Journal of Information Technology Review

DLINE JOURNALS

Print ISSN: 0976-3511
Online ISSN: 0976-2922

JITR 2025: 16 (4)
https://doi.org/10.6025/jitr/2025/16/4/146-153

Research on Intelligent Algorithm-based Swimming Athlete Pose Recognition and Correction Method

Xinying Bo, Bodong Zhang
Guangzhou sport university, 510500
Guangzhou, Guangdong, China
z475300831@gmail.com

ABSTRACT

In order to better assist swimming athletes in correcting their abnormal postures, we have developed a pose recognition technology based on depth images. This technology uses threshold algorithms for data preprocessing and Kalman filters to filter noise. We also employ Gaussian distribution functions to capture dynamic changes and utilize the SURF algorithm to remove blurry parts. This technology can help us better assist swimming athletes in correcting their abnormal postures and improve their daily training. Using the Euclidean distance method, we can accurately estimate the distance between two adjacent reference points and employ feedback monitoring techniques to correct improper postures. Through simulations, we have found that this new deep level image skeleton tracking technology can effectively capture the dynamics of athletes and accurately detect their poses, demonstrating high accuracy and stability.

Keywords: Swimming Competition, Pose Movements, Skeleton Tracking, Image Detection and Recognition

Received: 27 April 2025, Revised 4 August 2025, Accepted 15 August 2025

1. Introduction

The World Swimming Championships are organized by world class professionals who highly value the performance of the participants. They believe that only through correct swimming postures can ideal victories be achieved. Therefore, they have high training requirements for the participants to ensure that they can achieve the best performance [1]. Accurate observation and correction are essential steps for swimming athletes to continuously improve their skills and achieve victories. Although coaches can provide verbal guidance in specific environments, this method is not sufficient to help students grasp “feedback information” and effectively handle various complex situations [2]. Furthermore, using manual demonstrations to guide students’ learning also has certain shortcomings as students often struggle to learn and master key skills, hindering true learning and the ability to overcome challenges. Utilizing advanced multimedia technology to correct athletes’ improper postures is highly necessary [3]. From the perspectives of video recording, image flipping, real-time shooting, and review, it is possible to accurately identify athletes' improper postures, which is a hot topic discussed by industry professionals and scholars. The research significance of this topic cannot be underestimated [4]. In the field of athlete pose recognition, technologies such as Li G, Zhang C. have important applications [5]. Long T. uses target tracking algorithms to ensure accurate tracking, achieving precise follow up [6]. Wang J, Qiu K, Peng H, et al. utilizes dynamic module matching algorithms to effectively detect and correct athletes' improper postures [7]. Additionally, Long F employs SAA7111 and FPGA to achieve automated data acquisition and storage without using DSP, and uses DMA for high speed data transmission, achieving accurate tracking and correction [8]. Through PCI cards, we can accurately detect and correct the postures of swimming athletes. This efficient data analysis method using DSP technology enables fast and accurate data transmission and analysis. Although traditional measurement techniques can provide a certain level of accuracy, they can still be distorted due to the complexity of external factors such as dimensions, noise, and distance [9]. Images, as an effective means of data collection and processing, play a crucial role, especially in the field of visual error correction when photographers capture motion scenes using photographic devices or other technologies. The research achievements of visual error correction technology will become a core issue in the field of multimedia vision [10].

With the rapid advancement of technology, fitness activities have far exceeded traditional physical exercises, and their precision and complexity have greatly improved. Therefore, using precise image recognition technology to help students better grasp and understand these movements can reduce visual errors during exercise [11]. By accurately processing images of fitness movements, we can provide more precise videos to provide the best experience for the audience. Furthermore, we propose a new visual error calibration technology that can be widely used in daily life. To address the shortcomings of traditional techniques, we have developed a novel computer model that utilizes depth learning based skeleton tracking technology and an improved algorithm based on SIFT [12] to effectively capture images of moving targets [13]. It can also adjust rotation angles, sizes, brightness, and other parameters as needed to ensure the realism of the model. Additionally, it can suppress noise, thereby achieving accurate monitoring of swimming postures. This allows athletes to clearly identify their shortcomings and make prompt corrections, leading to effective training outcomes.

2. Swimmer Pose Recognition Correction Method

2.1 Principles of athlete posture recognition and correction

By using Formula (1), we can accurately locate the various key points of the swimming athlete during the pose recognition and correction process by treating the foreground pixel points $Q=(QR, QG, QB)$ and the background pixel points $W=(ER, Ec, EB)$ as parameters.

$$G(x, y, z) = \frac{Q \cdot W}{(\Delta_{T} \quad {}_{D} \quad {}_{0} \quad {}_{LD} \quad {}_{L1} \quad {}_{Ro} \quad {}_{R1})} \tag{1}$$

Analyzing the located key points, we can extract the characteristic points of the athlete's limbs at various joints during swimming movements using Equation (2).

$$X^2 + Y^2 + Z^2 = \sqrt{\frac{\text{shead}}{\pi}} \times \frac{(x, y)}{R} \frac{(x - r \cdot \Phi, y + r)}{\Phi = (1 + \sqrt{5}) / 2} \tag{2}$$

In this paper, R refers to the range of motion of the athlete, $(x\text{-}r, y+r)$ represents the position of the left button, and (1+5)/2 represents the x-coordinate of the left side. Based on the principle of symmetry, we can calculate the x-coordinate of the right side. Assuming ($x+y$^2) represents the length of the athlete's forearm, and $x=(f/z)$, $y=(f/z)y$, we can obtain the coordinates of the four buttons: left, right, top, and bottom, and calculate the distance of each button. By using Equation (3), we can construct a model that accurately detects and corrects the athlete's pose.

$$P = \frac{G(x, y, z) \times \left(X^2 + Y^2 + Z^2\right)}{\left(x^2 + y^2\right) \times \left(X, Y, Z\right)} \times \left(B_0, B_1, B_2, B_3\right) \tag{3}$$

During swimming, the human body is prone to affine deformation, resulting in low contrast motion feature points. Traditional methods extract these feature points and compare them with the correct pose to achieve pose recognition and correction, which leads to the inability to detect and correct errors in real time [14, 15]. This paper proposes a swimming athlete pose recognition and correction method based on depth image skeleton tracking. By using gray level co occurrence matrix, we can better analyze the texture of the animation. This matrix captures the similarities between different pixels in the image and maps their positions and sizes. In this way, we can better describe the texture of the image and more accurately predict their motion trajectories [16, 17]. With the gray level co occurrence matrix, we can start from the beginning of each pixel, and the distance between each pixel can be determined by the matrix. Here, (i, j) represents the relationship between two pixels with different gray values, and $f(x, y)$ represents the distance between two pixels, i.*e*., a, y, which indicates their continuous relationship [18, 19].

2.2 Swimming Athlete Pose Recognition and Correction based on Depth Image Skeleton Tracking

With the rapid development of IT technology, the hardware infrastructure of DV video has reached a certain level, enabling effective detection and recording of athletes' movement trajectories. It has been widely applied in various fields, especially in the swimming domain. Firstly, the images are obtained using Kinect technology. Secondly, the Kalman predictor is used in combination with the SIFT algorithm to implement timely detection and recognition of swimming target poses. Based on the image deep cut tracking algorithm, the position of the athlete's target action is estimated by deep cutting the image, transforming complex prediction problems into relatively easy classification problems. The human swimming pose is used as the target for tracking and detection. To better achieve motion target monitoring, Kinect is used to obtain deep cut images. The acquisition method is as follows: 1) Use the sensors on the left and right sides of Kinect to send and receive infrared rays from different directions. Kinect achieves three dimensional spatial segmentation by using two components: the left side of Kinect transmits "light coding" to the surrounding space through the infrared emitter of Kinect in the blue band, and the right side of Kinect captures the blue band of Kinect through the infrared receiver of Kinect. Through the three dimensional spatial segmentation of Kinect, the three dimensional spatial segmentation results of Kinect can be obtained and utilized. Accurate tracking of the human skeletal structure is performed based on the collected relevant information. Based on the obtained information parameters, images of each skeletal point are extracted. Firstly, the human image is extracted from the background environment. Secondly, the repeated parts of each body part are carefully examined from the front, side, and top down angles, collecting 20 key points and combining them into a complete skeletal structure diagram. The Kinect camera can extract useful information from multiple depth cut images, label and classify them to better predict the position of human body movements. Additionally, these labeled images can be combined with three dimensional nodes to better simulate realistic human movements. Figure 1 shows the matching diagram of skeletal points formed by the target image.

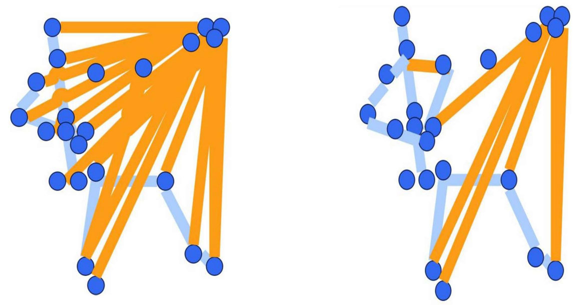

Figure 1. Skeletal point matching map formed by target image

With the image extraction techniques described above, we can accurately determine the position of the motion target, which will provide strong support for subsequent swimming error detection and recognition.

2.3 Improvement of SIFT Image Recognition Algorithm Combined with Kalman Predictor

Using three dimensional stereoscopic technology, we can extract the target image and search for swimming poses to obtain their positions, relevant information, as well as the positions and location information of the swimmers. The threshold algorithm can effectively combine the established target module with the unknown image module. The SIFT algorithm can extract partial feature points of the moving target from multiple video images and accurately locate the optimal extremal points. This allows us to obtain unaffected data such as the pose, location information, and rotation angle of the swimming motion target. Additionally, it can accurately match the same motion target in multiple video images.

Firstly, by using the scale space operation, the given signal can be represented as a set of signals, where $f(X)$ ($X \in R^2$) represents a scale space, $s(X, t)$ represents a time node, and $X(X \in R^2)$ represents a spatial coordinate. It can be represented as $s(X, t) = I(X, t)$, where $I(X, t)$ denotes the scale space of the original image signal when the scale is t. To obtain Gaussian difference video images, Gaussian functions with different α can be obtained from the original image, and the results can be calculated through Gaussian filtering at different scales. After determining the optimal point in the scale space through *DG*, the feature points to be recognized are obtained. Although the feature points to be recognized include all the information of the motion video image, there are still some marginal and low contrast points. Since these low contrast points and marginal points will have a certain impact on feature retrieval and recognition of the motion target image, it is necessary to remove these marginal and low contrast points. *SURF* method is introduced for screening, and the *SIFT* image feature point detection and matching method is employed to ensure that the range of changes in tilt angle, affine transformation degree, and noise level remains within a certain range under different rotation angles, scale spaces, and brightness variations.

However, in order to obtain the feature points of motion video images with a large range of view, more time is required, and real time requirements cannot be met. If we can determine the position of the video image in advance and only need to capture the feature points within the approximate range of the moving image, we can quickly match the feature images, greatly reducing the time for image matching and accelerating the detection and recognition speed of the motion target. By combining the feature extraction technique of the *SIFT* algorithm with the matching algorithm, an effective process for detecting the swimming motion target pose can be constructed. The correction of athlete's pose recognition is achieved through feedback monitoring principles.

3. Experimental Design and Results Analysis

3.1 Motion Blur Image Restoration

By using a new technique to improve athletes' motion performance, we can enhance their motion performance through a method called visual error correction. This technique extracts the high frequency layer information of the athlete's motion performance from a blurry image using a sliding window of a certain radius, and segments it into 1/4 regions for detection. By measuring the rich edge indices for different sliding windows, we found that among these values, higher values such as 13.0106 and 11.1146 indicate more edge information present in those areas. Therefore, we can use these values to evaluate the effectiveness of gymnastic exercises and their blur kernels. By enhancing the information in the edge regions, dynamic scenes can be better restored, ensuring not only visual clarity but also significantly reducing processing costs and greatly improving work efficiency, which is a prominent advantage. After precise calibration, our dynamic images appear very clear, without any jagged edges or blurriness, indicating the success of our correction technique and the reliability of our correction algorithm. Through comparison, we found that the algorithm can accurately capture the swimming poses of the motion target and perform rapid recognition. To test the reliability of the algorithm, we conducted a series of tests, including the first shooting and tracking at each step. By applying the algorithm that combines the Kalman predictor with SIFT to the swimming poses of the athletes, we observed significant improvements, as shown in Figure 2.

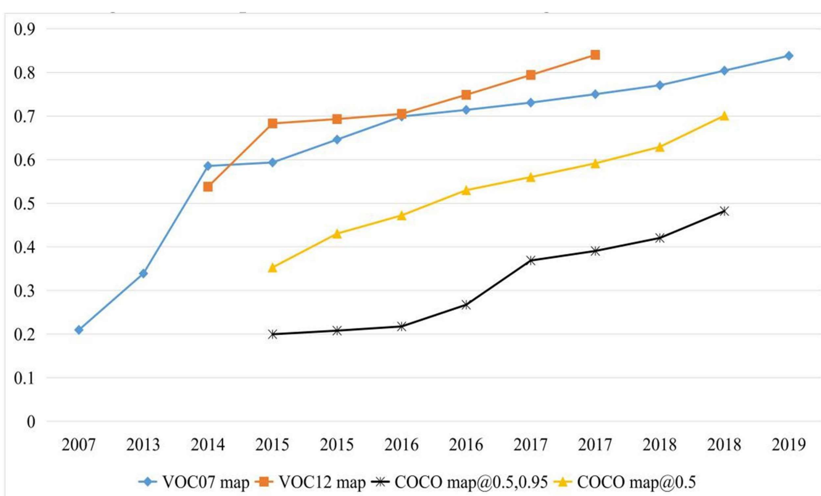

Figure 2. Comparison of Image Recognition between Traditional Algorithms and Our Algorithm

3.2 Image Recognition Performance

By applying the calibration scheme proposed in this paper to practical fitness motion scenarios, we can evaluate its accuracy by testing its signal to noise ratio, normalized mean square error, model consistency, and image quality.

By applying the Ziegler Nichols algorithm, we can effectively control the system to achieve accurate tracking

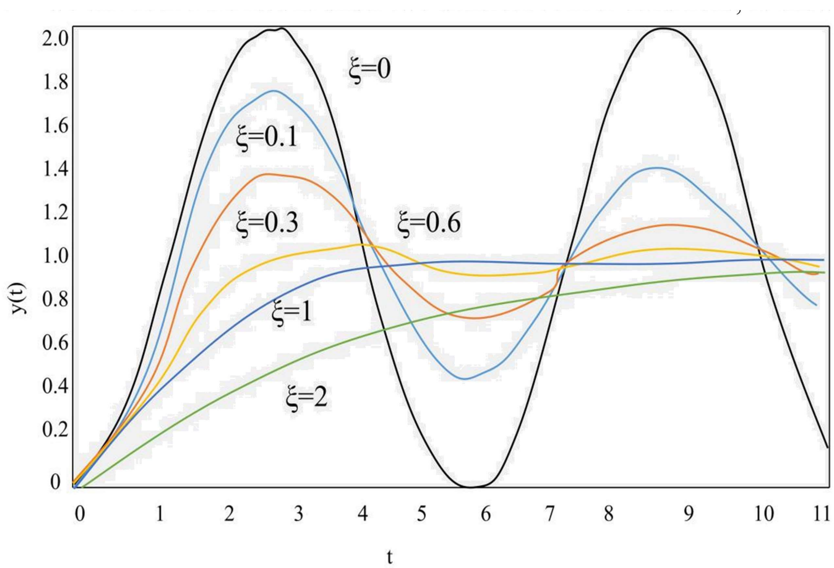

Figure 3. Step Response Curve of the System

and recognition of objects, and evaluate its image processing capabilities. To better evaluate the step response between the *PID* control system and the fuzzy *PID* control system, we took the following steps: k=1.44, h=0.75, k=0.7. By comparing the *PID* control parameters of the two groups, we can observe the changes in their step responses. First, we can use the traditional control method to control the system, and then we can use the new control method to control the system. Finally, we can use the new control method to control the system and obtain a better step response. By applying median filtering, we can obtain the results under two different control conditions, as shown in Figure 3.

According to Figure 3, using the traditional PID control method, the adjustment of the dynamic swimming posture tracking system reduces the time consumption from 2s to 24%. By adopting the fuzzy *PID* approach, the time consumption is reduced from 1.2s to 12%, while the adjustment time is reduced from 1.9*s* to 4*s*. This significantly improves the performance of the camera in tracking the motion target. With the use of specific algorithms, we can accurately recognize objects in motion and quickly locate them. Our recognition algorithm can accurately track and identify the position of objects, while improving their peak signal to noise ratio, structural similarity index, and image quality. All these indicate that our recognition algorithm significantly reduces the normalized mean square error, thereby validating the effectiveness of our recognition algorithm. When all submodules are successfully implemented, we can analyze dynamic images by calculating fractal codes and determining the offset between them. By applying interpolation, we can achieve optimal correction results.

4. Conclusion

This paper has focused on utilizing various techniques such as sampling and interpolation to reconstruct and enhance details and features in blurry dynamic scenes, thereby eliminating their impact on human vision. Through practical testing, we found that this technology is highly effective in addressing visual issues for human perception. However, this study still has its limitations. To address this, we propose a new deep image based skeletal tracking technique for training the body posture of swimmers. Through simulation, we found that this technique can better capture and assess the athletes' physical states, demonstrating higher accuracy and improved stability.

References

[1] Yuan, R., Zhang, Z., Le, Y., et al. (2021). Adaptive recognition of motion posture in sports video based on evolution equation. *Advances in Mathematical Physics, 2021*, 1–12.

[2] He, Y. (2022). Research on leg posture recognition of sprinters based on SVM classifier. *International Journal of Biometrics, 14*(3–4), 367–382.

[3] Tsai, M. F., Huang, S. H. (2022). Temporal-variation skeleton point correction algorithm for improved accuracy of human action recognition. *International Journal of Pattern Recognition and Artificial Intelligence, 36*(10), 2500355.

[4] Yin, J. (2021). The method of table tennis players' posture recognition based on a genetic algorithm. *International Journal of Biometrics, 13*(2–3), 243–257.

[5] Li, G., Zhang, C. (2019). Automatic detection technology of sports athletes based on image recognition technology. *EURASIP Journal on Image and Video Processing, 2019*, 1–9.

[6] Long, T. (2020). Research on application of athlete gesture tracking algorithms based on deep learning. *Journal of Ambient Intelligence and Humanized Computing, 11*, 3649–3657.

[7] Wang, J., Qiu, K., Peng, H., et al. (2019). AI coach: Deep human pose estimation and analysis for personalized athletic training assistance. In *Proceedings of the 27th ACM International Conference on Multimedia* (pp. 374–382).

[8] Hailong, L. (2021). Role of artificial intelligence algorithm for taekwondo teaching effect evaluation model. *Journal of Intelligent and Fuzzy Systems, 40*(2), 3239–3250.

[9] Bansal, S., Gupta, A., Garg, A. K. (2022). The intelligent water drops algorithm for burst scheduling in OBS network. *Materials Today: Proceedings, 48*, 1059–1063.

[10] Lin, T., Liu, X. (2021). An intelligent recognition system for insulator string defects based on dimension correction and optimized faster R-CNN. *Electrical Engineering, 2021*(1), 103.

[11] Long, F. (2021). Simulation of English text recognition model based on ant colony algorithm and genetic algorithm. *Journal of Intelligent and Fuzzy Systems, 2021*(4), 1–12.

[12] Zhang, X., Feng, Z., Yang, X., et al. (2021). An intelligent error correction algorithm for elderly care robots. *Applied Sciences, 11*(16), 7316.

[13] Srikanth, P., Koley, C. (2021). An intelligent algorithm for auto-recognition of power system faults using superlets. *Sustainable Energy Grids and Networks, 26*(22), 100450.

[14] Jindal, A., Aggarwal, N., Gupta, S. (2018). An obstacle detection method for visually impaired persons by

ground plane removal using speeded-up robust features and gray level co-occurrence matrix. *Pattern Recognition and Image Analysis, 28*, 288–300.

[15] Wei, B., Zhao, X., Wang, L., et al. (2019). Analysis of gear surface morphology based on gray level co-occurrence matrix and fractal dimension. *PLOS ONE, 14*(10), e0223825.

[16] Bin, Y., Cao, X., Chen, X., et al. (2020). Adversarial semantic data augmentation for human pose estimation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16* (pp. 606–622). Springer International Publishing.

[17] Li, S., Gunel, S., Ostrek, M., et al. (2020). Deformation-aware unpaired image translation for pose estimation on laboratory animals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 13158–13168).

[18] Liu, H., Chen, Y., Zhao, W., et al. (2021). Human pose recognition via adaptive distribution encoding for action perception in the self-regulated learning process. *Infrared Physics & Technology, 114*, 103660.

[19] Nosrati, M. S., Abugharbieh, R., Peyrat, J. M., et al. (2015). Simultaneous multi-structure segmentation and 3D nonrigid pose estimation in image-guided robotic surgery. *IEEE Transactions on Medical Imaging, 35*(1), 1–12.