2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI) | 979-8-3503-0115-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/ACEDPI58926.2023.00035

A Method of Yoga Action Pattern Recognition Based on Computer Vision Technology

Ling Ma[a], Chao Huang[b,*]
Wenhua College, Wuhan, Hubei, 430074 China
[a]165576803@qq.com, [b]975044561@qq.com
*Corresponding Author

***Abstract*—Object detection and recognition using computer vision has been a very interesting and challenging research field in the past three decades. Classification and target location based on machine learning and computer vision technology has always been a hot topic, and great achievements have been made. Firstly, the classification analysis method is used to analyze and study the yoga movement and theory based on computer vision technology, including human eye tracking and recognition, face recognition, head movement tracking and recognition, gesture recognition and posture recognition. Based on the task of yoga movement pattern recognition, a yoga movement pattern recognition method based on computer vision technology is proposed. The improved model is based on the framework and structure of the network. A certain number of candidate regions are proposed and classified through feature extraction, and then these regions are output as the detected bounding box. The posture action diagram of yoga, that is, ordinary RGB color image, is collected by RGB camera, and the bone data is extracted from RGB image by bone extraction model. The RGB image data and bone data are input into the joint model, and the joint model will output the category of Yoga action and the score of this action.**

Keywords—Computer vision technology, yoga; movement patterns, recognition methods

I. INTRODUCTION

Yoga can reduce anxiety and depression, and improve various symptoms including psychological and pain syndrome, musculoskeletal and nervous system diseases, autoimmune and immune syndrome. As a result, this fashionable fitness exercise has been widely carried out in the elective courses of physical education in universities and colleges, and it is very popular among students [1]. With the rapid development of artificial intelligence technology, many new frontier technologies are widely popularized and applied in daily life. Human posture recognition, human motion recognition, human motion detection and other technologies have broad application prospects in human motion behavior analysis, medical rehabilitation training, physical education and other aspects [2]. Every day in the internet, gigabytes of images are produced, which contain a lot of information. We need to process these images in order to browse and retrieve them effectively and use the information in them. The progress of image retrieval is also inseparable from the development of database management system and the effective promotion of computer vision [3]. Most of the information received by human beings comes from vision, and image retrieval based on vision is particularly important. At present, all kinds of human motion recognition methods mainly focus on recognizing simple human body parts based on external features, but in the real environment, it is often necessary to recognize human motion in video, so the research on human motion recognition has strong practical significance. The motion recognition system based on computer vision involves many disciplines such as computer science, artificial intelligence, neurobiophysics, cognitive psychology, art, etc. In this paper, the exploration of a new man-machine interaction mode and the research of robot intelligence technology are bound to have different significance and influence on each of the above disciplines [4].

Vision is an important way for human beings to perceive the external world and obtain environmental information. Because of its unique space-time characteristics, visual perception plays an important role in human perception system, and the information obtained from the visual system is more intuitive and rich [5]. With the rapid development of Internet and multimedia technology, the amount of information in the real world is growing rapidly, but the ability of human beings to obtain information through vision for analysis and processing can not be improved. Computer vision refers to the use of computers to realize human visual function-the perception, recognition and understanding of three-dimensional scenes or objects in the objective world. Action recognition based on computer vision is to use computer vision technology to obtain images in the objective world, then use different algorithms for action recognition, and make the system make different responses according to the meaning of different actions. Through the way of computer vision, we can recognize the yoga movement pattern, track the movement of human joints, establish the geometric model of human body, and analyze the human movement, which is helpful to master the human movement state and improve the movement performance. As a pillar of the national economy and an important form of movement in the material production sector, it continues to make important contributions to the national economic development [6]. Action recognition based on computer vision is to let the machine match the video containing human actions with a predefined group of action category labels, so as to realize the automatic annotation and classification of video. Firstly, the collected color image is grayed, so as to reduce the amount of computer calculation. Then the background of the generated gray image is eliminated by Gaussian mixture model to obtain the black-and-white video image. After median filter denoising, the foreground target in the image is obtained [7].

II. COMPUTER VISION TECHNOLOGY

A. Fundamental

Computer vision refers to the perception, recognition and understanding of three-dimensional scenes in the

979-8-3503-0115-1/23/$31.00 ©2023 IEEE
DOI 10.1109/ACEDPI58926.2023.00035

objective world by using computers to realize human visual functions. Computer vision enables computer to recognize three-dimensional environmental information on the basis of two-dimensional image information, including sensing geometric information of objects in the environment, describing, storing, identifying and understanding these information. That is, the computer vision technology is used to acquire and recognize the actions of human bodies in the environment, and then make corresponding responses. Action recognition based on computer vision technology mainly uses computer vision technology to detect the moving area of human body and track the movement, and uses different action recognition algorithms to recognize and understand the meaning of yoga action. Figure 1 shows the general method flow of yoga action recognition.

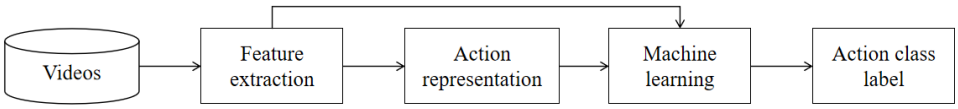

Figure 1. General method flow of yoga action recognition

Human motion recognition takes low-level video data as input, and finally outputs high-level semantic knowledge, so that computers have visual comprehension ability similar to human beings. The representation and modeling of human motion is a crucial step in the process of human motion recognition. The original video data contains a series of consecutive frames, and each frame is composed of RGB values of points. Only by properly mapping the original data can more refined and reasonable features containing information be formed [8]. Human structure model features can describe the human structure model by detecting and tracking the key parts of the human body, such as joints and extremities, and using corresponding parameters to describe the states of these parts. An important feature of human visual perception is that it doesn't immediately process the whole scene, but focuses on a certain part of the visual space, scans the images in a certain order, and moves from one area to another. Get the information of a certain local area at a certain moment, and then combine the information of these areas for overall judgment and feeling. Computer is used to simulate human visual function, including perception, recognition and understanding of the real world. Computer vision technology enables computers to recognize three-dimensional environmental information through two-dimensional images. Through the image acquisition card, the obtained image information is imported into the computer for storage. Then, the stored information is processed in the computer with the help of relevant image processing algorithms. Finally, the results are fed back visually.

B. Image Processing

This paper summarizes the research content of yoga movement pattern recognition based on computer vision technology, and uses the image semantic hierarchy model to re tier it. The intelligent recognition process of yoga movement information is divided into bottom visual layer, middle visual layer and high visual layer. The bottom visual layer includes a series of basic knowledge of image processing and recognition, such as moving target detection and segmentation, and extracts the common visual feature semantic information of images; The middle visual layer includes image processing knowledge such as motion tracking and matching to extract relevant motion semantic information; The high-level vision layer belongs to image understanding, including high-level image recognition knowledge such as target type recognition, behavior recognition, scene semantic understanding and description. Digital image processing is the process of converting image signals into digital signals and processing them by computer. Digital image processing refers to the processing of digital images with computers. For the input low-quality images, digital image processing carries out a series of transformations or processes such as denoising, edge processing, feature extraction, graying and image segmentation, so as to obtain new images with stronger brightness and higher quality or meet the needs of further research [9]. Digital image processing technology focuses on the transformation between images. Its main goal is to process digital images to improve the visual effect of images, and lay the foundation for subsequent high-level image analysis and understanding or automatic target recognition. Figure 2 shows the accuracy of image action recognition.

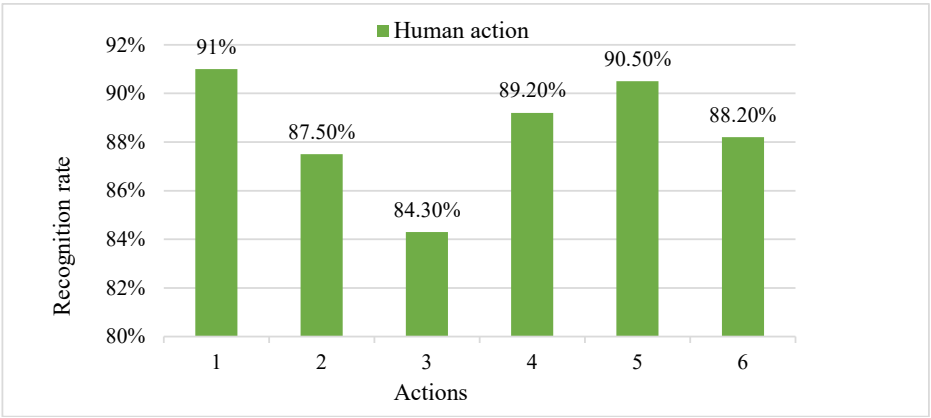

Figure 2. Image action recognition accuracy

Suppose an M * n image f (x, y) passes through a weighted mean filter w (s, t) of M * n (M and N are odd), and the processing process is:

$$g(x,y)=\frac{\sum_{s=-at=-b}^{a}\sum^{b}w(s,t)f(x+s,y+t)}{\sum_{s=-at=-b}^{a}\sum^{b}w(s,t)} \quad (1)$$

Where, a = (m-1) / 2, and B = (n-l) / 2; W (s, t) is the mask coefficient; The numerator is used as the template coefficient and the pixel block centered on (x, y) is convoluted; The denominator normalizes the weighting coefficient. A fully filtered image is obtained by performing on x = 0, 1, 2,..., M-1 and y = 0, 1, 2,..., N-1. The important preprocessing methods in image preprocessing can lay a solid foundation for subsequent image segmentation, recognition and analysis.

III. Recognition method of yoga movement pattern

A. Yoga action recognition algorithm design

At present, with the continuous development of computer vision technology, multimodal recognition technology has become the most commonly used method in yoga gesture recognition. First of all, the RGB camera is used to collect yoga posture and action pictures, that is, ordinary RGB color images, then the RGB images are converted into bone images by the bone extraction model, and then the RGB images and bone images are input into the joint model. The joint model will output the categories of yoga actions. Table 1 shows the data set of yoga postures.

Table 1. Dataset of yoga poses

Action 1	Counts	Action 2	Counts	Action 3	Counts	Action 4	Counts
Mountain pose	400	Walking stick	400	Magic chair	400	Supine	400
	400		400		400		400
	400		400		400		400

When recognizing yoga movements, the data of two modalities, RGB data and bone data, are used. The data of bones is a matrix composed of Euclidean distances of joint points of each bone, as shown below.

$$G\_Data=\begin{bmatrix}\| J_2J_1 \|... \\ ... \quad ... \\ \| J_nJ_1 \|\| J_nJ_{n-1} \|\end{bmatrix} \quad (2)$$

Where J_n represents the coordinates of the nth joint point; $\| J_nJ_1 \|$ represents the Euclidean distance from the nth joint point to the first joint point. Because the classification needs to identify the types of yoga and the scores of yoga movements, it is necessary to design a multi-level model when designing the data input of the joint model. The format of the input data is as follows:

$$D=\{Y_1,\ Y_2,...,Y_n\}$$
$$Y_i=\{S_1,S_2,S_3\}$$
$$i=(1,2,...,n) \quad (3)$$

Among them, yn represents the type of yoga movements; S_1, S_2 and S_3 represent excellent, good and medium respectively. D data has two labels, namely category label and rating label. In order to verify the effect of the model, the training data set is sent into the bone extraction model, and the corresponding bone posture is extracted. At this time, training data are divided into RGB data and bone data. RGB data and bone data are processed according to the same proportion to generate multimodal data. Table 2 shows the comparison of data results.

Table 2. Comparison of data results

Modules	RGB data					
	Action	Score	Action	Score	Action	Score
Mountain pose	100%	76.4%	97.8%	88.2%	100%	92.4%
Walking stick	100%	74.3%	86.5%	76.1%	100%	88.6%
Magic chair	100%	75.1%	93.2%	86.3%	100%	89.2%
Supine	100%	75.5%	86.7%	75.7%	100%	88.4%

The comprehensive evaluation index of design yoga posture recognition is W. W is obtained according to the judgment of yoga posture category and score. The calculation formula of each model W is as follows:

$$W=\frac{1}{n}\sum_i A_iS_i$$
$$(i=1,2,...,n) \quad (4)$$

In which, n stands for yoga category; The accuracy rate of the category that A_i represents the class i yoga

posture; S_i represents the accuracy of the score of yoga posture of class i. When the gray value of image pixels is 255, it can be regarded as a moving target point. Then we analyze the connectivity of the binary image R_n to get the final image R_n.

$$R_n(x,y)=\begin{cases}255, D_n(x,y)>T\\ 0, \text{other}\end{cases} \quad (5)$$

Synthesize the W of each model, as shown in Figure 3.

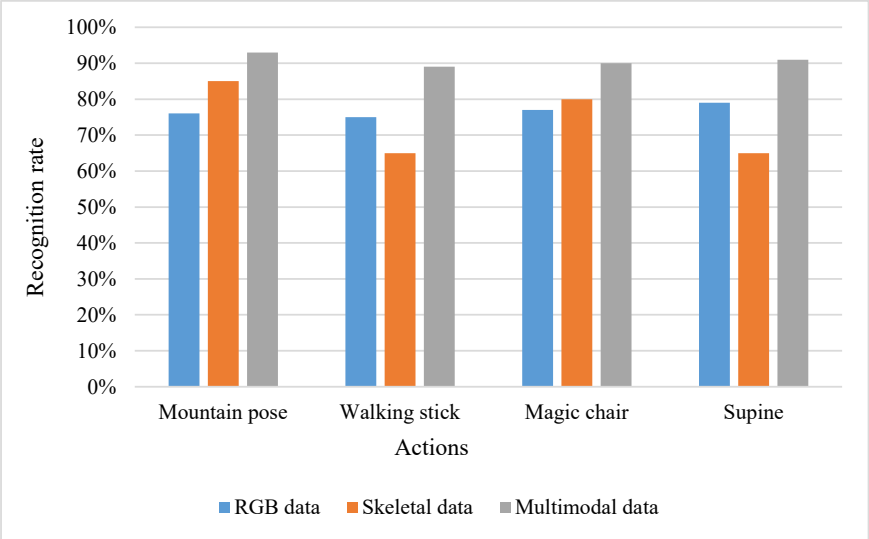

Figure 3. Recognition rate of each action pattern

Multi-modal data shows great advantages in the joint model, which combines the advantages of RGB data and bone data. At the same time, one model is used to complete the recognition task, which can quickly and accurately complete the classification and grading of yoga movements.

B. Realization of yoga action pattern recognition

Computer vision technology includes many algorithms that can understand images (including pictures and videos). It is a lot of innovative key technologies. The improved network based on computer vision technology has stronger learning ability, can segment the target region more accurately, improve the segmentation accuracy, and better fit and predict the segmentation edge and target edge. Yoga movement recognition based on computer vision technology is mainly divided into four stages: data acquisition, feature extraction and movement representation, abnormal data detection and repair, and movement classification and recognition. The methods in each stage are coordinated. Data acquisition is the basis of feature extraction, action classification and recognition. The quality of the collected action data directly affects the training and model quality of the subsequent action recognition model. The quantity, clarity and diversity of the collected data will directly affect the subsequent labeling and identification work [10]. Feature extraction is an important link connecting the two stages of data acquisition and entity recognition model training and evaluation that is, labeling the target entity in the collected two-dimensional color image to support the training of entity recognition model. Abnormal data detection and repair is to detect and repair the abnormal data formed due to joint point misidentification in the feature extraction stage, so as to ensure the quality of the data. Action classification and recognition is the final stage of this method, and the recognition effect is closely related to the quality of feature data formed in the first three stages. This stage includes two processes: model training and call, and model evaluation, so as to realize the classification and recognition of various actions. However, in practical application, due to the differences in application requirements, shooting equipment and conditions, there may be differences between the actual action video and the corresponding action video in the database used for training. For such problems, we call the marked database the source domain, and the target action data whose distribution is not completely consistent with it the target domain. Conduct cross domain action recognition, eliminate and mitigate the impact of inconsistent distribution, and migrate the existing action recognition knowledge in the source domain to the target domain. Under the framework of the algorithm, the description of human action crosses the obstacles of perspective, and realizes the purpose of information sharing from different perspectives.

IV. Conclusions

With the continuous development of computer technology, intelligent machines and equipment have gradually attracted extensive attention of human beings, and at the same time, the technology of human motion recognition has also been paid more and more attention. At present, all kinds of human motion recognition methods mainly focus on recognizing simple human body parts based on external features, but in the real environment, it is often necessary to recognize human motion in video, so the research on human motion recognition has strong practical significance. This chapter summarizes and explains the research background and significance of yoga movement recognition and target detection, and introduces the basic structure of the algorithm used in this paper, as well as the basic techniques of training optimization. Based on the

theoretical basis of computer vision technology, combined with the current related network technologies and algorithms, and combined with the recognition and detection network, this paper realizes the yoga movement recognition and detection. The research of human motion recognition based on computer vision is in the ascendant, full of challenges and opportunities, and has a broad application prospect. Computer science has a long way to go before the ultimate goal of "achieving human visual cognitive ability".

ACKNOWLEDGEMENTS

Experimental Research on Mixed Teaching in University Yoga Elective Course.

REFERENCES

[1] Luo Guoqiang, Li Jiahua, Zuo Wentao, et al. Research on the steps and methods of gesture recognition based on computer vision technology. Wireless Internet Technology, vol.17, no.3, pp.2, 2020.

[2] Zhang Feiyun, Zhou Shuai. Intelligent recognition of wrong actions of sports players based on computer vision technology. Adhesion, vol.44, no.12, pp.4, 2020.

[3] Mao Zhengchong, Han Yi. Research on the Application of Computer Vision Technology in Vehicle Recognition. Application of Single Chip Microcomputer and Embedded System, vol.17, no.6, pp.6, 2017.

[4] Mao Xia, Wang Lan, Li Jianjun. A Human Action Recognition Framework Based on RGB-D Feature Fusion. Computer Science, vol.45, no.8, pp.6, 2018.

[5] Cen Hong. Research on Image Recognition and Restoration of Computer Vision Technology. Computer Programming Skills and Maintenance, no.12, pp.3, 2020.

[6] Zhang Shujun, Lan Shanzhen, Bu Qi, et al. A brief description of action recognition method based on deep learning. Journal of Communication University of China: Natural Science Edition, vol.26, no.5, pp.7, 2019.

[7] Chen Zhuoyan, Guo Jianliang, Chen Yongqing, et al. An automatic detection system for bicycle pedal pairing based on computer vision. Journal of Ningbo Institute of Technology, vol.34, no.1, pp.7, 2022.

[8] Wang Quansheng. A method to improve the accuracy of computer vision recognition of red. Computer Knowledge and Technology: Academic Edition, vol.17, no.25, pp.3, 2021.

[9] Bi Xuechao. Research on dance video action recognition technology based on computer vision. Electronic Design Engineering, vol.28, no.7, pp.5, 2020.

[10] Luo Mingming, Zhu Feng, Wang Dongsheng. Research Review of Automatic Recognition Method of Ship Image Based on Computer Vision. Software Guide, vol.17, no.7, pp.4, 2018.