF MOQURRAB received the master's degree from the National University of Computer and Emerging Sciences and the Ph.D. degree from the COMSATS Institute of Information Technology. He is currently a Distinguished Academician, an Educator, and a Researcher, has made significant contributions to the fields of computer science and artificial intelligence. As a Research Assistant Professor with Gachon University, South Korea, and formerly an Assistant Professor with Air

University, Pakistan, he has showcased expertise in a broad spectrum of subjects. His extensive publication record features innovative solutions for preserving privacy, advanced data analytics, and deep learning models. His research has focused on critical aspects of privacy and security in various domains.

...