J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 1983--1993

Journal of Combinatorial Mathematics and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Human Motion Morphology Capture Based on Computer Vision and Image Segmentation Algorithms

Yubao Zhang[1,✉]

[1] *School of Design and Communication, Zhejiang Fashion Institute of Technology, Ningbo, Zhejiang, 315211, China*

ABSTRACT

Although human motion form capture is widely used in multiple fields, it often requires a significant amount of time and cost to learn how to operate the device during use. Therefore, this article attempted to apply computer vision (CV) technology and image segmentation algorithms to human motion form capture technology, simplifying the operation scheme and improving recognition accuracy and efficiency. This article provided an in-depth analysis of human motion form capture technology. Firstly, it identified several parts of the current human motion form capture technology that can be optimized, and introduced the effects of these optimized parts on human motion form capture in sports training. This article took the form capture of aerobics athletes as a sample and extracted 50 keyframe images containing aerobics scoring actions from 100 aerobics activity videos. The extraction interval for these keyframe images was at least 10 seconds. Next, this article used histogram equalization to enhance the image, while segmenting and recognizing the human motion forms of the five types of actions in the keyframe images, highlighting the level of action standards of athletes in aerobics. Finally, this article selected 6 key frame images containing different movements of aerobics athletes for comparative experimental analysis. In this experiment, both commonly used optical unlabeled capture techniques and motion morphology capture techniques combining CV and image segmentation algorithms were used to capture the human body in the image. The addition of CV technology and image segmentation has improved the overall performance of human motion morphology capture technology by approximately 26.02%. The integration of CV technology and image segmentation algorithms into human motion form capture technology has greatly improved image processing efficiency. At the same time, CV technology and image segmentation algorithms have also enabled better image processing accuracy in human motion form capture.

Keywords: Motion Capture, Computer Vision, Image Segmentation, Human Movement Analysis

✉ Corresponding author.
E-mail address: zbaob@163.com (Y. Zhang).
Received 15 September 2024; Revised 05 November 2024; Accepted 20 February 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-111

1. Introduction

The study of human body morphology is a research direction that started relatively late. In this research path, it mainly focuses on the analysis of human body morphology, thus spreading to the fields of Kinematics and medicine. It can determine various aspects of human body information through research on human movement patterns. At the same time, the capture of human motion forms is a focus issue in current related fields, and this type of motion form capture usually has deep-seated applications in many sports fields such as sports and dance. In the field of sports, human motion form capture is usually used to capture the form of excellent athletes. CV technology and image segmentation algorithms can enhance the clarity of captured images while improving image processing accuracy, helping raters rate athlete movements.

At present, researchers in the field of morphology have analyzed the human motion capture technology and determined the advantages, disadvantages, and application scope of the current human motion capture technology. For example, Alexander Mathis explored how to capture target shapes in motion. By taking advantage of the depth learning algorithm in image analysis, he proposed a unmarked motion capture technology, and determined the multiple advantages of this motion capture technology [1]. Eline, Van der Kruk studied the accuracy of human motion form capture technology in the practical application process in the field of sports. He determined the difference of these motion capture through the research on motion capture techniques that have been used to capture athletes' movement patterns for a long time [2]. Xiaowei Zhou analyzed the application of a human motion form capture technology that uses neural network based image geometry processing. Through the analysis of the effect of the motion capture technology in two-dimensional and three-dimensional image processing, he determined that the technology has higher motion capture accuracy [3]. Gergely, Nagymate conducted research on the application and multifaceted performance of a motion capture technique that references biomechanical theory. After analyzing the data of this motion capture technology in multiple literature databases, he validated its performance in various aspects such as labeling accuracy and data processing [4]. Kang Chen proposed a new optical motion capture technology. Compared with the traditional optical motion capture technology, this motion capture technology mainly upgrades the internal human skeleton positioning module, so as to achieve a faster and more accurate human motion capture technology [5]. Gustavo Ramos Dalla Bernardina analyzed the performance differences between human motion capture technology and a motion optical capture technology in commercial applications. He mainly evaluated the effectiveness of two capture techniques in identifying and capturing various joints in the human body, in order to determine that human motion form capture techniques in commercial applications have higher capture accuracy [6]. Mreza Naeemabadi analyzed the impact of bone processing algorithms in marker based human motion capture technology on human capture accuracy. He determined through experiments and literature analysis that errors in bone processing in motion form capture technology can significantly reduce capture accuracy [7]. However, currently, most motion capture technologies have shortcomings such as weak linkage with image processing technology, resulting in poor readability of images captured from motion.

Researchers in the information field have studied the combination of CV or image processing technology and motion capture technology, which can improve the efficiency and application range of motion capture. Steffi L. Colyer mainly analyzed the application of motion capture in medical rehabilitation and sports fields, and explored the application and technical integration solutions of computer vision methods in unlabeled motion capture, proposing more ideas for the development of motion capture [8]. The feasibility and reliability of the application of a motion capture system optimized by computer vision technology in the detection of falls in the elderly have been determined by K. I. S. H. A. N. P. R. A. S. A. D Gunale through relevant experiments [9]. Cong, Zhang analyzed the role of image segmentation in the training of basketball players' throwing movements.

He mainly used image segmentation algorithms to recognize the movements of athletes and proposes a highly reliable method for recognizing basketball players' throwing movements [10]. Magdalena, Mazur-Milecka proposed the use of thermal imaging and image segmentation for motion detection and tracking of animals in the laboratory. He also confirmed that the motion detection and tracking technology composed of image segmentation has better performance through scientific control [11]. Hadish Habte Tesfamikael proposed a control scheme for electric wheelchairs based on image segmentation for eye movement detection and tracking, and determined that this control scheme can provide more convenience for the elderly and disabled [12]. This human motion capture technology, combined with CV technology and image processing technology, can not only be used in more movements, but also complete image processing at a faster speed.

This article studies the human motion form capture technology that combines CV technology and image segmentation algorithms, mainly by arranging multiple sensing devices around athletes. These sensing devices collect and capture the movement patterns of athletes, which is based on optical form capture. In the aerobics movement, the paper use the human motion form capture technology that combines CV technology and image segmentation algorithm to capture the athlete's form in multiple key frame images in the aerobics movement video. Firstly, it uses image histogram equalization to enhance the image, and then segments the athletes in the keyframe image to highlight the movement characteristics of athletes in different aerobics movements. Therefore, this article combines CV technology and image segmentation algorithms to capture human motion morphology, which can not only play an important role in athlete training. It can also assist raters in accurately scoring athletes' performance in various sports events.

2. Human Motion Form Capture

In the study of a human motion capture scheme based on CV technology and image segmentation algorithms, this article uses the human motion in aerobics as the target to capture the shape of aerobics athletes during their exercise process. In the evaluation and training of aerobics activities, the standard degree of upper and lower limb movements and trunk movements of aerobics athletes is generally analyzed. The lower limb movements mainly cover the athlete's pace rhythm, muscle and strength performance, while the upper limb movements mainly include the athlete's arm swing and basic gymnastics skills. The final torso movement mainly provides continuous protection for the athlete's movement process.

In the process of capturing the movement form of aerobics athletes, it is first necessary to set up a Mannequin, which can point out the key parts of the body of aerobics athletes and provide assistance for subsequent motion capture. The dataset used in this study is sourced from Kaggle's Sports Image - Image Classification dataset (available at https://www.kaggle.com/datasets/sheikhzaib/sports-image-image-classification?resource=download). The labeling results of the human body labeling model during aerobics exercise are shown in Figure 1.

Fig.1. Schematic diagram of the labeling results of the human body labeling model in the process of aerobics

In the process of obtaining this image, it is also necessary to establish different coordinate systems based on the different sites. This coordinate system can determine the spatial position of the target. Generally, CV technology is used to establish the coordinate system of the image and the coordinate system of the pixels. The coordinate system of the image is shown in Figure 2.

Fig. 2. Schematic diagram of the coordinate system in the aerobics image

In this process, it is necessary to mark the targets in the images and collect images from different locations and devices in the aerobics venue. Among these devices, there is a constraint relationship between the two corresponding image collection device points. Through a geometric operation, the motion with the corresponding constraint perspective can be planned, and the calculation formula is shown in Formula (1).

$$p = k\left(fp + t\right) \tag{1}$$

The point p represents the intersection of two perspectives on the target character; k represents the parameters of the image collection device; f and t represent the corresponding relationship between the two perspectives. The calculation formula for the coordinates of two perspective points is shown in Formula (2).

$$x_1 = k^{-1}p, x_2 = kp \tag{2}$$

A mathematical relationship that determines the relationship between the two perspectives by calculating the above formulas (1) and (2) can be used to plan the positions of the two perspectives. The calculation formula for this planning is shown in Formula (3).

$$Y = x_2^t \cdot t \cdot fpx_1 \tag{3}$$

These formulas mainly plan the perspective arrangement in the image based on the three-dimensional coordinate axis of the image, which can convert the characters in the two-dimensional image into three-dimensional character images after image processing and CV technology processing. This three-dimensional character image can enhance the credibility of the scoring staff in the objective evaluation of the action standards of aerobics athletes in aerobics, and can also point out non-standard movements of aerobics athletes in their daily training.

At the same time, the detection method of marker points in human motion also needs to set a threshold during the process of capturing motion forms. When the credibility of human marker point detection is less than the set threshold, it is necessary to use the currently obtained image pixels as untrusted pixels in the process of motion form capture and detection, and delete the same type of data in the human marker point sample database based on the multi-dimensional information of untrusted pixels.

The detection of human motion marker points in images is mainly achieved by extracting the morphological features of the target human motion. Generally, data mining algorithms based on association rules are used to construct a feature vector for detecting the credibility of marker points. At the same time, the dimension of this feature vector is reduced to establish a sample data set for the target human motion form and a sample data set for human marker points. The first step in feature extraction is to calculate the correlation of the target human motion form feature parameters, as shown in Formula (4).

$$y = \frac{\sum\left(a_n + b_n\right)}{k\sqrt{A^2 + 1}} \tag{4}$$

Among them, a_n represents the feature parameters of the human body's marker points, and k represents the feature proportion of human motion morphology. At the same time, the coefficient of feature correlation is calculated based on the characteristic parameters of different human body movements in aerobics, as shown in Formula (5).

$$S = \frac{2\left(Q_i + Q_0\right)}{Q_0} \tag{5}$$

After extracting the morphological features of the human body in aerobics, it would continue to use random selection in the k-means algorithm to select defects in the image. At the same time, key frames in aerobics videos can be extracted, and the human behavior dataset and feature vector dataset of human form in the key frame images can be calculated to further optimize the recognition effect of human marker point models in aerobics, and the two datasets can be gridded. The performance of this grid processing in images and the capture of character motion patterns is shown in Figure 3.

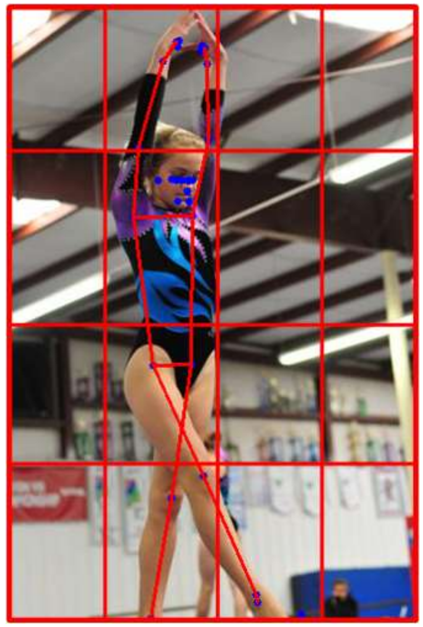


Fig. 3. Schematic representation of grid processing in images and representation of character motion form capture

The above marker point detection method mainly obtains a vector map of the target person's marker points in the image through feature matching. Each human marker point in the image corresponds to a vector map. At the same time, this vector map would generate a reliable distribution probability table for detecting character marker points in the image based on the three-dimensional position of the coordinate axis constructed by computer vision. The trusted marker points can be determined based on the maximum value of the image's trusted distribution probability, and through such reciprocating calculations, the trusted coordinate sample dataset of all trusted marker points in the coordinate system of the two-dimensional target image can be obtained. Each set of trustworthy coordinate sample datasets is composed of the coordinate system horizontal and vertical coordinate data of 12 marked points in the keyframe images of aerobics videos. The vector map of the target human body's marker points is set between two human marker points, and the maximum value of their values is determined by the probability of credibility in the two directions of the horizontal and vertical coordinate axes, generating a key point vector map.

3. CV and Image Segmentation of Algorithms

CV technology is generally an intelligent technology in modern society that utilizes professional computer software to transform, analyze, and process images. This technology can convert unprocessed images into the desired image format for users [13-14]. In the process of capturing human motion patterns in aerobics in this article, CV technology can be used to adjust the grayscale values of key frame images in aerobics videos. This adjustment is different from the grayscale adjustment in traditional image processing, which can convert the grayscale value of the image into the corresponding quantization value. Therefore, this CV technology also has better performance and accuracy in the process of image analysis and processing, which can convert relatively poor quality images of the original target into clearer images of the target, intuitively displaying the motion form of the target.

The image segmentation algorithm is the fundamental technology in CV technology, and image segmentation can adjust the grayscale, spatial texture, and feature values in the image according to the needs of the user. This adjustment divides the image into disjoint image regions at multiple levels, thereby highlighting the target character features in the keyframe image [15-16]. Before the image segmentation algorithm processes the image, it is generally necessary to use other processing methods in image processing technology to preprocess the image. This processing generally involves processing multiple attributes such as grayscale and contrast of the image. This article first performs image histogram equalization on keyframe images extracted from multiple aerobics exercise videos, and the equalization results are shown in Figure 4.

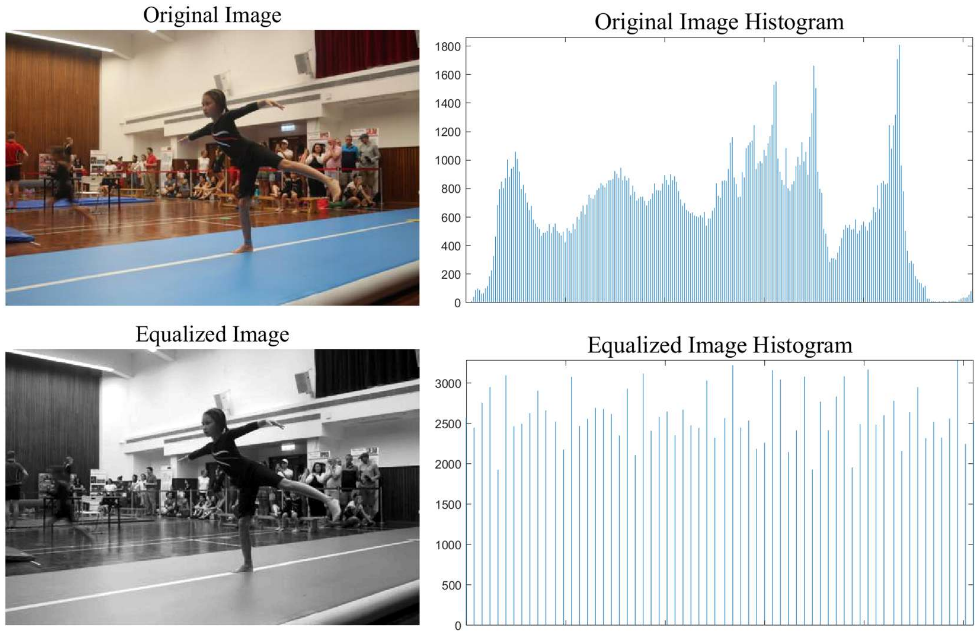
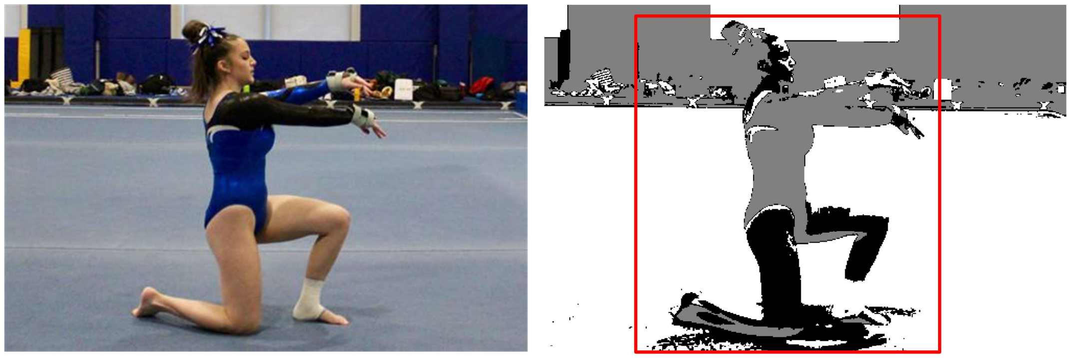


Fig. 4. Schematic diagram of key frame image histogram equalization of aerobics exercise video

In this Histogram equalization, it can be observed from the histogram of the original image that the contrast of the original image is mainly concentrated in the area of 100 to 200. After Histogram equalization, the contrast of the image is relatively evenly distributed in each part. The result of this distribution is that the dark part of the key frame image of the aerobics video is highlighted in the image. This also makes the dark details of the image easier to observe. At the same time, the image after Histogram equalization balances the contrast and brightness in the aerobics video key frame image, but the body part in the balanced key frame image is too bright. This is mainly because most of the pixels in the original aerobics keyframe image are concentrated in a small interval. The accumulation probability of pixels with larger pixel values in the keyframe image tends to a unified value, so the brightness of the processed dark pixels would be significantly improved. In addition, obvious edge phenomenon also appears in the darker image part of the original key frame image, mainly because the difference of various parameters of the key frame image after Histogram equalization is further increased after being mapped.

Then, the image segmentation mode based on K-means algorithm in image segmentation technology is used to segment the keyframe images extracted from aerobics exercise videos. The image segmentation effect based on K-means algorithm is shown in Figure 5.



The original image K-means custering image segmentation

Fig. 5. Schematic diagram of image segmentation effect based on K-means algorithm

In the process of using the K-means algorithm to segment keyframe images, it is necessary to simultaneously analyze the numerical performance of four aspects: Peak Signal to Noise Ratio (PSNR) of keyframe images, structural similarity of keyframe images, target clarity, and target feature values to analyze the image segmentation results. This analysis can evaluate the effectiveness of image K-means segmentation clustering algorithm, and the evaluation of image segmentation results is shown in Table 1.

Table 1. Effect evaluation table of image k-means clustering segmentation mode

Metrics\Images	The original image	Segmented image
PSNR (dB)	17.82	21.36
Structural similarity	0.61	0.79
Target clarity	0.73	0.65
Target eigenvalue	1.34	1.95

The PSNR in Table 1 mainly represents a ratio of the noise power that affects the quality of image segmentation. The unit of this ratio is generally expressed in decibels, while structural similarity, target clarity, and target feature values are only measurement indicators, and there is currently no representation unit. According to the data analysis in Table 1, it can be seen that after K-means clustering image segmentation, the PSNR, image structure similarity, and character feature values of the aerobics key images have been enhanced to some extent. This enhancement indicates that the image segmentation effect is relatively excellent. However, the clarity of the character target in the image is further reduced, which greatly reduces the intuitive resolution of the image target human body.

In calculating the structural similarity between PSNR and image segmentation, formulas (6) and (7) are mainly used to calculate the numerical values of PSNR and structural similarity. Among them, Formula (6) is mainly used to calculate PSNR.

$$P = \log_{10}\left(\frac{L^2}{M}\right) \tag{6}$$

L represents the maximum intensity of the noise power of the image, while M represents the Mean squared error. Next, calculate the image similarity, which generally calculates the similarity between two images, as shown in Formula (7).

$$S = \frac{2\sigma_n + d_2}{\sigma^2 + d} \tag{7}$$

Then, image segmentation was performed on the keyframes in the aerobics exercise video using both region segmentation and threshold segmentation modes. The results of these two image segmentation methods are shown in Figure 6.

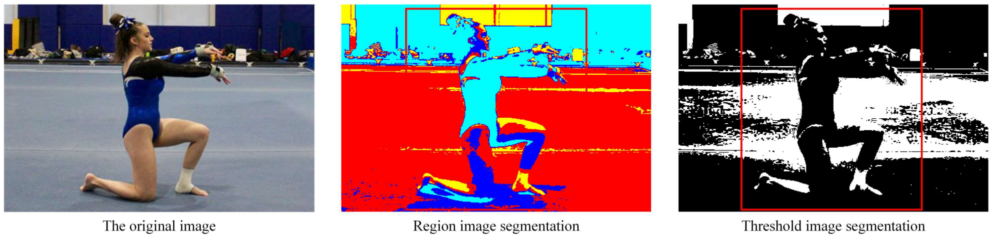


Fig. 6. Schematic diagram of the image segmentation effect of the key frames in the aerobics video based on the region segmentation and threshold segmentation modes

In an image, it is generally composed of regions with different characteristics, which satisfy special consistency. Based on the consistency of the image regions, prescribed judgment methods can be used to divide different regions in the image. Region-based image segmentation assigns specified attributes to different pixels of keyframe images in aerobics videos. Then, pixels with the same properties can be connected according to the criteria of the image segmentation algorithm to form an image segmentation region.

Finally, six movements in the 50 aerobics action images extracted in this paper were captured in motion capture of human motion patterns, and the visibility of the images was enhanced by first performing histogram equalization on the six key frame images. Then, it is possible to capture and mark the relatively clear points of the 12 human body marker points of aerobics athletes. This type of marking can help aerobics raters to rate the performance of athletes in the process of aerobics more quickly and accurately. The key frame image recognition of aerobics athletes including 6 types of aerobics movements is shown in Figure 7.

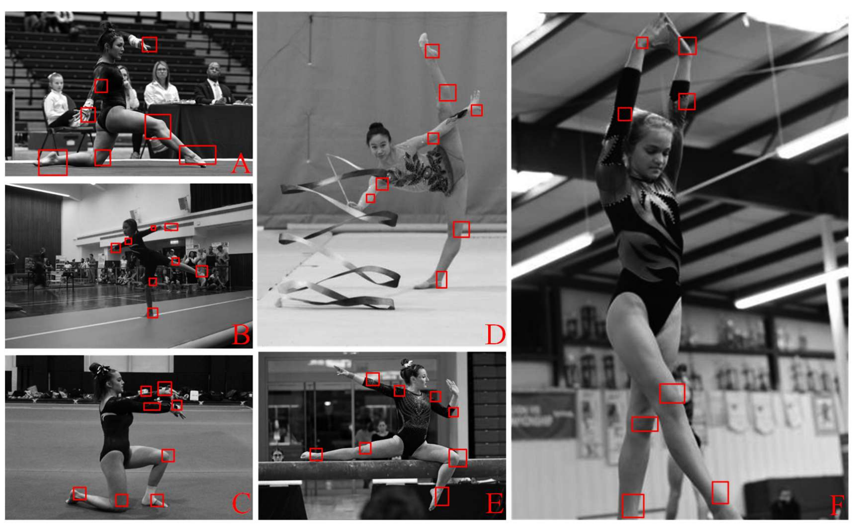


Fig. 7. Schematic diagram of key frame image recognition of aerobics athletes including 6 types of aerobics movements

In this paper, after image enhancement, the six types of movements are numbered from A to F. In the process of motion capture, the commonly used optical unmarked human motion form capture technology 1. This article combines CV technology and image segmentation algorithm to analyze the recognition quantity of 12 main joints in the human body in these 6 types of movements using human motion shape capture technology 2, as shown in Table 2.

Table 2. Joint recognition performance of optical markerless human motion capture technology and motion shape capture technology of CV and image segmentation algorithm

Action type\Technology type	Number of joints (piece)	Motion capture technology a (piece)	Motion capture technology b (piece)
A	12	8	12
B	8	4	6
C	11	6	10
D	20	6	9
E	12	5	10
F	10	6	7

From the data in Table 2, it is evident that the human motion form capture technology combined with CV technology and image segmentation algorithm has relatively better performance in identifying and capturing the main joints of aerobics athletes. This advantage mainly comes from a series of image processing operations on key frame images of aerobics athletes using image segmentation technology, which enables human motion form capture technology to recognize and capture athletes' joints in relatively blurry areas. Especially when the color and background color of the joints in the gymnastic clothing of aerobics athletes tend to be similar, the commonly used optical unmarked human motion form capture is difficult to recognize and capture the joints of aerobics athletes. Finally, it was determined that the commonly used optical unlabeled human motion form capture technique a in Figure 7 identified a total of 35 joints for aerobics athletes in 6 images. The number of joints recognized by the human motion shape capture technology b, which combines CV technology and image segmentation algorithms, was 54 for aerobics athletes. The total number of joints that can be recognized in these 6 images, except for those obstructed by the viewing angle, was 73 (image D contains the recognition of two aerobics athletes in total). From this, it can be determined that the recognition effect of optical unmarked human motion capture technology was approximately 47.95%. The recognition effect of human motion capture technology combining CV and image segmentation algorithms was approximately 73.97%, which confirms that the combination of CV technology and image segmentation has improved the comprehensive performance of human motion capture technology by 26.02%.

4. Conclusions

This article explained the application of CV technology and image segmentation algorithms in current society, and determined the feasibility and reliability of these two technologies in human motion form capture schemes. This article mainly analyzed the application of a human motion form capture scheme that combined CV technology and image segmentation algorithm in aerobics. This analysis can further demonstrate the degree of performance optimization in various aspects of the human motion capture scheme that integrates CV technology and image segmentation. In aerobics sports, this human motion form capture scheme that combines CV technology and image segmentation can be widely applied in aerobics action standard scoring and aerobics action standard training. CV technology is mainly responsible for regulating the perspective of image acquisition in the capture scheme of the target human body's motion form in aerobics. This regulation can draw a bone map of the 12 main joints of a character. This type of human labeled bone map can help capture motion patterns and quickly recognize the human body in key frame images of aerobics videos. The image segmentation algorithm mainly processes the key frame images of aerobics in the form capture scheme of aerobics, highlighting the human body feature values in the images and assisting machines in recognizing the human body in the images. From this, it can be determined that the combination of CV technology and image segmentation algorithms has a good optimization effect in capturing human motion morphology. It can improve the accuracy of image processing during the process, while also enhancing the athlete's movement standards in daily training during sports activities.

References

[1] Alexander Mathis, Steffen Schneider, Jessy Lauer, and Mackenzie Weygandt Mathis. "A primer on motion capture with deep learning: principles, pitfalls, and perspectives." Neuron 108.1 (2020): 44-65.

[2] Eline, Van der Kruk, and Marco M. Reijne. "Accuracy of human motion capture systems for sport applications; state-of-the-art review." European journal of sport science 18.6 (2018): 806-819.

[3] Xiaowei Zhou, Menglong Zhu, Georgios Pavlakos, Spyridon Leonardos, Konstantinos G. Derpanis, and Kostas Daniilidis. "Monocap: Monocular human motion capture using a cnn coupled with a geometric prior." IEEE transactions on pattern analysis and machine intelligence 41.4 (2018): 901-914.

[4] Gergely, Nagymate, and Rita M. Kiss. "Application of OptiTrack motion capture systems in human movement analysis: A systematic literature review." Recent Innovations in Mechatronics 5.1 (2018): 1-9.

[5] Kang Chen, Yupan Wang, Song-Hai Zhang, Senzhe Xu, Weidong Zhang, and Shi-min Hu. "MoCap-Solver: A neural solver for optical motion capture data." ACM Transactions on Graphics (TOG) 40.4 (2021): 1-11.

[6] Gustavo Ramos Dalla Bernardina, Tony Monnet, Heber Teixeira Pinto, Ricardo Machado Leite de Barros, Pietro Cerveri, and Amanda Piaia Silvatti. "Are action sport cameras accurate enough for 3d motion analysis? A comparison with a commercial motion capture system." Journal of Applied Biomechanics 35.1 (2019): 80-86.

[7] Mreza Naeemabadi, Birthe Dinesen, Ole Kæseler Andersen, and John Hansen. "Influence of a marker-based motion capture system on the performance of Microsoft Kinect v2 skeleton algorithm." IEEE Sensors Journal 19.1 (2018): 171-179.

[8] Steffi L. Colyer, Murray Evans, Darren P. Cosker, and Aki I. T. Salo. "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system." Sports medicine-open 4.1 (2018): 1-15.

[9] K. I. S. H. A. N. P. R. A. S. A. D., Gunale, and Prachi Mukherji. "Indoor human fall detection system based on automatic vision using computer vision and machine learning algorithms." J. Eng. Sci. Technol 13.8 (2018): 2587-2605.

[10] Cong, Zhang, Miao Wang, and Limin Zhou. "Recognition method of basketball players' throwing action based on image segmentation." International Journal of Biometrics 15.2 (2023): 121-133.

[11] Magdalena, Mazur-Milecka, and Jacek Ruminski. "Deep learning based thermal image segmentation for laboratory animals tracking." Quantitative InfraRed Thermography Journal 18.3 (2021): 159-176.

[12] Hadish Habte Tesfamikael, Adam Fray, Israel Mengsteab, Adonay Semere, and Zebib Amanuel. "Simulation of eye tracking control based electric wheelchair construction by image segmentation algorithm." Journal of Innovative Image Processing (JIIP) 3.01 (2021): 21-35.

[13] Ben G, Weinstein. "A computer vision for animal ecology." Journal of Animal Ecology 87.3 (2018): 533-545.

[14] Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, et al. "Attention mechanisms in computer vision: A survey." Computational Visual Media 8.3 (2022): 331-368.

[15] Sohom Ghosh, Nibaran C Das, Ishita Das, and Ujjwal Maulik. "Understanding deep learning techniques for image segmentation." ACM Computing Surveys (CSUR) 52.4 (2019): 1-35.

[16] Xin Zheng, Qinyi Lei, Run Yao, Yifei Gong, and Qian Yin. "Image segmentation based on adaptive K-means algorithm." EURASIP Journal on Image and Video Processing 2018.1 (2018): 1-10.