IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 20, 2020, accepted July 2, 2020, date of publication July 8, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007897

Construction of Virtual Video Scene and Its Visualization During Sports Training

RUI YUAN[1,2], ZHENDONG ZHANG[1], PENGWEI SONG[1,4,5], JIA ZHANG[3], AND LONG QIN[4,5]

[1]School of Physical Education, Zhengzhou University, Zhengzhou 450001, China
[2]Marine Sports, Pukyong National University, Busan 48513, South Korea
[3]Division of Physical Education, Chung-Ang University, Seoul 06874, South Korea
[4]Division of Physical Education, Keimyung University, Daegu 42601, South Korea
[5]Division of Physical Education, Guangxi Science and Technology Normal University, Guangxi 549199, China

Corresponding author: Zhendong Zhang (zzd0506@zzu.edu.cn)

This work was supported in part by the Philosophy and Social Science Planning Project of Henan Province (Research on the Development Strategy of Leisure Sports in Henan Province in the Process of Building a Well-off Society in an All-round Way) under Project 2019BTY008, and in part by the Humanities and Social Sciences Project of Henan Provincial Department of Education (Research on the Trinity Linkage Mechanism of College Football in Henan Province from the Perspective of Ecology) under Project 2020-ZZJH-455.

ABSTRACT This article studies the actual captured human motion data for human motion synthesis and style transfer, constructs a scene of motion virtual video, and attempts to directly generate human motion style video to establish a sports style transfer model that combines and self-encoding. The original human motion capture data mapped to the motion feature space for style transfer synthesis. The coding network used to map the high-dimensional motion capture data to the low-dimensional feature space, and the motion style transfer constraints established in the feature space, and the human body motion results after the style transfer obtained by decoding. This paper proposes a pixel-level human motion style transfer model based on conditional adversarial networks and uses convolution and convolution to establish two branch coding networks to extract the features of the input style video and content pictures. The decoding network decodes the combined two features and generates a human motion video data frame by frame. The Gram matrix establishes constraints on the encoding and decoding features, controls the movement style of the human body, and finally realizes the visualization of the movement process. The incremental learning method based on the cascade network can improve the high accuracy and achieve the posture measurement frequency of 200Hz. The research results provide a key foundation for improving the immersion sensation of sport visual and tactile interaction simulation.

INDEX TERMS Virtual video, scene construction, movement process, visualization.

I. INTRODUCTION

Motion capture technology is known as dynamic capture. The motion capture devices used to record the movement of people or objects. It widely used in motion science, virtual reality, animation, and video games. In the early production of movies and animations, the movement of multiple camera characters required, and the character data shot is difficult to reuse [1]. Using motion capture technology to record the motion data of the character's character, and use the computer for processing, it can achieve multiple reuses [2]. Because of the high efficiency of motion capture technology, it can not only realize motion synthesis but also create new motions according to the designer's ideas. In the video game, you can use motion capture technology to capture human motion data to build athletes, martial artists, and other game character models [3]. For example, issued arcade games use passive optical system tags to collect human motion data, and adventure games use motion capture. The device collects human movements and completes the character and plot production of the characters in the game [4]. In film production, the use of motion capture technology can improve the efficiency of film production. In science fiction movies, computer-generated virtual characters used to replace characters in traditional animation. In motion science, motion capture technology used to achieve gait analysis, tracking, and calibration of the human body's whole body, lower body and feet, and other areas, providing researchers and biokinematicians with accurate 3D

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

tracking data to facilitate [5]. The human body researches have various factors of biomechanics. In the field of virtual reality and augmented reality, the use of motion capture devices to model the motion behavior of human bodies or objects allows users to interact with virtual scenes in real-time through their actions, which is very useful for visual perception or training simulation in virtual scenes [6]–[10]. These application technologies use motion capture equipment to digitally collect human motion, and then use computer technology to synthesize and render the collected motion data. Similar to human motion capture devices, these facial motion capture devices are usually composed of multiple joints and rigid links. Angle sensors are installed in the joints and fixed at the mouth, eyes, etc. of the person. When motion occurs, the angle sensor can measure the change in angle and calculate the position and trajectory of the fixed point in space based on the length of the connecting rod.

Human motion synthesis is the use of data-driven and constrained control methods to post-process motion capture data to synthesize high-quality motion sequences to apply the motion sequence to multiple scenes, increase the multiplexing capability of motion data, and provide interactive control way to edit sports characters [7]. The main methods of motion editing include animator production, inverse kinematics methods, variation methods, data-driven methods, and statistical methods. The method created by the animator requires high technical requirements and consumes a lot of time [11]. The data-driven method uses machine learning methods to build models and uses the captured human motion data to train the model, and controls human motion in a data-driven manner [12]. Statistical methods process motion data through motion mixing or dimensionality reduction methods, and perform timing calibration and mixing in motion to generate new human motion. Bulgarian uses the motion parameters as sampling signals to establish the mapping relationship between different motions, uses multi-resolution filtering to deal with the motion parameter problems, and then uses the multi-target interpolation technology based on the principle of dynamic time warping algorithm to achieve human motion mixing [13]. We used in the Fourier series expansion method to analyze human motion capture data set that needs to be measured, the human motion data into the time domain and frequency domain, and then interpolated different motions in the frequency domain to obtain new motion. Users were allowed to perform interactive editing and real-time control [14]. It is difficult to process more complex movements by directly synthesizing new sport through signal processing, and it is difficult to obtain better results for sports data with large differences in sports styles [15].

It used a set of motion capture data to the human body captured by the camera to build a local model and then used this model to generate a character with full-body animation [16]. This method uses two cameras and a reflector to achieve different behaviors in real-time. M. Di Renzo also used a local model to study human motion synthesis. Based on the sparse accelerator input, the cross-domain search method used to build an inert neighborhood map online [17]. The independent component analysis method is also a common method for data-driven synthesis of human motion. Motions divided motion data into different visual components, and migrated its style components from one motion to one motion [18]. The above methods of using linear mixing to synthesize new human movements have certain limitations which require cumbersome data preprocessing [19]. We can only synthesize relatively simple human movements and the output results need to be further processed manually before they can be used in movies and computers Animation and other real applications [20]. The motion data established a relationship between different motions through inverse kinematics and radial basis functions to generate new motions in real-time motion interpolation [20]. This kind of motion synthesis method combining inverse kinematics and method not only satisfies the characteristics of actual human emotion but also can capture the characteristics of complex motion by learning the motion parameters. It conducted further research on this method [21]. Firstly, the basic degree function used to interpolate the degrees of freedom of each motion with the interpolation of each example to reduce the number of interpolations. The motion space designed by the division interpolated to obtain new motion [22]. Because the method lacks a mechanism to deal with data noise and fluctuations, which make it prone to overfitting, the Gaussian process method has also proposed to mix different types of motion. The study used the method to fit the model's motion data, and analyzed the correlation between motion sequences, and generated new motion by interpolation [23]. The researchers used a Gaussian process latent variable model to map high-dimensional motion data to a low-dimensional space and combined with inverse kinematics methods to directly control the moving characters in the low-dimensional space [24]. Similarly, it also used the method to model the timing characteristics of motion data, and used historical motion sequences to learn the motion posture of the next frame [25]. Because of the synthesis and control of moving characters in different venue environments, some researchers studied the gait of game animation characters and proposed a cyclic motion para metritis data structure for synthesizing cyclically changing movements [26]. It can control the gain of human movement through arbitrary terrain. However, these kernel-based methods often require huge memory consumption and meet real-time applications, and the actual application scenarios have certain limitations [27].

The main reason for the difficulty of synthesizing motion gestures is that the generation of the next motion state tends to take an average [28]. For example, when the prediction of a gesture may be left or right, two possible averages may be obtained finally, causing left and right swings [29]. These problems are that the rotation and translation information of the input control signal is not considered in the model training, resulting in the model preferring to predict the time series rather than synthesizing the attitude, so the input of the control signal for interactive motion synthesis is very important [30]. Therefore, this paper first establishes a self-encoding generation network pre-trained using results-based management (RBM) and attempts to directly synthesize the gesture of another move into the sequence of the control signal from directly inputting the motion sequence as the control signal [31]. Then, concerning the time sequence of the motion sequence, this paper takes the motion sequence as input, so that the model can extract time-series features [32]. Secondly, because of the problem that the abstraction of sport style leads to the difficulty of defining the output of sport style, this paper combines technology and cyclic constraints to achieve end-to-end movement sequence style transfer. Finally, for the generation of human motion in real application scenarios, this paper combines convolution and long short-term memory (LSTM) to establish a style transfer model to directly synthesize human motion styles on video.

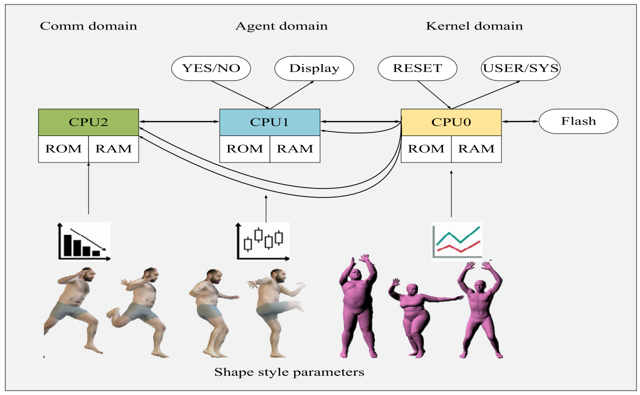

FIGURE 1. The analytical framework of human motion synthesis and style migration database.

II. ANALYSIS OF VIRTUAL SPORTS VIDEO SCENE DESIGN

A. ANALYSIS OF HUMAN MOTION SYNTHESIS AND STYLE TRANSFER DATABASE

The motion capture database realizes the recording of human motion fragments through motion capture devices so that it can be simulated and simulated using devices such as computers. In scenes film, the animation needs to create a virtual persona, the motion capture by the number of databases of reuse, can greatly reduce production costs, and the use of motion capture technology can be synthesized many action scenes realistic attitude can not be achieved. The reasonable use of the motion capture database helps to improve the ability of the model. In this study, we mainly use the motion capture data set for experiments and try to apply the video data directly to the real video scene. Therefore, this chapter mainly introduces human motion capture database and simple human video database, as well as data preprocessing methods.

Motion capture technology is to use motion capture equipment to collect human body motion data of real scenes and apply the collected motion capture data to a movie, animation, and other application scenarios. The collected motion capture data need to further processed to realize motion capture data reuse [33]. Motion capture databases included Carnegie Mellon University's motion database and Human. Cumorah consists of sports data composed of major categories and subcategories, including different sport types such as running, walking, jumping, and different sports venues such as playgrounds and obstacles. Human is widely used in motion prediction with 11 human characters and 17 motion scenes, and also provides real motion images corresponding to each motion posture, as shown in Figure 1. The format directly describes the movement through the three-dimensional coordinate method of storing the movement data, which is convenient for the user's intuitive understanding. RBM is through the LSTM showing skeletal motion and files LSTM data representing motion to describe each of the motion, since the motion capture data acquired for a plurality of data of different composition or other objects, resulting in different movements that may exist between the same sports skeleton and a different sports skeleton. Therefore, RBM can be used to flexibly represent different sports data. It is convenient for us to redirect the motion based on the bone information about the motion and realize the multiplexing of the motion. A common motion capture data representation format is Bah, which adds the hierarchical information of the motion skeleton to the motion data, and calculates the child node information according to the displacement of the parent node. Bah is also composed of two parts. The top represents the skeleton of the movement and the initial state of the movement in a hierarchical manner, and the bottom data section includes the entire movement sequence. Hierarchy represents the start of skeleton data, after root is the name of the root node, after offset is the translation information of each node, after channels is the channel information of the data, after joint is the name of the child node [34]. End site represents the end information of the current node. Motion indicates the start of motion data. Frames are the number of motion frames, and frame time indicates the motion sampling rate. The original motion capture data used in this article is in LSTM format, and the tool provided by cap can convert the RBM format to LSTM.

This paper is mainly used for cap motion capture database study. The first initial RBM converted the data format of the motion LSTM format, and the LSTM data format conversed to 3D representation, and reduce the number of joints. The original joint point has a total of 31 nodes. The name of each node is shown in Table 1. Among them, there are multiple nodes at the joints such as feet and hands. In practice, only one node can be used to express the overall style of movement. Redundant points reduced the complexity of training data. Therefore, the nodes used to represent the motion sequence of each frame.

Because different types of human motion capture data use different human skeletons, the different skeletons in the training data cause the positions corresponding to each joint to be different, which makes the model difficult to train, so the original LSTM motion capture data needs to be mapped to the same skeleton. Yamani uses methods to model the

TABLE 1. Complete human skeleton model description table.

Serial number	1	2	3	4
description	Root	Upper back	Lower back	Thorax
Serial number	6	7	8	9
Skeleton	hand	L-hand	R-wrist	L-wrist
Serial number	11	12	13	14
Skeleton	fingers	R-fingers	L-fingers	R-fingers
Skeleton	fingers	R-fingers	L-fingers	R-fingers

position association, joint angle error, and joint motion distance of motion capture data, and maps motion capture data from different skeletons to the same skeleton through inverse kinematics. The pretreatment method converted all nodes of 3-dimensional coordinates.

Because it is difficult to directly apply motion capture data to real application scenarios, this article attempts to directly synthesize the human body style in the image video, and transfers the style of the video data to the human body in the image. It lets the human body in the image make similar body style movements in the video. The data of represented by three -dimensional nodes, and the human body image corresponding to each pose processed to the pixels of the image. Image and video data such as LSTM and RBM processed in the same way, and each frame of the image input of the model processed. An inverse operation can restore the final generation result of the model.

B. MULTITASK FEATURE MINING WITH AN ATTENTION MECHANISM

The target tracking task processes the tracking video frame by frame, so the single input of the tracking model is a picture. The mainstream idea of the current tracking method is Tracking-by-Detection. Comparing the similarity between the target and the search area, the twin network is an excellent structure, which can extract the generalized feature expression of the target and the tracking image under the condition of satisfying real-time. As shown in Figure 2, this paper designs a complete tracking model structure based on the twin network structure.

In the model designed in this paper, the template image and the tracking image extracted by a pair of twin convolutional neural networks with shared parameters. We put the output elements into the candidate area generation network for target classification and coordinate regression. Among them, the branch for template image feature extraction can be regarded as a twin sub-network, which can be divided into a target image branch and a template frame image branch. The target image is a part of the template frame image, which calculated by the network here. The purpose is to make full use of the coordinate monitoring information of the target in the template frame to guide the subsequent candidate regions to generate the regression of the network to the coordinate frame. The features extracted by convolutional neural network (CNN) of the target image will be updated once based on its coordinate information in the template frame. The updated parameters used candidate regions to generate the convolution kernel parameters in the network regression branch, and the structure of the classification branch is like Siamese-RPN. It classified each candidate frame according to the characteristics of the target image, select the positive sample with the highest similarity to the target as the classification result, and combine with the coordinate frame calculation of the regression branch to obtain the tracking structure of the current frame.

The features extracted from the template frame and the search frame through the above-mentioned convolutional neural network will go through a layer of convolutional layers for channel enhancement, and then input into the candidate area generation network for coordinate frame calculation. Wherein the direct branch target classification characteristic spectrum and the search image cross-correlation, the regression branch target feature as a convolution pattern further required following the actual value of the size of its coordinates a secondary gradient descent, updated by the regression parameters branch for calculation. Generating in the candidate area network, the network search for each pixel

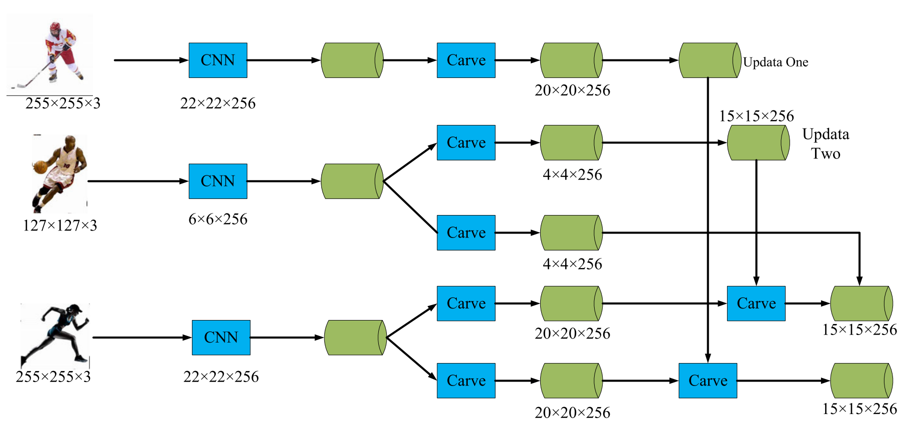

FIGURE 2. The structure of a twin network tracking model based on meta-learning.

area provided to enhance the coordinate frame calculation accuracy. Each anchor assumed that the a priori coordinate frame is the same as a reference. The different prior frames of an anchor point have different aspect ratios, which provide a reference for tracking the possible shape of the target, making the tracking model robust to the shape change of the target. At the same time, in the regression branch, the network only needs to calculate the relative coordinate frame. Due to the offset of the a priori frame without direct regression of coordinates, the output range constrained to any value from the search area to a normal distribution with an average value close to 0. When the predicted value is not much different from the true value, the adjustment process can be regarded as a linear process, and good results can be obtained through regression algorithm modeling, and the network is easier to train.

The priori box is the probability of the positive samples, i.e., the output of each block before PI. When the a priori frame is a positive sample, the criterion for determining is a positive sample which is the a priori frame y. Real coordinate frame is larger than a predetermined threshold value. It is less than a predetermined threshold value that is determined as negative samples, using the formula can be expressed as:

$$\min_{P} \sum_{i=1}^{n} \frac{1}{2} \left\| y_i - X_{P_i} \right\|_2^2 1 + \chi \left\| P \right\|_{2,1} \tag{1}$$

We can transform the problem into a binary classification problem. The commonly used loss function is the cross-entropy function:

$$\overline{w_{i,j}} = \begin{cases} \sin(w_{i,j})(\left|w_{i,j}\right| - \lambda), & \left|w_{i,j}\right| \gg \lambda \\ 0, & \left|w_{i,j}\right| < \lambda \end{cases} \tag{2}$$

Regression output branches contain 4k a channel vector, the output of each frame before [dX, day, dW, dH], which represented of the prior frame relative to the frame offset from the actual coordinate. At the time, we can regard the coordinate calculation as linear regression, so the loss function between the parameter and the true value uses SmoothL1 loss.

$$\overline{h_{i,j}} = \begin{cases} h_{i,j}, & \left|h_{i,j}\right| \gg \lambda \\ 0, & \left|h_{i,j}\right| < \lambda \end{cases} \tag{3}$$

There are four regressions in the regression branch, so the loss function of the regression branch is the sum of the SmoothL losses of the four variables:

$$\sigma = \frac{median(|di(j)|)}{0.625} \tag{4}$$

If we assume a priori frame width and height coordinates, coordinates and real coordinates of the frame width and height, then the formula (5) in Standardized is:

$$\begin{aligned} M_L(\mathrm{w}) &= \frac{1}{2}\sum_{i,j} \mathrm{P}_{ij}(\mathrm{w}^T x_i - \mathrm{w}^T x_j)^2 \\ &= \frac{1}{2}\sum_{i,j} \mathrm{P}_{ij}(\mathrm{w}^T x_i - \mathrm{w}^T x_j)(\mathrm{w}^T x_i - \mathrm{w}^T x_j) \\ &= \frac{1}{2}\mathrm{w}^T(\sum_{i,j} \mathrm{P}_{ij}(x_i - x_j)(x_i - x_j)^T)^{w} \\ &= \frac{1}{2}\mathrm{w}^T(\sum_{i,j}\mathrm{P}_{ij}(\mathrm{x}_i x_i^T) + \sum_{i,j}\mathrm{P}_{ij}(\mathrm{x}_j x_j^T) - 2\sum_{i,j}\mathrm{P}_{ij}(\mathrm{x}_i x_j^T))^{w} \\ &= \frac{1}{2}\mathrm{w}^T(2XDX^T - 2\mathrm{XSX}^{\mathrm{T}})^{\mathrm{w}} \\ &= \mathrm{w}^T X(\mathrm{D} - S)\mathrm{X}^{\mathrm{Tw}} \end{aligned} \tag{5}$$

After designing the respective loss functions of the classification and regression branches, we can obtain the final loss function of the optimizer, that is, the weighted sum of the classification loss and the regression loss:

$$M_N(w) = \frac{1}{2}\sum_{i,j}(1-\mathrm{P}_{ij})^2(\mathrm{m}_i - \mathrm{m}_j) \tag{6}$$

In the tracking task, we want to train a good performance tracking model, the need for video datasets tagged for training and verification. We denote the video data set a segment separate tagged video data is M, then the model may be desired by the subsequent learning information in the first frame. It can be a good fitting frame marked similar coordinate frame.

$$tr(w^M XL_N \mathrm{X}^{\mathrm{Mw}}) = tr(w^M XL\mathrm{X}^{\mathrm{Mw}}) \tag{7}$$

The loss functions between the result obtained in the model and the real annotation can be minimized. This parameter is the optimal value we hope to achieve through training. The tracking process of the tracking model for the video can be regarded as a One-Shot-Learning task. The first frame of data is the only training sample. The purpose of the meta-learning offline training is to get an optimizer, making the video first frame of samples as input to the optimizer, to obtain the desired optimum parameters. The formula can be expressed as word when the parameter of the optimizer φ satisfies the formula (7), the parameter θ as the output of the optimizer satisfies the formula (8):

$$M_N(w) = \frac{M_N(w)}{M_L(w)} = \frac{tr(\mathrm{w}^T XL\mathrm{X}^{\mathrm{Tw}} + \mathrm{w}^T X_l \mathrm{X}_l^T \mathrm{X}^{\mathrm{Tw}})}{tr(\mathrm{w}^T XL_N\mathrm{X}^{\mathrm{Tw}} + \mathrm{w}^T X_b \mathrm{X}_b^T \mathrm{X}^{\mathrm{Tw}})} \tag{8}$$

According to the usual method of updating the model parameters deep learning metalworking thinking, we use the model of independent learning element related methods, optimal parameters obtained by gradient descent, so that the parameter update mode is $\alpha\nabla\theta L$, where α is the learning rate. Optimization on the whole training loss function can be described according to frame sample. To satisfy the subsequent frame sample, the method of gradient descent the process of parameter updating is:

$$\varphi(t) = \sqrt{2}\sum_{n} p_1(n)\varphi(2t - n) \tag{9}$$

The loss function calculated during actual training is:

$$M_N(w) = w^M XL_N \mathrm{X}^{M\mathrm{w}} \tag{10}$$

The updated iteration process has undergone two gradient descent optimizations. We call them internal training and external training. The learning rates of the two layers of training optimization are and β, respectively. The purpose of internal optimization is to calculate the temporary parameters for evaluating the current video test frame. This parameter only used for the current video. The temporary parameters of each video are not related. The purpose of external optimization is to train and update the final parameters of the model. In actual training, to get better results, the number of internal optimization updates usually increased appropriately, but not too much. Generally, a few updates can get better results, and the increase in the number of times will not significantly improve. Instead, it will increase the amount of calculation. In the process of updating external parameters, the specific optimization methods of parameters usually extended to more advanced optimization methods based on gradient descent to achieve training stable, such as learning rate decay, adding momentum, and other techniques.

III. VISUAL DESIGN ANALYSIS OF SPORTS TRAINING PROCESS

A. ANALYSIS OF MOTION TRAINING PROCESS BASED ON DEEP SELF-CODING AND SPATIOTEMPORAL CONSTRAINTS

Using the method of deep learning can extract abstract motion capture data features well, and establish appropriate constraints on the motion features to achieve a synthesis between different motions. Through the extraction of time-series features of human motion, the quality of motion synthesis further improved. Using the motion sequence of the current frame of the motion capture data, a motion style transfer model based on deep self-coding and spatio-temporal features established. The model is also based on self-encoding. Using self-encoding, the data can bedecoded, which is convenient for mapping the original motion data to the feature space for motion synthesis. At the same time, the structure of the self-encoding adjusted, and the input result of the past motion sequence added to the self-encoding network so that it can learn the timing characteristics of the original motion data.

The process of the sports style transfer model established in this paper is shown in Figure 3. The entire style transfer model consists of three parts, that is, RBM pre-training network, sport style extraction network, and sport style transfer network. Among them, RBM pre-training is like the model. To extract the time series feature of human motion, motion sequence added to the RBM pre-training network. The motion style extraction network can be divided into style feature extraction and motion feature extraction. Both networks use fast motion sequences to extract motion timing features, the parameters of the two networks shared. The sport style transfer network is to establish the movement synthesis constraint and use the movement characteristics obtained by the movement style extraction network to perform the movement style transfer synthesis. When performing sport style transfer synthesis, firstly pre-process the original motion capture data, and randomly divide it into style sport and

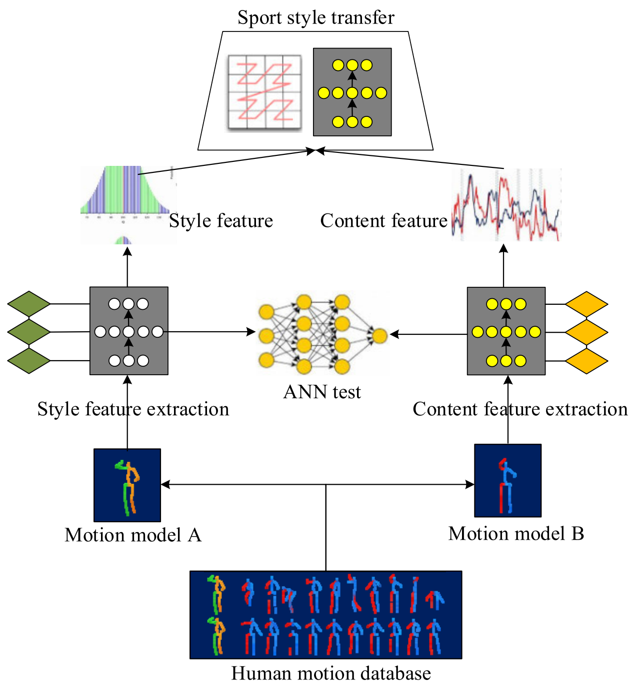

FIGURE 3. Flow chart of the motion training process based on deep self-coding and Spatio-temporal features.

content sport during training. At the same time, all the motion capture data used to train the RBM pre-training model, and the pre-trained parameters used to initialize the feature extraction network. Then, the style extraction network used to extract the feature motion and content motion respectively, and the high-dimensional motion data mapped to the low-dimensional feature space. Finally, obtained sport style features used to establish constraints, and the sport style transfer realized through the sport style transfer network.

In the sport style transfer model combining RBM and self-encoding, each frame of the entire motion sequence directly input as one by one. This method is convenient to operate and can extract the characteristics of each frame of motion, but the model does not consider the motion sequence. Timing relationship, there is a large dependency between each frame of motion, and this dependency also reflects the consistency and naturalness of the motion data. Therefore, this paper proposes a self-encoding feature extraction network that incorporates a time series. This network can not only extract the spatial features of motion data but also express the relationship between each frame of the motion sequence.

The main structure of the self-encoding feature extraction network incorporating time series is shown in Figure 4. The input of the network consists of two parts, the current motion T and the previous motion t. The main network transformed from self-encoding, and the past motion sequence added as an input to the self-encoding network structure. Current motion of a single frame inputs the frame motion to the current motion before t frames motion data. The network model as a sequence of movements in the past, the network

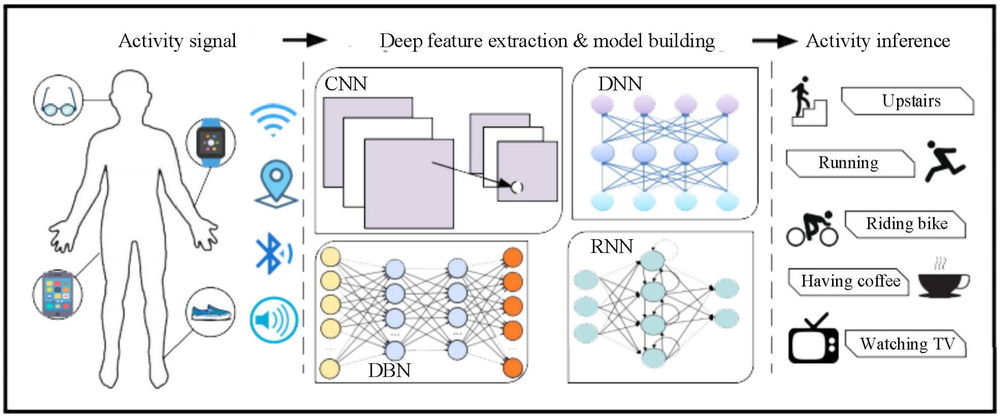

FIGURE 4. Self-coding feature extraction network with time series.

can be extracted to the motion. The entire network model can be divided into an encoding stage and a decoding stage.

In the encoding stage of the network, network maps the input original motion data to the feature space, and finally obtains the function of the network in the encoding stage according to the input current motion. Using the past motion sequence of the current motion, a self-encoding feature extraction network fused with time sequence constructed to map the original motion capture data to the feature space, and the spatial structure and time correlation of the original motion data retained. Aiming at the generation of human sport style transfer, the extracted motion features used to establish style transfer constraints to realize the movement style transfer model, and further transfer the style of one game to another game. The main structure of the sport style migration network with spatiotemporal feature constraints proposed is shown in the Figure 4. The overall network structure is similar to self-encoding. In the encoding and decoding stages of self-encoding, the past motion sequences added as input, so that the network can learn the motion for the timing characteristics of the data. As shown in Figure 4, the feature extraction network can be used to obtain the content features and style features of the original motion data. The content features, style features, and hidden layer features of the style migration network used to form a sports style migration constraint, and the constraint used to transfer the style. The model updated to achieve the transfer of sport style.

B. ANALYSIS OF VISUAL PROCESSING OF MOTION

The flow of the sport style transfer model based on deep self-encoding and time-series features is shown. First, it uses the self-encoding feature extraction network that combines time-series features to obtain the motion features of the two original motion data inputs, and then use the obtained features to construct style transfer constraints. We transfer the style of one sport to another through the sport style transfer model.

In the visual-inertial fusion poses measurement process, the frequency of the inertial data is higher than the frequency of the tag poses data, which makes it impossible to establish a direct time series prediction model of high-frequency inertial data and low-frequency pole data. To solve this problem, an incremental learning method based on cascaded networks proposed, which estimates the increment of time series variables in a short period based on high-frequency input, and optimizes network parameters using real values of long periods to gradually learn the increase in time series variables. To achieve a high-frequency estimation of variables, when the operator interacts with the virtual scene using magnetically operated instruments, the binocular camera collects images of the scene, the convolutional neural network captures the acceleration and angular velocity data of the magnetically interactive instruments. The visual control unit calculates the position information of the magnetic interactive instrument, and the inertial control unit calculates its posture information. The calculated pose information, combined with the original acceleration and angular velocity data, used by the server to perform high-frequency estimation of the pose information of the magnetic interaction device. After the server uses the visual-inertial fusion pose measurement method proposed in this chapter to complete the update of the pose information. According to the pose measurement information, the corresponding feedback force applied to the magnetic interactive device and each of the coil array calculated. The motion visualization interface intelligently adjusts the coil current to generate the required effective magnetic field. The magnetic interaction equipment receives the feedback force in this magnetic field and transmits it to the operator. Besides, the virtual scene is visually updated, and finally, the user experiences visual-tactile interaction, as shown in Figure 5.

In video scenes, the target features will undergo various changes over time, including various conditions such as deformation, rotation, occlusion, scale transformation, etc. Therefore, the features of the target in the current frame will have a certain degree of difference from the initial features. The degree may increase over time. In the visual target tracking task, the model often updates the observation model online to adapt to the changes of the target during the tracking process and maintain the discriminative ability of the model.

TABLE 2. Test results of various tracking algorithms on OTB.

Algorithms	This paper	Muster	SimeseFC	MEEM	STRUCK	C-COT	MDNet	GOTURN	TGPR	KCF
OTB-CVPR	0.456	0.532	0.782	0.653	0.833	0.675	0.781	0.972	0.782	0.667
OPB-100	0.570	0.673	0.667	0.764	0.731	0.539	0.588	0.759	0.781	0.664
Real time (YES/NO)	YES	NO	YES	NO	NO	YES	NO	YES	NO	YES

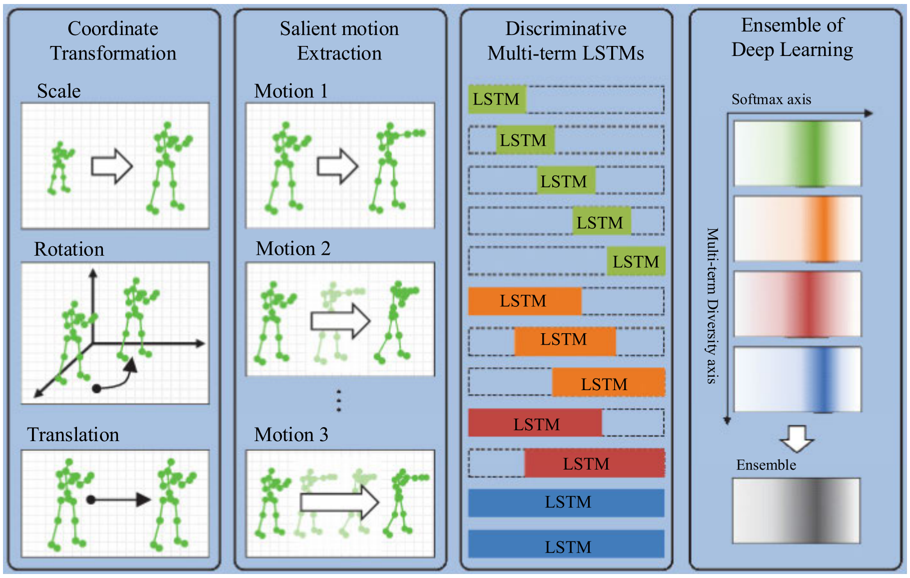

FIGURE 5. Motion visualization design.

The target tracking algorithm based on template matching; model updating is a relatively common setting. At present, there is no unified standard and the method of updating the model every other frame is usually used. However, in many twin network structure models including Siamese-FC and Siamese-RPN, generally, only the characteristics of the first frame used for matching. If the tracking results used to update the parameters, they can adapt to the changes of the target representation to a certain extent. The continuous update will introduce more and more noise, that is, when the result is in error, it will be continuously amplified, so that the model is gradually updated in the wrong direction, and eventually the target is completely lost and the model drifts. At the time of the common natural objects tracking, tracing algorithm twin's network structure-based approach does not use an online update to satisfy the stability in most cases with the track.

The lighting effect of each triangle mesh is determined by the normal vector of the plane and the normal vector of the vertices, so it is necessary to calculate the plane normal vector and the vertex normal vector of the three-dimensional model of the human body. The augmented reality system registers virtual objects to enhance the information. If a user wants to interact with a computer-generated virtual object, the human body model and the virtual object registered in the same one. At this time, the registered human body model regarded as a virtual object, so that the talent can perceive the depth of the virtual object in the virtual object. To embody real interaction, human models and virtual objects simulated and fused with the real height, and they must satisfy the properties of inertia and collision, material, and quality.

Although the three-dimensional interface with different distances can adjust the physical size to balance the screen size, the experience will still be different when browsing. The movement change rate of the 3D interface with different distances will be different. When the mobile phone moves back and forth or left and right, the change rate of the 3D interface on the screen is different. When the phone moves back and forth (close to or away from the 3D interface), the screen size will change, zoom in when close, and become smaller when away. The change rate of the interface at different distances is different. The closer the distance of the three-dimensional interface is the greater the rate of change of movement. As shown in Figure 5, for the farther interface, the number of icons changes less after the mobile phone moves (from 8 to 6), and the number of icons near the interface changes more (from 8 to 4).

When tracking common natural objects, the tracking algorithm based on the twin network structure does not adopt the online update method to meet the stable tracking in most cases, thereby avoiding the model drift caused by the frame-by-frame update. However, after analyzing and testing the text video, we found that in the data set, the text target very frequently appears a part of the target of drifting out of the picture, and the next target is usually very narrow and long. When most of the target image lost, its characteristics will be a large change from the initial time caused the tracking effect to decrease. To solve the above problems, we need to update the model without drifting to ensure the stability of the online update. Therefore, we cannot use the traditional method of updating frame by frame based on the tracking results, but choose to model the model in time series. It uses a combination of long and short-term updates, and comprehensively uses the information of the first few frames of the current tracking frame to update.

In the training phase, the network still uses the meta-learning update strategy to train the parameters, but to ensure the uniformity of time series modeling, the random network frame extraction method commonly used in the twin network structure model can no longer be used for training but is optimized internally. To extract the image sequence calculated, and every frame parameter of a batch update module updates the line. The line loss function module updated before computing the characteristic features of the target template and the second frame obtained with labeled features. In the tracking phase, the model will retain the processed template features as a small batch of data with a capacity of frames of the target template's feature map. When the saved data exceeds the capacity, the oldest test frame data will be discarded and the remaining frames of the template The data is used as the input of the long and short-term memory network in the online update module, and the output feature map is used as the template feature for the search of the target. The entire calculation process is the feedforward propagation of the network without involving parameter updates, so it does not affect the tracking model. Real-time has too much impact.

IV. RESULTS AND ANALYSIS

A. COMPARISON OF TRAINING RESULTS DURING EXERCISE

This article uses the input content movement as right-turn walking, and the style movement as standing to imitate monkey movement and zombie movement, and then conduct style transfer tests on using past movement sequences and not using past movement sequences, which further proves that the past movement sequences used in this article can be The model can deal with timing features, and the results are shown in Figure 6. Without using a motion sequence in the past, the style of the input motion standing monkey generated by monkeys turn right to go to small changes in the trajector. If the past motion sequence is used, the generated monkey trajectory to turn right is similar to the input trajectory. For zombies-style motion input, if the past motion sequence not used, the trajectory of the generated zombie walking right is random. If the past motion sequence is used, the trajectory of the generated zombie's right turn is similar to the input trajectory. Finally, to test the effect of past motion sequences on the generation of motion slips, simply judge by

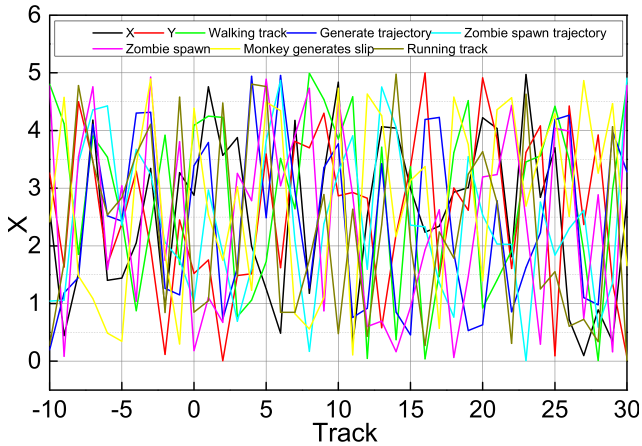

FIGURE 6. Test results of past motion sequences.

the position difference between the previous frame and the next frame of the motion. If a forward motion, the difference between the back frame and the previous frame is less than 0, it means the slippage phenomenon occurs, and the results are shown in Figure 6. When the past motion sequence is used, the slippage phenomenon is significantly less than that without the past motion sequence.

The parameters of the twin feature extraction network are initialized by pre-training on the convolutional neural network dataset. CNN is a video data set used for visual target tracking tasks. Its video features are highly diverse, including a total of 563 target categories and 87 active modes in real life, when compared to CNN. CNN-100 is currently the most abundant video data set. As training data, it can greatly improve the robustness of the model and reduce the risk of overfitting during training. During the test, the first frame image of the current test video used as the template frame, and similarity matching and coordinate regression performed with all subsequent frames to be tracked. The output result evaluated with the calculation overlap in the test set. This design based on meta-learning twin web tracking algorithm is mainly in the evaluated on the dataset by plotting overlap at different thresholds. Indicators derived algorithm tracking performance evaluation system. The proposed tracking algorithm tested on 100, and the visualization results are shown in Figure 7.

The Success Plot curve in Figure 7 is drawn using the open-source test code officially published by CNN. It was that the accuracy of the algorithm proposed in this paper has achieved a certain degree of accuracy improvement over its benchmark algorithms Siamese and Siamese. We analyze this improvement mainly due to it benefits from the learning of the template frame coordinate supervision information by the algorithm. The Siamese algorithm only predicts the position of the target through cross-correlation operations but cannot perform regression calculations on the coordinate frame. Therefore, the model cannot adapt well when the object is deformed. Siamese can coordinate frame calculation, only the classification information of the target is used, and the

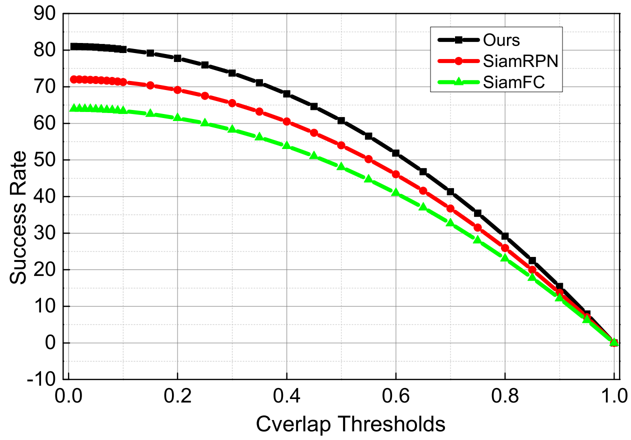

FIGURE 7. Herein algorithm OTB-100 on AUC results.

foreground and background are simply distinguished, so the tracking effect will decrease when interference objects like the target appear. The algorithm in this paper makes full use of the template frame supervision information and updates the parameters of the convolution kernel of the regression branch in the network according to the ground truth of the coordinates of the target, which improves the accuracy of the network to the coordinate frame regression.

In the CNN, many of the current mainstream tracking algorithm testing tools provided in the official results available to the user as a comparison reference, but most algorithms are missing CNN-100 experiments on, so this article CNN -13 on also made the corresponding test And compared with several mainstream tracking algorithms with better results, the results are shown in Figure 8.

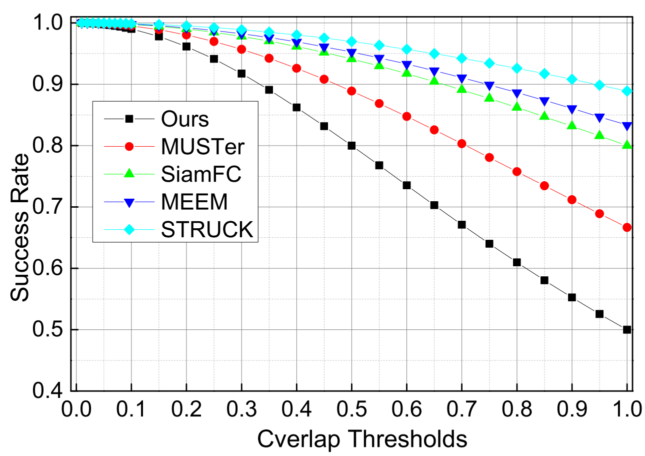

FIGURE 8. The algorithm in OTB-CVPR13 on AUC results.

Besides, although some algorithms not included in the official optical transceiver board (OTB) test tool, the test results on the OTB data set given in related papers. Figure 9 lists the algorithms in this paper and the major mainstream tracking algorithms. The performance comparison of the public results on the OTB data set can be seen that the algorithm of this

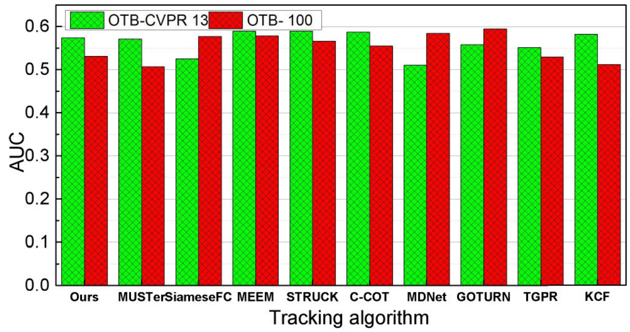

FIGURE 9. Each class in the tracking algorithm OTB test results.

paper has achieved very good results in both accuracy and speed in the current authoritative data set.

Since the target detection process sometimes blocked by other objects or the mobile phone's movement rate is too fast, the target image area blurred so that some objects not detected during target detection. In this case, the number of detected boxes is less than the number of predicted boxes, resulting in the lack of matching targets for the prediction box that should be matched with the target box.

The actual video tracking results of the tracking algorithm in the OTB-100 data set in this paper show that the first picture in each video scene is the first frame training sample, the target to be tracked is marked with a green coordinate frame, and the next three pictures are the algorithm. In the tracking results, real annotations also marked in green coordinate boxes, and the prediction results marked in red to coordinate boxes. It was that the tracking algorithm in this paper can achieve good results on OTB, the current mainstream authority video data set, and can effectively track the movement of most objects in the video.

B. PERFORMANCE RESULTS ANALYSIS

The visual-inertial fusion poses measurement in this chapter uses the binocular visual motion tracking module as the main pose measurement method, and its performance will affect the accuracy of visual-inertial fusion poses measurement. In evaluating the visual-inertial fusion based on incremental learning proposed, the performance of the pose measurement method evaluated the performance of the binocular visual motion tracking module. To evaluate the positioning accuracy of the binocular visual motion tracking module, a calibration platform is designed. To ensure that the test environment of the binocular visual motion tracking module is the same as its application environment, Autodesk Inventor design the calibration platform according to the workspace of the motion interactive scene. The designed calibration platform contains cylindrical positioning holes, and these cylindrical positioning holes evenly distributed on the plane. An interactive instrument with coloring marking points designed. The bottom end of the interactive instrument is spherical, called a positioning ball, and the positioning ball can be placed directly in the cylindrical positioning hole. The above design is to ensure that the positioning ball can be fixed in a certain position. Besides, the interactive instrument ball with the positioning ball can rotate freely in the positioning hole, so that the interactive instrument can be placed in the positioning hole in different postures. According to the statistical results in Figure 10, it was that the average standard deviation of the three-dimensional position obtained by the binocular visual motion tracking module is 0.28mm, indicating that the binocular visual motion tracking module used in the motion visualization interactive system in this chapter has high positioning tracking Precision. This provides accurate real values of position data for the visual-inertial fusion pose measurement method based on incremental learning proposed in this chapter. On this basis, the performance of the visual-inertial fusion poses measurement method based on incremental learning will be evaluated next.

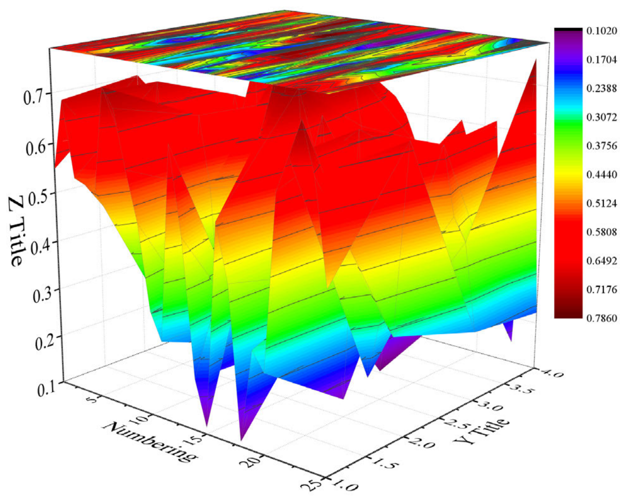

FIGURE 10. Binocular vision motion tracking module positioning information statistics.

To verify the performance of the incremental learning method based on the cascaded network, the constructed cascaded network net is used to realize the incremental pose estimation at different growth rates (ratio), and the growth rate of the motion target attitude estimation frequency is set to ratio = 2, 4, 6, 8, 10, that is to increase the frequency of attitude estimation by 2 times, 4 times, 6 times, 8 times and 10 times, and trained 5 cascaded networks net respectively.

It showed the mean absolute error between the predicted and true values of the pose data obtained by the net model at different growth rates. As shown in Figure 11, it can be seen in three dimensions to the attitude-angle component with increasing growth rate, the average absolute error between the predicted value and the true value of the posture angle is gradually increased. It shows the mean absolute attitude angle. The maximum value of the error is less than 0.02 degrees. Therefore, the incremental learning method based on the cascade network proposed in this chapter can improve the frequency of pose measurement while maintaining high accuracy.

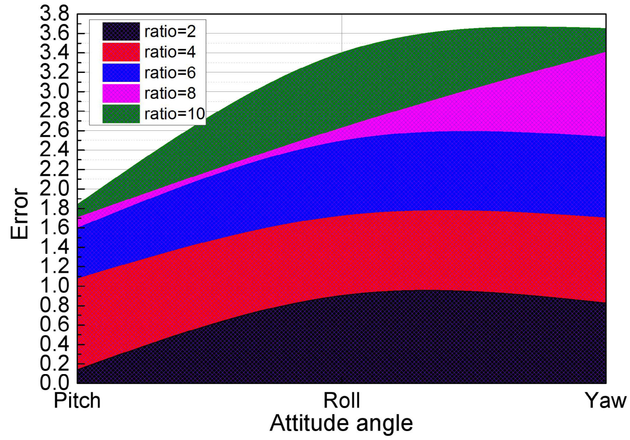

FIGURE 11. Statistics of mean absolute error of attitude increment estimation under different growth rates.

The performance of visual-inertial fusion poses measurement method based on incremental learning evaluated. The cascaded network net trained to estimate the pose. Figure 12 shows the comparison between the predicted value of the post data and the true value. It can be seen that for the pose The three components of the angle, the predicted value of the attitude data obtained by the cascaded network net is still very close to the true value when the estimated frequency is increased by 10 times, indicating that the incremental learning method proposed in this chapter has better performance in attitude estimation.

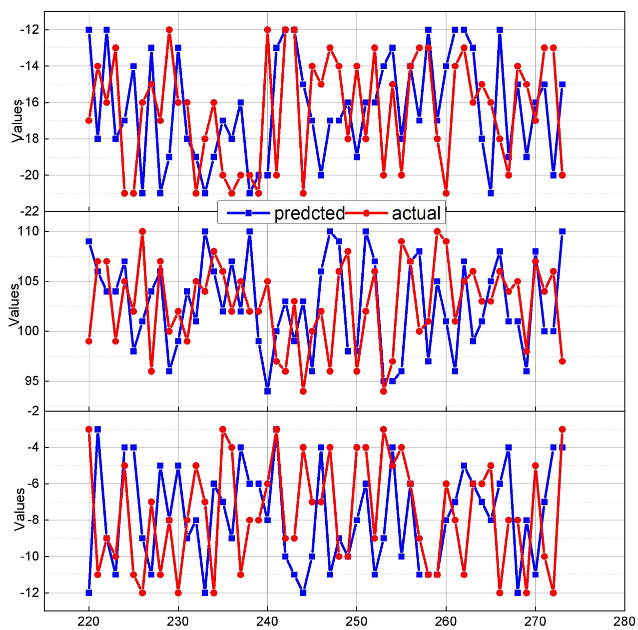

FIGURE 12. Comparison of the predicted and actual values of attitude angle when ratio=10.

To evaluate the performance of position estimation based on the incremental method, the operation introduced in the pose estimation process, and a nested cascade network net constructed to alleviate the error caused by the initialization operation. To verify the performance of the embedded nested cascade network net based on the θ- incremental learning method, each planet contains 10 the net subnetworks. The net position of the model obtained the mean absolute error between the predicted value and the real value data in Figure 13. We found that for each degree of freedom, the described θ-nested cascade network, the average absolute error of the position data of the 11 components x, y, s between the predicted values of the three initialization methods obtained is very close to the true value. Construct incremental learning methods. The results show that the position estimation is very robust. The planet position estimation model average absolute error is less than 1mm, reflecting the improved position estimating a frequency 10 which folds the positioning accuracy is still high.

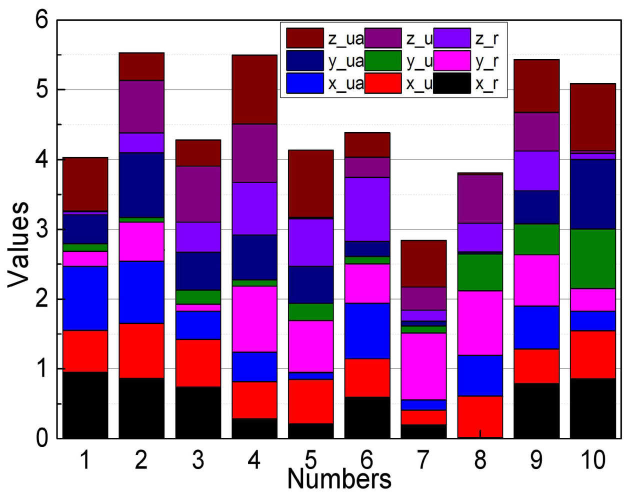

FIGURE 13. **Three Species initialization method in position and attitude measurement frequency.**

Otherwise, it is determined that the target has disappeared, and the position of the selection box not predicted by Kalman filtering. However, for the newly appeared target detection frame, which does not match the existing prediction frame, and still does not match the prediction frame within the disappearance threshold, it judged that the detection frame detects the newly appeared target. The disappearance threshold dynamically changed according to the selection box in the visual area. If the position at the time of disappearance is at the center of the vision, there is a high probability that it detected but not disappeared, and a larger threshold for disappearance can be given.

The problem of visual-inertial fusion poses measurement defined as a regression problem, and an incremental learning method based on the cascade network is proposed to gradually learn the increment of the time series variable. Based on the proposed incremental learning method, a cascaded network net and a nested cascaded network net are constructed to estimate the increments of attitude data and position data respectively to quickly and accurately track the magnetic interactive instruments in the motion visualization interactive system to improve motion visualization Realism and stability of interactive experience.

The visual-inertial fusion poses measurement method proposed in this chapter applied to a motion-based visual-tactile interaction system. The experimental results show that while maintaining high accuracy, the average absolute error of position and attitude estimation is less than 1mm and 0.02. The method can achieve a posture measurement frequency of 200 Hz, which shows that the incremental learning method based on the cascade network proposed in this chapter can increase the posture measurement frequency while maintaining a high posture measurement accuracy, and can lay the foundation for immersion motion visualization interaction.

The advantage of this method is that the number of three-dimensional unstructured mesh nodes on each computing node is basically equal, which increases the data communication between computing nodes. Considering that the calculation of inactive nodes is basically negligible, the method of weighted partitioning is adopted. The weight function of the active node is much larger than the weight function of the inactive node, thus ensuring that the number of active nodes in each partition is substantially equal. After the partition is completed, a data transfer list between partitions is also established. After the node and the time step of each physical time step are iteratively completed, the values of each sub-area are sent back to the computing master node.

V. CONCLUSION

This article studies the construction of the scene of sport virtual video and compares and analyzes the visual algorithm model of its sports training. The human movement style is an indispensable part of the fields of animation, movies, and games. The display needs to be achieved through the collection of human motion data of the real scene, and then through manual processing in the later stage. Therefore, this article conducts an in-depth study of human movement synthesis and style transfer. By extracting the spatiotemporal features of human movement data, a style transfer model established to automatically synthesize the required human movement style data. Aiming at the abstraction of the original motion capture data, this paper proposes a combination of RBM and self-encoding motion style transfer model to map the original human motion data to the feature space, establish style transfer constraints in the feature space, and map the features to the original human motion space to achieve style transfer. The model uses the self-encoding network to have the characteristics of the symmetric structure of encoding and decoding, which is convenient to encode the original human motion data, map it to the feature space, and then obtain the generated human motion data by decoding. At the same time, the Gram matrix used to establish constraints on the extracted motion features, and the control network is used for style transfer. Finally, experiments show that the model has a certain style of transferability and can transfer the style of one movement to another. The application of motion capture data in real application scenarios such as movies are complicated and requires further processing in the later stage. This paper establishes a pixel-level human motion style transfer model based on conditional confrontation networks and attempts to directly generate video human motion that can be directly applied to reduce the manual operation of designers. The model combines convolution and LSTM, and it can not only learn the spatial information in the video data, but also learn the temporal correlation between each frame of video, and adopt an unsupervised training method, without mentioning the provision of real style transfer. Experimental results show that this method can transfer the motion style of one human body in the video to the human body in another image. The research in this paper will have a huge impact and reference value on the future development of virtual video technology. In the future, we will continue to study the construction of sports virtual video scenes, and make a more in-depth comparison and analysis of the visual algorithm model of its sports training.

REFERENCES

[1] M. Renzo, H. Haas, and P. Grant, "Spatial modulation for multiple-antenna wireless systems: A survey," *IEEE Commun. Mag.*, vol. 49, no. 12, pp. 182–191, Dec. 2011.

[2] E. Basar, M. Wen, R. Mesleh, M. D. Renzo, Y. Xiao, and H. Haas, "Index modulation techniques for next-generation wireless networks," *IEEE Access*, vol. 5, p. 16693–16746, 2017.

[3] Y. Ding, K. J. Kim, T. Koike-Akino, M. Pajovic, P. Wang, and P. Orlik, "Spatial scattering modulation for uplink millimeter-wave systems," *IEEE Commun. Lett.*, vol. 21, no. 7, pp. 1493–1496, Jul. 2017.

[4] Y. Naresh and A. Chockalingam, "On media-based modulation using RF mirrors," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4967–4983, Jun. 2017.

[5] D. N. Viet, M. Di Renzo, V. Basavarajappa, B. Bedia Exposito, J. Basterrechea, and D.-T. Phan-Huy, "Spatial modulation based on reconfigurable antennas: Performance evaluation by using the prototype of a reconfigurable antenna," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 149, Jun. 2019.

[6] D.-T. Phan-Huy, Y. Kokar, K. Rachedi, P. Pajusco, A. Mokh, T. Magounaki, R. Masood, C. Buey, P. Ratajczak, N. Malhouroux-Gaffet, J.-M. Conrat, J.-C. Prevotet, A. Ourir, J. De Rosny, M. Crussiere, M. Helard, A. Gati, T. Sarrebourse, and M. Di Renzo, "Single-carrier spatial modulation for the Internet of Things: Design and performance evaluation by using real compact and reconfigurable antennas," *IEEE Access*, vol. 7, pp. 18978–18993, Jan. 2019.

[7] G. Lavigne, K. Achouri, V. S. Asadchy, S. A. Tretyakov, and C. Caloz, "Susceptibility derivation and experimental demonstration of refracting metasurfaces without spurious diffraction," *IEEE Trans. Antennas Propag.*, vol. 66, no. 3, pp. 1321–1330, Mar. 2018.

[8] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A new wireless communication paradigm through software-controlled metasurfaces," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 162–169, Sep. 2018.

[9] M. D. Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, J. D. Rosny, A. Bounceur, G. Lerosey, and M. Fink, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 129, May 2019.

[10] V. S. Asadchy, M. Albooyeh, S. N. Tcvetkova, A. Díaz-Rubio, Y. Ra'di, and S. A. Tretyakov, "Perfect control of reflection and refraction using spatially dispersive metasurfaces," *Phys. Rev. B, Condens. Matter*, vol. 94, no. 7, Aug. 2016, Art. no. 075142.

[11] K. Achouri and C. Caloz, "Design, concepts, and applications of electromagnetic metasurfaces," *Nanophotonics*, vol. 7, no. 6, pp. 1095–1116, Jun. 2018.

[12] S. Bhat, U. R. Acharya, N. Dadmehr, and H. Adeli, "Clinical neurophysiological and automated EEG-based diagnosis of the alzheimer's disease," *Eur. Neurology*, vol. 74, nos. 3–4, pp. 202–210, Jun. 2015.

[13] J. D. Griffin and G. D. Durgin, "Complete link budgets for back scatter radio and RFID systems," *IEEE Antennas Propag. Mag.*, vol. 51, no. 2, pp. 11–25, Apr. 2009.

[14] P. del Hougne, M. Fink, and G. Lerosey, "Optimally diverse communication channels in disordered environments with tuned randomness," *Nature Electron.*, vol. 2, no. 1, pp. 36–41, Jan. 2019.

[15] S. V. Hum and J. Perruisseau-Carrier, "Reconfigurable reflectarrays and array lenses for dynamic antenna beam control: A review," *IEEE Trans. Antennas Propag.*, vol. 62, no. 1, pp. 183–198, Jan. 2014.

[16] S. Narayanan, M. Di Renzo, F. Graziosi, and H. Haas, "Distributed spatial modulation: A cooperative diversity protocol for half-duplex relay-aided wireless networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 2947–2964, May 2016.

[17] W. Tang, J. Y. Dai, M. Chen, X. Li, Q. Cheng, S. Jin, K.-K. Wong, and T. J. Cui, "Programmable metasurface-based RF chain-free 8PSK wireless transmitter," *Electron. Lett.*, vol. 55, no. 7, pp. 417–420, Apr. 2019.

[18] L. Subrt and P. Pechac, "Intelligent walls as autonomous parts of smart indoor environments," *IET Commun.*, vol. 6, no. 8, pp. 1004–1010, May 2012.

[19] G. H. Elzwawi, H. H. Elzwawi, M. M. Tahseen, and T. A. Denidni, "Frequency selective surface-based switched-beamforming antenna," *IEEE Access*, vol. 6, pp. 48042–48050, 2018.

[20] T. J. Cui, M. Q. Qi, X. Wan, J. Zhao, and Q. Cheng, "Coding metamaterials, digital metamaterials and programmable metamaterials," *Light, Sci. Appl.*, vol. 3, no. 10, p. e218, Oct. 2014.

[21] L. Zhang, X. Q. Chen, S. Liu, Q. Zhang, J. Zhao, J. Y. Dai, G. D. Bai, X. Wan, Q. Cheng, G. Castaldi, V. Galdi, and T. J. Cui, "Space-timecoding digital metasurfaces," *Nature Commun.*, vol. 9, no. 1, p. 4334, Oct. 2018.

[22] H. Yang, X. Cao, F. Yang, J. Gao, S. Xu, M. Li, X. Chen, Y. Zhao, Y. Zheng, and S. Li, "A programmable metasurface with dynamic polarization, scattering and focusing control," *Sci. Rep.*, vol. 6, no. 1, p. 35692, Oct. 2016.

[23] C. Liaskos, S. Nie, A. Tsioliaridou, A. Pitsillides, S. Ioannidis, and I. Akyildiz, "A novel communication paradigm for high capacity and security via programmable indoor wireless environments in next generation wireless systems," *Ad Hoc Netw.*, vol. 87, pp. 1–16, May 2019.

[24] K. Achouri, M. A. Salem, and C. Caloz, "General metasurface synthesis based on susceptibility tensors," *IEEE Trans. Antennas Propag.*, vol. 63, no. 7, pp. 2977–2991, Jul. 2015.

[25] B. Ying and A. Nayak, "Anonymous and lightweight authentication for secure vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 12, pp. 10626–10636, Dec. 2017.

[26] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface EMG signal based on fuzzy entropy," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 2, pp. 266–272, Jun. 2007.

[27] H. Azami and J. Escudero, "Coarse-graining approaches in univariate multiscale sample and dispersion entropy," *Entropy*, vol. 20, no. 2, p. 138, Feb. 2018.

[28] H. Zhong, B. Huang, J. Cui, Y. Xu, and L. Liu, "Conditional privacy-preserving authentication using registration list in vehicular ad hoc networks," *IEEE Access*, vol. 6, pp. 2241–2250, 2018.

[29] A. Studer, F. Bai, B. Bellur, and A. Perrig, "Flexible, extensible, and efficient VANET authentication," *J. Commun. Netw.*, vol. 11, no. 6, pp. 574–588, Dec. 2009.

[30] M. Vidailhet, S. Rivaud, N. Gouider-Khouja, B. Pillon, A.-M. Bonnet, B. Gaymard, Y. Agid, and C. Pierrot-Deseilligny, "Eye movements in parkinsonian syndromes," *Ann. Neurol.*, vol. 35, no. 4, pp. 420–426, Apr. 1994.

[31] X. Lin, X. Sun, X. Wang, C. Zhang, P. H. Ho, and X. Shen, "TSVC: Timed efficient and secure vehicular communications with privacy preserving," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4987–4998, Dec. 2018.

[32] K.-A. Shim, "$\mathcal{CPAS}$: An efficient conditional privacy-preserving authentication scheme for vehicular sensor networks," *IEEE Trans. Veh. Technol.*, vol. 61, no. 4, pp. 1874–1883, May 2012.

[33] D. He, S. Zeadally, B. Xu, and X. Huang, "An efficient identity-based conditional privacy-preserving authentication scheme for vehicular ad hoc networks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2681–2691, Dec. 2015.

[34] N. Kshetri and J. Voas, "Blockchain-enabled E-Voting," *IEEE Softw.*, vol. 35, no. 4, pp. 95–99, Jul. 2018.

RUI YUAN was born in Henan, China, in 1993. She received the bachelor's and master's degrees from Zhengzhou University, in 2015 and 2018, respectively. She is currently pursuing the Ph.D. degree with Pukyong National University, South Korea. She has been published a total of three articles. Her research interests include sports training and sports psychology.

ZHENDONG ZHANG was born in Beijing, China, in 1961. He received the bachelor's degree from Wuhan Sport University, in 1983. He currently works with Zhengzhou University. He has hosted a general project of the China National Social Science Foundation. He has published a total of 12 articles on CSSCI. His research interests include sports training and social sports.

PENGWEI SONG was born in Henan, China, in 1991. He received the bachelor's and master's degrees from Zhengzhou University, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. degree in private training with Keimyung University and Zhengzhou University. He also works with Guangxi Science and Technology Normal University. He has published a total of five articles. His research interests include sports training, sports psychology, social sports, and private training.

JIA ZHANG was born in Henan, China, in 1992. He received the bachelor's degree from Zhengzhou University, in 2015, and the master's degree from Yanbian University, in 2018. He is currently pursuing the Ph.D. degree in sports psychology with Chung-Ang University, South Korea. He has published a total of two articles. His research interests include sports psychology and physical education of children.

LONG QIN was born in Henan, China, in 1990. He received the bachelor's degree from Zhengzhou University, in 2015, and the master's degree from Wuhan Sports University, in 2018. He is currently pursuing the Ph.D. degree with Keimyung University. He also works with Guangxi Science and Technology Normal University. He has published a total of three articles. His research interests include sports training, sports psychology, and social sports.