Method for extracting key features of sports wrong actions in teaching

Yunfei Lu[1]

Abstract. To improve the extraction accuracy of key features of wrong movements in physical exercise, an extraction method based on the key features of wrong movements in physical exercise in two stages of local extremum area was proposed. Firstly, the modeling principle of 3D vision inspection of wrong movements in physical exercise was introduced, and the 3D recognition model was constructed for wrong movements in physical exercise; secondly, the probability calculation was provided for each extremum area feature constructed with feature calculation, and the feature with the local maximum probability was selected as the output of the first stage and the input of the second stage so as to achieve the effective identification of the feature model; finally, the effectiveness of algorithm is verified with simulation experiment.

Key words. Local extremum, Two stages, Physical exercise, Wrong movements, Key fea-tures
.

1. Introduction

With the continuous promotion of technical level of physical exercise, the demonstration of right movements of physical exercise has been the key subject in the research in the field of physical education during the teaching process of physical exercise techniques. During the actual physical teaching process, as the cognition and understanding level of students varies obviously, partial students may be subject to more wrong movements, and they are slower than others in mastering the right physical exercise movements. In such cases, how to correct the wrong movements in the physical exercise has been the main problem to be solved in such field; the 3D visual-inspection modeling method for wrong movements in physical exercise can provide the 3D visual discrimination function for wrong movements in physical exercise; the establishment of 3D visual inspection model for wrong movements in physical exercise is the effective approach to the problem mentioned above, which has attracted the attention of many experts and scholars. As the 3D visual inspection model of wrong movements in physical exercise is of deep and long development

---

[1]Wuhan University of Science and Technology, Wuhan, Hubei, 430081

significance, it has been the key subject for the research of industry insiders, which has won the wide attention; and many better methods are also proposed.

The 3D visual-inspection modeling method based on the wrong movements in physical exercise of OpenCV is proposed in Literature [8]. Such method is used to acquire the outline information of movements in current physical exercise frame with the method of temporal difference; and the area beyond the physical-exercise movement outline of this frame of physical exercise image is used to update the wrong movements in physical exercise based on the unified training- smoothing so as to set up the 3D visual-inspection model for wrong movements in physical exercise. This method is simple, but subject to the problem of large error rate in modeling. Literature [9] is the discussion about the 3D visual-inspection modeling method for wrong movements in physical exercise based on Barnard algorithm. This method is applied to the extraction of feature points from the 3D visual physical-movement image of two adjacent physical movements, so that the sports field of sparse physical-exercise movements can be estimated in the 3D visual motion feature space; and the movement features extracted shall be subject to the C mean-value fuzzy clustering so as to use the result of clustering to set up the 3D visual-inspection model for wrong movements in physical exercise. This method is of strong expandability, however, in case the comparison model is set up with the current algorithm, the visual discrimination function for wrong movements cannot be set up, which is subject to large error in modeling. Literature [10] is the research about 3D visual-inspection modeling method for wrong movements in physical exercise based on 3D parameters. As for such method, the camera is used to collect the image sequence of physical movements for wrong movements in single-frame physical exercise, so that the knowledge marked in movements of physical exercise can be used to extract the features of 2D and 2D physical exercise actions and to establish the 3D visual-inspection model of wrong movements of physical exercise. The identification accuracy of such method is high, but the identification process is complex, and the identification error is large.

As for problems mentioned above, an extraction method of key features based on the wrong movements of physical exercise in two stages of local extremum region is proposed; the feature calculation shall be used in probability calculation for each feature of extremum region constructed, and shall also be used to identify the 3D model of wrong movements of physical exercise; then the performance verification shall be made through simulation.

2. 3D visual-inspection modeling principle for wrong movements in physical exercise

The objective is to set up a contour model for wrong movements in physical exercise of explicit-formulation integration; such model consists of three parts: in which, the human body parameter b_t and wrong-movement type vector s_t are subject to certain coupling relationship; another element is the geometric transformation parameter α_t.

The establishment of human body parameter of lower dimension means that two important problems should be solved: one is the selection of inserting method; the

other one is to select the embedding based on kinematical data (joint angle information), or to select the embedding based on visual data (outline) [7]. The method of spline fitting is applied to the parameterization of uniform average manifold of visual data (outline), as shown in Fig. 1. The $b_t \in \mathrm{R}^3$ represents the embedding-type coordinates of human body parameters at the time of t, then the uniform average manifold can be mapped into the 3D embedding space from the parameter space with one 1D parameter $\beta_t \in \mathrm{R}$ and spline fitting function $f : \mathrm{R} \to \mathrm{R}^3$. The form of mapping function is as follows:

$$y_t^k = \gamma_k (b_t) = C^k \cdot \psi (b_t) \ . \tag{1}$$

Where, $y_t^k \in \mathrm{R}^D$ is the wrong movement of No. k person at the time t, the b_t is the point corresponding to manifold embedding; C^k is the coefficient matrix mapped of $D \times N$ dimension, which depends on the specific wrong movement of human body. The form of nonlinear function $\psi (\cdot) : \mathrm{R}^3 \to \mathrm{R}^N$ is as follows:

$$\psi (x) = [k (x, x_1) , \cdots k (x, x_N)]^T \ . \tag{2}$$

Where, $k (\cdot)$ is the kernel function, and $x_1, \cdots, x_N$ is the sampling point on manifold embedding. The function $\psi (\cdot)$ means the kernel mapping, which uses the N kernel function to make the manifold model fit in the embedding space; the Gaussian kernel function or thin plate spline function is applied hereof.
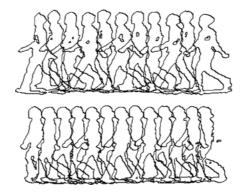

Fig. 1. Outline edge sequence of wrong movements

As for physical movement $1, \cdots, k$, the nonlinear mapping coefficient matrix $C^1, C^2, \cdots, C^k$ is given; the type parameter of wrong movement is mapped to the coefficient space through the unsymmetrical bilinear model:

$$[c^1, \cdots, c^k] = AS \ . \tag{3}$$

Where, c^k is the DN dimensional vector, which is the column vector representation of matrix C^k; matrix A is the coefficient matrix of $DN \times K$ dimension; the type matrix $S = [s^1, s^2, \cdots, s^k]^T$ contains the orthogonal matrix of type vector of wrong movement. The decomposition of features mentioned above can be achieved

through the singular value decomposition (SVD). Therefore, the type vector s^k of wrong movement can be acquired through the stacking vector $c^k = As^k$, thus the coefficient matrix C^k of No. k person can be obtained.

Withal, the wrong movement outline [8] of the No. k person can be acquired:

$$y_t^k = \mathcal{A} \times_1 s^k \times_2 \psi(b_t) . \tag{4}$$

Where, $\mathcal{A}$ is the third-order tensor matrix of $D \times K \times N$ demention, which can be acquired by restacking coefficient matrix A. $\times_1$ and $\times_2$ are the tensor product under different modes.

Pattern vector s of wrong movements is the linear combination of standard orthonormal basis of wrong movements in space. However, the dimension of the pattern vector depends on number of people and training cycle. Therefore, the dimension of vector s may be very high. As for high-dimension pattern vector s_t, it is very difficult to carry out tracking because it is easy to be caught in local minimum. Therefore, the pattern vector s_t is expressed as a group of protruded linear combination of pattern which is obtained from training. In case the type of pattern is $s^q, q = 1, \cdots, Q$ and the protruded linear weight is $\lambda = \left[\lambda^1, \cdots, \lambda^Q \right]$, then the pattern vector s can be expressed as:

$$s = \sum_{q=1}^{Q} \lambda^q s^q, \ \sum_{q=1}^{Q} \lambda^q = 1 . \tag{5}$$

Where, Q represents the number of pattern types. In general, it can be known by clustering algorithm in data space of pattern. Then the model in Equation (1) can be expressed as follows:

$$z_t = T_{\alpha_t} \left(\mathcal{A} \left[\sum_{q=1}^{Q} \lambda^q s^q \right] \times \psi(f(\beta_t)) \right) . \tag{6}$$

Four parameters such as scaling parameters S_x and S_y as well as conversion parameters T_x and T_y are used to realize parameterization of conversion change α and then the tracking problem of wrong movements is changed into the problem to estimate parameters α_t, β_t and z_t in given and observed image z_t of physical exercise.

3. Process of feature detection

3.1. Range of extremum

We can use one image I of physical exercise as one mapping $I \colon \mathcal{D} \subset \mathcal{N}^2 \to \mathcal{V}$, where, as for colorful image of physical exercise, the value of $\mathcal{V}$ is $[0, 255]^3$ in general. The information channel of image I of physical exercise is one mapping $C \colon \mathcal{D} \to \mathcal{S}$, where, $\mathcal{S}$ is a completely ordered set and $f_c \colon \mathcal{V} \to \mathcal{S}$ is the projection of pixel value

for a completely ordered set. In case A represents an adjacent area $A \subset \mathcal{D} \times \mathcal{D}$, then in the Thesis, 4 situations of pixel will be taken into consideration and pixels with coordinates of $(x \pm 1, y)$ and $(x, y \pm 1)$ are adjacent matrix of pixel (x, y). The area $\mathcal{R}$ for image I of physical exercise is one continuous subset of $\mathcal{D}$:

$$\begin{aligned} &\forall p_i, p_j \in \mathcal{R}, \exists p_i, q_1, q_2, \ldots, q_n, \\ &p_j : p_i A q_1, q_1 A q_2, \cdots, q_n A p_j \, . \end{aligned} \tag{7}$$

The outer boundary $\partial \mathcal{R}$ is a group of adjacent pixel sets not belonging to $\mathcal{R}$:

$$\partial \mathcal{R} = \{p \in \mathcal{D} \backslash \mathcal{R} := \exists q \in \mathcal{R} : \ pAq\} \, . \tag{8}$$

Extremum area is one area with pixel of outer boundary higher than pixel of the area:

$$\forall p \in \mathcal{R}, q \notin \partial \mathcal{R} : \ C(q) > \theta > C(p) \, . \tag{9}$$

Where, θ represents threshold value of extremum area. In the Thesis, we will consider RGB and HIS color space and additional strength gradient channel (∇), where, "gradient" pixel distribution is carried out for maximum intensity difference through adjacent pixel.

$$C_{\nabla}(p) = \max_{q \in \mathcal{D}:pAq} \{|C_I(p) - C_I(q)|\} \, . \tag{10}$$

In images of real physical exercises, some features consist of smaller elements or single element which consists of several combined features. Images of physical exercises are preprocessed by decomposing Gaussian pyramid. In case one feature contains several elements and these elements are combined together and constitute one single area, then in pyramid at each level, only stroke width of features with a certain gap is amplified.

3.2. Incremental calculation of descriptors

The key for rapid classification of ERS is the rapid calculation of regional descriptor as feature of classifier. The extremum area r on threshold value θ can constitute combined area or more extremum areas on threshold value $\theta - 1$. One special descriptor and the inclusion relation between ERS can be used and it is also possible to calculate the value of descriptor by incremental calculation.

In case $R_{\theta-1}$ represents one set of extremum area on threshold value $\theta - 1$, then one extremum area $r \in R_{\theta}$ on threshold value θ can consist of regional pixel on threshold value $\theta - 1$ and pixel value on threshold value θ:

$$r = (\cup u \in R_{\theta-1}) \cup (\cup p \in \mathcal{D} : C(p) = \theta) \, . \tag{11}$$

Let us further assume descriptors $\varphi(u)$ of all extremum areas on threshold value $u \in R_{\theta-1}$. To calculate the descriptor $\varphi(r)$ of the area $r \in R_{\theta}$, it is necessary to combine descriptor of area $u \in R_{\theta-1}$ and pixel $\{p \in \mathcal{D} : C(p) = \theta\}$ to form the area

r:

$$\varphi(r) = (\oplus \varphi(u)) \oplus (\oplus \psi(p)) . \tag{12}$$

Where, the symbol $\oplus$ represents the operation of descriptor in combined area (pixel) and $\psi(p)$ represents the initialization function to calculate descriptor of given pixel p. In this case, the process of incremental calculation exists between this descriptors $\psi(p)$ and $\oplus$, as is shown in Fig. 2.

(a) Perimeter p (b) Euler number η
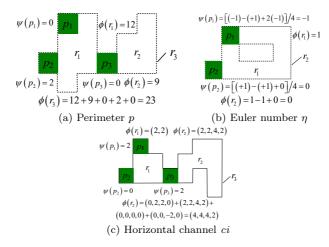
(c) Horizontal channel ci

Fig. 2. Incremental calculation of descriptors

It is obvious that descriptors in all extremum areas can be calculated by simply and successively increasing threshold value θ from 0 to 255. The calculation descriptor ψ of pixel is added to threshold value θ and then descriptor of key area φ is added to threshold value $\theta - 1$. Moreover, in case descriptor of single pixel $\psi(p)$ is calculated and combined operation $\oplus$ has fixed time complexity, then the calculation complexity of descriptors of all extremum areas for image of physical exercise of N pixels is $O(N)$ because $\varphi(p)$ for each pixel is only calculated for once. Therefore, the assessment on composite function can be completed in N times of calculation at most because the upper limit for number of extremum area is the pixel number of images of physical exercise. In the Thesis, we can use the following incremental calculation method to calculate descriptors:

(1) As for area parameter a namely the pixel number, the initialization function is one constant function $\psi(p) = 1$ and the combination operation $\oplus$ represents adding operation $(+)$.

(2) The bounding box $(x_{\min}, y_{\min}, x_{\max}, y_{\max})$ represents the top right corner and lower left corner of this area. The initialization function for pixel p of coordinate (x, y) is $(x, y, x + 1, y + 1)$ and the combination operation $\oplus$ is (min, min, max, max). The width ω and height h of the area can be respectively calculated as:

(3) Perimeter parameter p. The length of zone boundary is shown in Fig. 2a. When threshold value is added to pixel p, the initialization function $\psi(p)$ can deter-

mine change in length of perimeter:

$$\psi(p) = 4 - 2\left|q : qAp \wedge C(q) \leq C(p)\}\right| . \tag{13}$$

Where, the calculation complexity of $\psi(p)$ is $O(1)$ because each pixel has four neighboring pixels at most.

(4) Euler number η. Euler number is the topological feature of one binary image of physical exercise and its calculation equation is:

$$\eta = \frac{1}{4}(C_1 - C_2 + 2C_3)). \tag{14}$$

Where, C_1, C_2 and C_3 respectively represent the binary pixel number in image of physical exercise. The initialization function $\psi(p)$ can be determined by setting pixel p of threshold value $C(p)$ from 0 to 1:

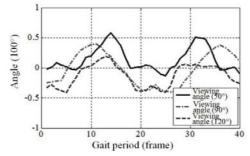$$\psi(p) = \frac{1}{4}(\Delta C_1 - \Delta C_2 + 2\Delta C_3)). \tag{15}$$
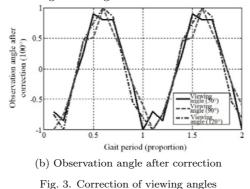
4. Experimental analysis

Setting of simulation hardware: operating system win7, CPU i5-4000 2.4GHz and RAM 8G ddr3 1600 and the simulation software is Matlab2012a.

4.1. Correction of observation angle

To verify effectiveness of proposed algorithm in identification of wrong movements under several monitoring cameras, it is required to verify effectiveness of algorithm for local extremum area in the algorithm first of all. Simulation of angle correction is carried out based on database of Casia-B wrong movements. The observation angles in conditions of different viewing angles (different video shots) are given in Fig. 3 without carrying angle correction and adopting angle correction algorithms proposed in the Thesis and the results are shown in Fig. 3 (a) (b).

Algorithm simulation results of local extremum areas are given in (a) and (b) in Fig. 3. It can be seen from Fig. (a) that in different viewing angles, observed swinging angle of lower limb is different. In case the swinging angle is not corrected, it is impossible to positioning and calculating angles of wrong movements in switchover of several shot cuts and thus the precision in identification algorithm of wrong movement is substantially reduced. Observation angles after algorithm of local extremum areas are given in (b). It can be seen from the figure that the observation angles in above-mentioned three viewing angles are basically the same. After treatment by this algorithm, features of wrong movement in different video shots can be unified, which can effectively improve precision in identification of wrong movements in switchover of several shot cuts.

(a) Observation angle of wrong movements in different viewing angles

(b) Observation angle after correction

Fig. 3. Correction of viewing angles

4.2. Identification and comparison of wrong movements

To verify effectiveness of proposed method, classification and identification of objects for wrong movements have been carried out based on database of Southampton indoor wrong movements. The database includes 160 video sequence data sources of wrong movements and there are 20 people with 8 video sequences for each person. Visual graph to select methods of image sequence of physical exercise to obtain the most discriminative character subset of wrong movements and correlation matrix extracted from features is shown in Fig. 4. The dark box reflects relatively higher degree of separation and it has relatively higher discriminability. At the same time, the box with bright diagonal reflects that the distance between the same type of wrong movements is zero and reflects that the degree of separation is the lowest.

To further assess identification ability of selected character subset, the classification accuracy (CCR) of character subset is calculated based on k-nearest neighbor classifier (KNN) and cross validation gauge. The simplicity and rapidity of KNN algorithm in classification application makes it convenient for use of existing methods. Therefore, in the Thesis, in classification of wrong movements, KNN algorithm is the foundation and cross validation rule is the auxiliary means to increase precision in classification and assessment. The comparison algorithm is the method in Literature [15] and [16] and classification and identification results are shown in Table 1.
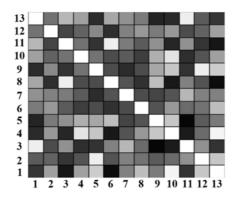
Fig. 4. Correlation matrix in identification of southampton wrong movements

Table 1. Classification and identification comparison results of wrong movements

Algorithm	Database of wrong movements	Accuracy in classification
Literature [15]	Southampton	83.7%
Literature [16]	Southampton	88.2%
Algorithm proposed in the Thesis (without correction of viewing angle)	Southampton	51.3%
Algorithm proposed in the Thesis (correction of viewing angle)	Southampton	96.3%

When $k = 5$ and the identification rate of 160 videos of sequential wrong movements in database of Southampton indoor wrong movements is given in Table 1, the identification rates of comparison algorithm are respectively 83.7% and 88.2%. However, as for algorithm proposed in the Thesis, when viewing angle is not corrected, the identification rate of wrong movement is 51.3% and the identification rate after correction of viewing angle is above 96%, which shows this algorithm is far better than comparison algorithm and verifies effectiveness of proposed algorithm. To further assess dataset in different scales and identification performance of proposed method, random excerpting is carried out for database of Southampton indoor wrong movements and $N = [30, 50, 70, 90, 110]$ subsets are used as simulation objects. Simulation comparison algorithm is subject to methods in Literature [15] and [16] and the comparison results of identification rate are shown in Fig. 5. The result in Fig. 5 is the average identification rate to operate above-mentioned algorithms for 10 times.

Accuracy in classification and identification of wrong movements by three algorithms in different scales of video datasets is given in Fig. 5. It can be seen from the figure that when there is less video flow (for example, the number of video flow is 30), classification and identification rate of three algorithms is low. When the number of algorithms proposed in the Thesis is 30, the identification rate is 75.8% and the classification and identification rate of the algorithm in Literature [15] is 52.3% while the classification and identification rate of the algorithm in Literature
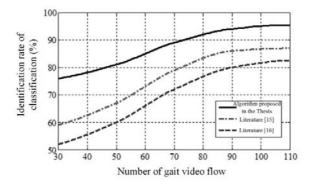
Fig. 5. Classification and identification rate of datasets in different scales

[16] is 58.7%. With increase in video flow in database, the classification and identification rates of several algorithms increase and become saturated in the end. When the number is 110 and close to original database, the identification rate of algorithm proposed in the Thesis is about 96% and the classification and identification rate of the algorithm in Literature [15] is 83% while the classification and identification rate of the algorithm in Literature [16] is 88% and all of them are close to classification and identification rate of video in original database. The reason of above-mentioned phenomenon is that randomly selected test database is not so good in conforming to rules for data and thus the identification rate is reduced. However, on the whole, classification and identification rate of algorithm proposed in the Thesis is better than that of comparison algorithm.

5. Conclusion

One extraction method based on the key features of wrong movements in physical exercise in two stages of local extremum area is proposed in the Thesis to built 3D identification model for wrong movements of physical exercises. Features are used for probability calculation of features in each extremum area and then the feature with maximum local probability is used as output in the first stage and input in the second stage so as to realize effective identification of feature model. In the next stage, we will design application research on the algorithm and then analyze and assess application effects of the algorithm.

References

[1] PAHLAVAN K, KRISHNAMURTHY P AND GENG Y: (2015) *Localization challenges for the emergence of the smart workd.* IEEE Access, 3(1), pp. 3058-3067

[2] LV, Z., TEK, A., DA SILVA, F., EMPEREUR-MOT, C., CHAVENT, M., & BAADEN, M.: (2013). *Game on, science-how video game technology may help biologists tackle visualization challenges.* PloS one, 8(3), e57990.

[3] LV, Z., HALAWANI, A., FENG, S., LI, H., & RÉHMAN, S. U.: (2014). *Multimodal*

hand and foot gesture interaction for handheld devices. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 11(1s), 10.

[4] WEISEN PAN, SHIZHAN CHEN, ZHIYONG FENG (2013) *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences, 7(2): 675-681.

[5] YINGYUE ZHANG, JENNIFER W. CHAN, ALYSHA MORETTI, AND KATHRYN E. UHRICH: (2015) *Designing Polymers with Sugar-based Advantages for Bioactive Delivery Applications*, Journal of Controlled Release, 219, 355-368.

[6] YINGYUE ZHANG, QI LI, WILLIAM J. WELSH, PRABHAS V. MOGHE, AND KATHRYN E. UHRICH: (2016) *Micellar and Structural Stability of Nanoscale Amphiphilic Polymers: Implications for Anti-atherosclerotic Bioactivity*, Biomaterials, 84, 230-240.

[7] JENNIFER W. CHAN, YINGYUE ZHANG, AND KATHRYN E. UHRICH: (2015) *Amphiphilic Macromolecule Self-Assembled Monolayers Suppress Smooth Muscle Cell Proliferation*, Bioconjugate Chemistry, 26(7), 1359-1369.

[8] LI L, ZHANG X, WANG Y G, ET AL.: (2008) *Nonparametric Motion Feature for Key Frame Extraction in Sports Video*[C]// Pattern Recognition, 2008. CCPR '08. Chinese Conference on. IEEE, 2008:1-5.

[9] ZHANG Y, XU C, LU H: (2010) *Extracting Key Sub-trajectory Features for Supervised Tactic Detection in Sports Video*[C]// International Conference on Pattern Recognition. IEEE Computer Society, 2010:125-128.

[10] ZHANG Y, XU C, LU H: (2010) *Extracting Key Sub-trajectory Features for Supervised Tactic Detection in Sports Video*[J]. 2010:125-128.

[11] NITTA N, BABAGUCHI N, KITAHASHI T: (2000) *Extracting Actors, Actions and Events from Sports Video - A Fundamental Approach to Story Tracking*[C]// International Conference on Pattern Recognition, 2000. Proceedings. IEEE, 2000:718-721.