Moving Target Location Method of Video Image Based on Computer Vision

Xiao-xia Li and Hai-yan Zhang(✉)

Huali College Guangdong University of Technology, Guangzhou 511325, China
dadawd654651@163.com

Abstract. By the localization and recognition of human moving target in video image, combined with the information of human motion feature in video image, the moving target localization and visual reconstruction is realized, this paper analyzes the feature quantity of moving objects in video image, improves the training level, and proposes a moving objects positioning technology of video image based on computer vision and 3D feature point reconstruction. According to the moving feature position of human body, the 3D information modeling and image acquisition of moving target is carried out by using video information acquisition and spatial feature scanning methods. The moving feature points of the collected moving target video image are calibrated and arranged, and the 3D edge outlines feature point set of human skeleton is extracted and represented as a high dimensional vector to form the regular feature database of moving target video image. The moving points in the regular feature database of moving target video image are fusion to realize the reconstruction of moving target video image and the location of moving target. The simulation results show that the method has good real-time and accuracy in moving target location of video image, has strong ability of 3D marking of human moving points, and has high accuracy of extracting moving motion features.

Keywords: Computer vision · Video image · Moving target · Location

1 Introduction

The movement away human body is random, scattered and nonlinear. In order to improve the ability of quantitative analysis of human motion and realize the scientific guidance of human motion training, it is necessary to reconstruct the three-dimensional image of the feature points of human motion [1]. With the development of video image tracking and scanning technology, the video image tracking method is used to reconstruct the feature points of human motion, to overcome the randomness and nonlinearity of human action, to avoid the problems of inaccurate, reconstruction of stiffness and rough edges of human motion, to use the video image tracking method of moving target feature points of video image to reconstruct the human motion, and to recognize the human action by video image tracking technology. To depict the regular characteristics of human motion from the process of human motion, so as to improve the scientific nature of human motion training, the study of video image moving target feature point tracking method has attracted great attention on experts and scholars.

Y.-D. Zhang et al. (Eds.): ICMTEL 2020, LNICST 326, pp. 182–192, 2020.
https://doi.org/10.1007/978-3-030-51100-5_16

In the traditional method, the tracking technology of the moving target characteristic point of the video image mainly comprises a human body motion action characteristic fusion scanning method based on a point cloud technology, a gesture information quantitative tracking identification method and a motion tracking identification method based on the computer vision feature extraction. The moving target feature quantity of the video image is analyzed by the method of image processing and attitude sensing and the like, the data mining and information fusion of the human body motion point are carried out, the sports skill and the level are improved, the relevant documents are researched and certain motion guidance and action correction performance are obtained, in which, the reference [2] method uses matrix f norm constraints for the residual term to solve the orthogonal prucker regression model, making the model very sensitive to some noise (e.g., illumination). Replace the original matrix f norm constraint with a more robust 1 norm constraint and propose a sparse orthogonal pruck regression model. The model can be solved by an effective alternating iterative algorithm. The experimental results show that the model can deal with the change of face attitude effectively. However, the accuracy of this method is general, and the accuracy of 3D marking and feature extraction is not high.

In reference [3], the method of facial expression modeling based on interval algebra bayesian networks is able to capture not only the spatial relation of the face, but also the complex temporal relation of the face, so that the face expression can be recognized more effectively. The method can improve the speed of training and recognition by using only tracking-based features and not manually marking peak frames. However, the accuracy of this method is general, and the accuracy of 3D marking and feature extraction is not high. the three-dimensional attitude information acquisition and three-dimensional reconstruction method of the motion characteristic is presented in the document [4], and the three-dimensional laser scanning method is adopted to acquire the motion characteristics of the human body target, and the tracking system of the target action is constructed, the method carries out the batch read-in of the human body motion characteristic information by using a sampling point noose interpolation method, realizes the three-dimensional information reconstruction and volume rendering of the human body dynamic information, improves the accuracy of the tracking and identification of the human body motion point, but the method cannot precisely register the control point when being interfered by the large motion characteristic, so that the dynamic accumulation of the motion action characteristic points is caused, so that the calculation cost is increased; the method for establishing the three-dimensional human body model of an athlete based on the laser scanning technology is proposed in the document [5], the human body model is established on the basis of the human skeleton model, the three-dimensional human body model reconstruction is carried out by the least square regression expansion algorithm, and the three-dimensional scattered video image motion target positioning is realized, and the problem of the method is that the tracking and identification accuracy of the high-dimensional motion pose information is not high [6].

In this paper, a moving target location technique for video image based on computer vision and 3D feature point reconstruction is proposed. Firstly, the 3D information modeling and image acquisition of moving targets is carried out by using video information acquisition and spatial feature scanning methods. According to the moving

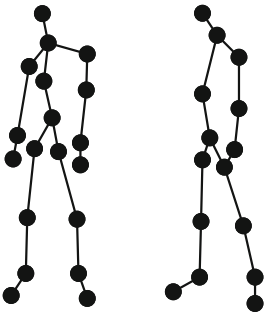
feature parts of human body, the moving feature points of the collected moving target video image are calibrated and arranged, and the 3D edge outlines feature point set of human skeleton is extracted and represented as a high dimensional vector to form a regular feature database of moving target video images. Then, the moving points in the regular feature database of moving target video image are fusion, and the moving target videos image reconstruction and moving target location are realized. Finally, simulation experiments are carried out to show the superior performance of this method of improving the accuracy of moving target location and recognition in video images.

2 Three-Dimensional Information Modeling of 3D Scanning Motion Target in Visual Space

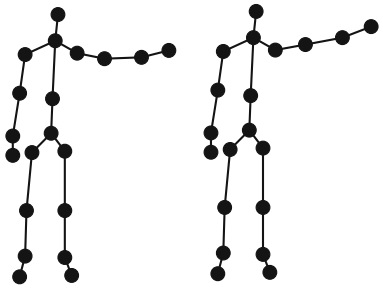
2.1 Image Acquisition Based on 3D Information Modeling of Moving Target

In order to realize video image tracking and recognition of moving target feature points in video image, firstly, the method of video information acquisition and spatial feature scanning is used to model the three-dimensional information of moving target, and through the three-dimensional image acquisition of human motion action, combined with the collected images, the motion feature points are calibrated and processed to realize the information fusion and tracking recognition of human body motion points [7]. The three-dimensional tracking and scanning method of visual space is used to obtain the dynamic information on human body, and the two-dimensional action manifolds analysis models on human motion point is constructed. The human motion studied in this paper mainly include standing, upper limb motion and lower extremity movement, and the three-dimensional scattered points of human motion point are calibrated, as shown in Fig. 1.

According to the calibration result of the human body motion characteristic point given in Fig. 1, the human body are characterized by a human skeleton, the visual space scanning image output path of the joint nodes u to v is represented by the $G = (V, E)$, taking the three-dimensional visual space scan image of the human body as the pixel sequence, $d_G(u,v)$, and input the original pixel feature data of the visual space scan, the formula $u^{(2)} = \left(u_1^{(2)}, u_2^{(2)}, u_3^{(2)}, u_4^{(2)}\right)$ it can be used to calculate the target sequence of human motion feature $u = (y_0, z_0, \lambda, \phi)^T$. The three-dimensional information modeling axis of the moving object can be calculated by using the least square method, the coordinate values, the $u^{(3)} = \left(u_1^{(3)}, u_2^{(3)}, u_3^{(3)}, u_4^{(3)}\right)$, and the direction of the vector $\overrightarrow{a} = \left(\cos u_3^{(3)} \cos u_4^{(3)}, \sin u_3^{(3)} \cos u_4^{(3)}, \sin u_4^{(3)}\right)$ along the axis of the human motion attitude distribution are calculated, thereby realizing the image acquisition of the three-dimensional information modeling of the moving target, it also provides accurate data input and entry for tracking feature points of moving objects in video images [8].



(a) Standing and lower limb movement



(b) Upper extremity exercise

Fig. 1. Calibration of motion characteristic points of human motion points

2.2 Calibration and Arrangement of Human Motion Feature Points

According to the motion characteristic part of the human body, the acquired moving target video image is calibrated and arranged, the action point inverse mapping on the two-dimensional manifold is realized, The three-dimensional model of cylindrical surface fitting algorithm is established, which can describe the coordinate position of visual space in spatial coordinate system $p^* = \left(X^{(cs2)}, \theta^*, \rho^*\right)$. The three-dimensional information modeling on the moving object of the three-dimensional model can be obtained by using the cylindrical surface fitting algorithm:

$$(\theta^e, \rho^e) = EFA(\theta^*, \rho^*) \tag{1}$$

Set the edge pixel set of the three-dimensional image, the line of the cylindrical surface of the motion space distribution is (θ^e, ρ^e) and the bus is parallel to the x axis. The following formula can be used to describe all the action vectors in the high-dimensional Euclidean space:

$$\left.\begin{array}{r} EX^{(cs2)} = \{x | x \in [0, h]\} \\ EY^{(cs2)} = \rho^e \cos\theta^e \\ EZ^{(cs2)} = \rho^e \sin\theta^e \end{array}\right\} \tag{2}$$

The set matrix which represents the characteristic points of human action is obtained by formula (2), and the X-axis direction translation is carried out with $\rho^e - R$ as the unit of displacement, and the three-dimensional manifolds distribution describing n human body cohesion actions is obtained as shown in Fig. 2.

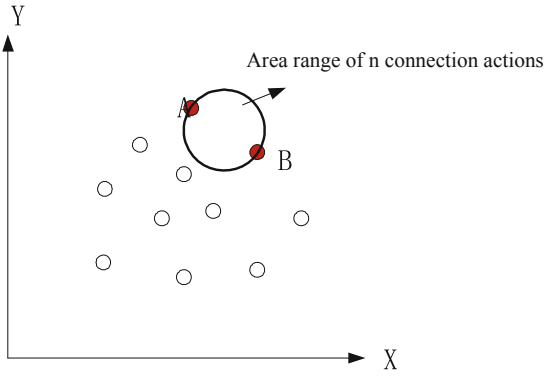


Fig. 2. Three-dimensional manifold distribution of human connection action

The training set of the human body motion characteristic part is extracted from the high-dimensional Euclidean space, a joint action vector is constructed in the area of the motion distribution range of the human body, and the feature values of the moving target feature points in the video image of M are small to small, and the human body movement feature points are shown in Fig. 3.

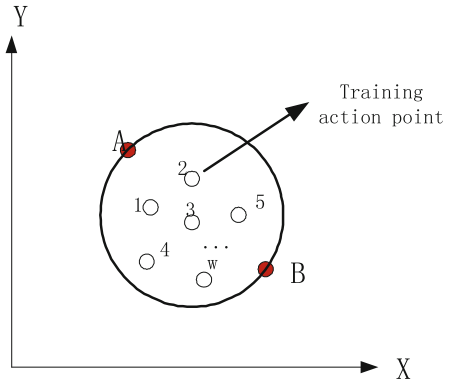


Fig. 3. Human motion feature point arrangement results

According to the arrangement result of Fig. 3, the maximum gray-level contour point mark is carried out, and the high-dimensional vector $I(i,j)$ of the moving point set of the three-dimensional human visual space scanning output is obtained as follows:

$$I(i,j) = \sum_{k=1}^{P} I_{(k)}(i,j) \times 2^{k-1} \tag{3}$$

Wherein, $I_{(k)}$ is a manifold vector of all three-dimensional scattered human moving sample points to a low-dimensional space, and the motion characteristic information is marked according to the three-dimensional edge contour feature point set of the human skeleton [9–11].

3 The Realization of the Moving Target of Three-Dimensional Scattered Video Image

3.1 Visual Space Fusion Processing of Moving Target Video Image

Based on the construction of human motion process and the extraction of moving feature points by using video information acquisition and spatial feature scanning method, the moving target location design of 3D scattered video images is carried out, and the regular feature database of moving target video image is formed by extracting the 3D edge outline feature point set of human skeleton [12]. The visual spatial information fusion method is used to track and recognize the moving points in the regular feature database of moving target video images. The 3D irregular point data onto human motion are regularized and fusion. After simple linear interpolation processing, the effective frames of the video image tracking images of moving human scattered points are obtained as:

$$\sigma(Z; D_X) = \sum_{i>j} \left| d_{ij}(Z) - d_X(x_i, x_j) \right|^2 \tag{4}$$

In the above formula, the $d_{ij}(Z)$ is the Euclidean distance from the pixel point, and $d_X(x_i, x_j)$ is the edge pixel point of the motion dynamic information registration. According to the characteristic space decomposition method, the motion action characteristic points of the three-dimensional human body model of the athlete are adaptively ordered, the template characteristic of the three-dimensional human body model is established, the acquired moving target video image is divided into pixel points. The method includes the following steps: obtaining a sub-block $M \times N$ of a human body motion characteristic point in a sub-block G of $G_{m,n}$ number 2-2, and carrying out discretization processing on the motion rules attribute of a human body motion point in

a two-dimensional quantization space [13], and obtaining the characteristic acquisition result of the motion characteristic point model of the athlete as follows:

$$G_{m,n} = \begin{pmatrix} g_{(m,n)}(1,1) & g_{(m,n)}(1,2) \\ g_{(m,n)}(2,1) & g_{(m,n)}(2,2) \end{pmatrix} \quad m = 1,2,\ldots,M; n = 1,2,\ldots,N; \tag{5}$$

Wherein:

$$g_{(m,n)}(u,v) = I_{(k)g}[2(m-1)+u, 2(n-1)+v] \tag{6}$$

In which, the $u \in \{1,2\}; v \in \{1,2\}$ represents the relevant factors of visual space fusion of moving target video image, and the three-dimensional visual space fusion method to conduct image retrieval. The posture and movement of the human body are constructed, and the second moment of image visual space fusion is obtained:

$$P_1 = \sum_{k=1}^{h} p_{(k)g}(i,j) \times 2^{k-1} \tag{7}$$

$$P_2 = \sum_{k=1}^{h} p_{(k)g}{}^{*}(i,j) \times 2^{k-1} \tag{8}$$

According to the multiple key points after image visual space fusion, the posture characteristics of human body are represented, which is used as a quantitative factor to realize video image tracking and recognition [14].

3.2 Moving Target Video Image Reconstruction and Video Image Tracking Recognition

The adjustment parameters of a three-dimensional scattered motion point are set up, and the visual reconstruction of the human motion attitude model are carried out by using the key point and frame point information feature matching method. The neighborhood characteristics of the original motion attitude data are used to identify the human action, and the inverse mapping of the human motion sample point is obtained as:

$$H(z) = P_1 \cdot \sum_{k=1}^{h} p_{(k)g}(i,j) \times 2^{k-1} / P_2 \cdot \sum_{k=1}^{h} p_{(k)g}{}^{*}(i,j) \times 2^{k-1} \tag{9}$$

The Euclidean distance between each video image tracking point and the pixel point is calculated. For k adjacent points, The autocorrelation feature matching method was used to obtain the 3D moving target video image reconstruction results:

$$x_i(t) = \sum_{k=1}^{p}\sum_{l=0}^{2}\varphi_{kl}[w_{i1}^{l},\cdots,w_{in}^{l}][x_1(t-k),\cdots,x_n(t-k)]^T - \sum_{k=1}^{q}\sum_{l=0}^{2}\theta_{kl}[w_{i1}^{l},\cdots,w_{in}^{l}] \tag{10}$$

Wherein, φ_{kl} is the stable pixel value in the process of optimizing the scattered points of human motion, θ_{kl} is the edge pixel set of k sample points, and $\varepsilon_i(t)$ represents the optimal reconstruction weight. For the human motion feature vectors of N input, the support vector machines classification method is used to classify the motion features. According to the video image tracking and recognition method, the tracking quantitative function of a moving action point is obtained as follows:

$$x_i(t) = [w_{i1}^{lk},\cdots,w_{in}^{lk}][x_1(t-k),\cdots,x_n(t-k)]^T \tag{11}$$

In which, $[w_{i1}^{lk},\cdots,w_{in}^{lk}]$ is a control kernel function of a three-dimensional edge contour feature point set, and $[x_1(t-k),\cdots,x_n(t-k)]^T$ is a motion action tracking control coefficient. The method comprises the following steps of: The moving points in the moving target video image regularization feature database are mapped to the low-dimensional space, performing feature compression, and reducing computational overhead, thereby obtaining a set of output samples of the video image tracking measurement of the video image moving target feature point as:

$$SSIM(x,y) = [l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[s(x,y)]^{\gamma} \tag{12}$$

The three-dimensional video images tracking information representing a human body motion point can be obtained by the above formula, and N motion identification elements are obtained in the feature space C, thereby realizing the three-dimensional scattered video image tracking control on the human body motion point [15].

4 Simulation Experiment and Result Analysis

In order to test the application performance of this method in video image tracking and recognition of moving target feature points in video image, simulation experiments are carried out. Matlab 2012 and Visual C simulation software are used for image processing and data analysis and analysis. The normal deviation of data sampling of moving action feature points is set to 0.21, and the allowable accuracy of the corresponding 3D reconstruction surface is 0.34. After sampling, the fitting parameter of point cloud is set to 12, the regression parameter of scattered point of human motion is set to 0.01, the adjacent point k of pixel partition is set to 12, the signal-to-noise ratio of 3D image acquisition in visual space is −12 dB, 302 × 250 pixel 3D mannequin image is used as the test set, according to the above parameter setting, the moving target feature point of video image is collected. The three-dimensional information modeling and image acquisition of moving target are realized, and the results are shown in Fig. 4.

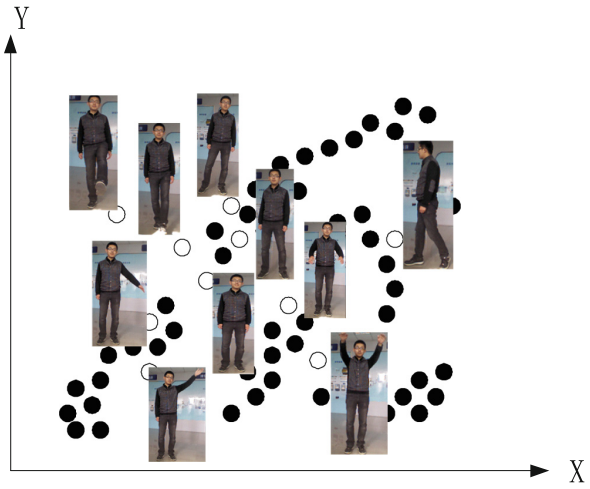


Fig. 4. Acquisition of the feature point of the moving target of the video image

The moving target feature point of the video image as given in Fig. 4 is a training set, and the human body motion feature point calibration and arrangement are carried out through the video image tracking technology, so that the three-dimensional visual space image reconstruction is realized, and the three-dimensional visual space reconstruction result of the human motion model is obtained as shown in Fig. 5.

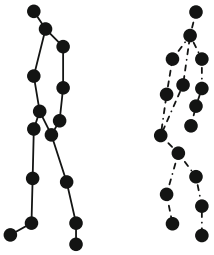


Fig. 5. 3D visual space reconstruction result of human motion model

The analysis of the results of Fig. 5 shows that the video image moving object feature point acquisition and the visual space image reconstruction are carried out by the method, the motion visual structure characteristic of the human body motion is effectively reflected, and the video image tracking recognition is carried out on the basis of the key motion feature point video image, The results are shown in Fig. 6.

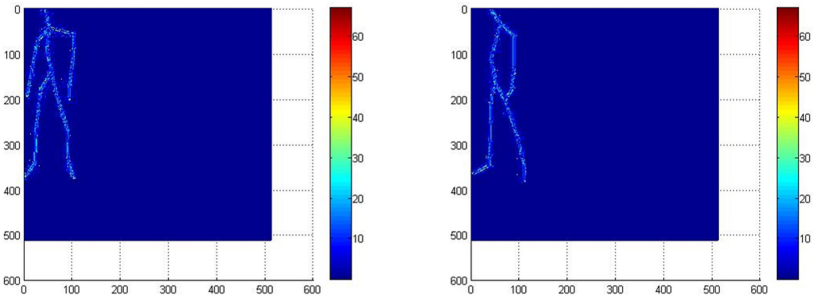


Fig. 6. Video image tracking recognition of human motion points

In order to quantitatively analyze the performance of the method in the realization of the tracking and recognition of human moving video images. Table 1 shows the comparison between the tracking offset, the calculation cost and the output signal-to-noise ratio of different methods. The method has the advantages of low offset of video image tracking, good description accuracy and short calculation time, so that the real-time and self-adaptive of the tracking is improved, the output signal-to-noise ratio is high, and the performance is better.

Table 1. Parameter performance comparison

Methods	Tracking offset/mm	Computational overhead/s	Output signal to noise ratio/dB
Proposed method	0.253	0.225	22.43
Spatial tracking method	2.922	12.264	9.46
Video tracking method	7.068	8.183	12.46

5 Conclusions

In this paper, the problem of tracking recognition and feature analysis of human motion points is studied. Combined with the human motion feature information in video image, the moving target location and visual reconstruction are realized, the moving target feature quantity of video image is analyzed, and the level of motion training is improved. A video image moving target location technology based on computer vision and 3D feature point reconstruction is proposed. The 3D information modeling and image acquisition of moving target are carried out by using video information acquisition and spatial feature scanning methods. The moving feature points of the collected moving target video image are calibrated and arranged. The 3D edge outline feature point set of human skeleton is extracted and represented as a high dimensional vector,

and the image fusion processing is carried out to realize the moving target video image reconstruction and moving target location. It is found that the real-time and accuracy of moving target location in video image is better, the deviation is low, and the adaptive ability is better.

References

1. Yan, S., Xu, D., Zhang, B., et al.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
2. Zhang, J.: Sparse orthogonal procrustes problem based regression for face recognition with pose variations. Comput. Sci. **44**(2), 302–305 (2017)
3. Qiu, Y., Zhao, J., Wang, Y.: Facial expression recognition using temporal relations among facial movements. Acta Electron. Sin. **44**(6), 1307–1313 (2016)
4. Li, S.T., Yin, H.T., Fang, L.Y.: Remote sensing image fusion via sparse representations over learned dictionaries. IEEE Trans. Geosci. Remote Sens. **51**(9), 4779–4789 (2013)
5. Li, B., Sang, J., Ning, J.: Analysis of accuracy in orbit predictions for space debris using semianalytic theory. Infrared Laser Eng. **44**(11), 3310–3316 (2015)
6. Yin, M., Liu, W., Zhao, X., et al.: Image denoising using trivariate prior model in nonsubsampled dual-tree complex contourlet transform domain and non-local means filter in spatial domain. Optik-Int. J. Light Electron Opt. **124**(24), 6896–6904 (2013)
7. Li, X., Gong, X.: 3D face modeling and validation in cross-pose face matching. J. Comput. Appl. **37**(1), 262–267 (2017)
8. Wang, H., Jin, H., Wang, J., Jiang, W.: Optimization approach for multi-scale segmentation of remotely sensed imagery under k-means clustering guidance. Acta Geodaetica et Cartographica Sin. **44**(5), 526–532 (2015)
9. He, G., Xiong, W., Chen, L., Wu, Q., Jing, N.: An MPI-based parallel pyramid building algorithm for large-scale RS image. J. Geo-Inf. Sci. **17**(5), 515–522 (2015)
10. Li, B., Wang, C., Huang, D.S.: Supervised feature extraction based on orthogonal discriminant projection. Neurocomputing **73**(1), 191–196 (2009)
11. Hou, C., Nie, F., Li, X., et al.: Joint embedding learning and sparse regression: a framework for unsupervised feature selection. IEEE Trans. Cybern. **44**(6), 793–804 (2014)
12. Cheng, M.M., Mitra, N.J., Huang, X., et al.: Global contrast based salient region detection. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 569–582 (2015)
13. Liu, N., Han, J.: DHSNet: deep hierarchical saliency network for salient object detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 678–686. IEEE Computer Society, Washington, DC (2016)
14. Kim, W., Kim, C.: Spatiotemporal saliency detection using textural contrast and its applications. IEEE Trans. Circ. Syst. Video Technol. **24**(4), 646–659 (2014)
15. Yan, Q., Xu, L., Shi, J., et al.: Hierarchical saliency detection. In: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013, pp. 1155–1162. IEEE Computer Society, Washington, DC (2013)