HUMAN DETECTION AND MOTION RECOVERY BASED ON MONOCULAR VISION

DETECÇÃO HUMANA E RECUPERAÇÃO DE MOVIMENTO COM BASE NA VISÃO MONOCULAR

DETECCIÓN HUMANA Y RECUPERACIÓN DE MOVIMIENTO BASADA EN VISIÓN MONOCULAR

Original Article
Artigo Original
Artículo Original

Dongbo Liu[1]
(Physical Education Professional)

1. Chang'an University, Shaanxi, China.

Correspondence:
Dongbo Liu
Chang'an University, Shaanxi, China. liudongbocau@yeah.net

ABSTRACT

Objective: Provides interactive games and human animation real motion data and technical options. Therefore, how to complete the position, attitude detection, and motion recovery under monocular vision has become an important research direction. Methods: This paper improves the part-based human detection algorithm and uses the AdaBoost multi-instance learning algorithm to train the part detector. Results: The results show that obtaining blood pressure waveform based on monocular vision pulse wave is feasible and has generalization. Conclusions: The results show the feasibility and accuracy of the gait motion detection, motion recovery and analysis system for human lower limbs based on monocular vision. ***Level of evidence II; Therapeutic studies - investigation of treatment results.***

Keywords: Detection; Motion recovery; Vision, monocular.

RESUMO

Objetivo: Fornece jogos interativos e dados de movimento real de animação humana e opções técnicas. Portanto, como completar a posição, detecção de atitude e recuperação de movimento sob visão monocular tornou-se uma importante direção de pesquisa. Métodos: este artigo aprimora o algoritmo de detecção humana baseado em partes e usa o algoritmo de aprendizado de múltiplas instâncias AdaBoost para treinar o detector de partes. Resultados: Os resultados mostram que o método de obtenção da forma de onda da pressão arterial com base na onda de pulso de visão monocular é viável e se pode generalizar. Conclusões: Os resultados mostram a viabilidade e precisão do sistema de detecção, recuperação e análise do movimento da marcha para membros inferiores humanos com base na visão monocular. ***Nível de evidência II; Estudos terapêuticos- investigação dos resultados do tratamento.***

Descritores: *Detecção; Recuperação de movimento; Visão monocular.*

RESUMEN

Objetivo: Proporciona juegos interactivos y animación humana, datos de movimiento real y opciones técnicas. Por lo tanto, cómo completar la posición, la detección de actitud y la recuperación de movimiento bajo visión monocular se ha convertido en una importante dirección de investigación. Métodos: este documento mejora el algoritmo de detección humana basado en piezas y utiliza el algoritmo de aprendizaje de instancias múltiples AdaBoost para entrenar el detector de piezas. Resultados: Los resultados muestran que el método de obtención de la forma de onda de la presión arterial basado en la onda de pulso de visión monocular es factible y se puede generalizar. Conclusiones: Los resultados muestran la viabilidad y precisión del sistema de detección, recuperación y análisis del movimiento de la marcha para miembros inferiores humanos basado en visión monocular. ***Nivel de evidencia II; Estudios terapéuticos- investigación de los resultados del tratamiento.***

Descriptores: *Detección; Recuperación de movimiento; Visión monocular.*

DOI: http://dx.doi.org/10.1590/1517-8692202127042021_0113 Article received on 04/28/2021 accepted on 05/10/2021

INTRODUCTION

The extraction and processing of related person information in the video is important to use value.[1] By analyzing human gestures, action, and expression analysis, the computer can understand people's intent and achieve true intelligence analysis and processing.[2] The single eye is relatively small compared to the double eye, and the amount of calculation is relatively small, which is convenient for operation.[3] Golden motion extraction Based on single eye video, including recovery or reproducing real human motion data on video summary body models, this will provide a broader realistic motion data and interactive games and human animation technologies.

METHODS

Human detection based on monocular vision

Human detection technology is to search all human targets from images or videos to be detected. To this end, this chapter uses cascaded

Adaboosting learning methods to train human detectors.[4] Fast feature selection algorithm and cascade classifier training.

In view of the shortcomings of AdaBoosting in training speed, we use this fast feature selection method to improve the training of weak classifiers.The process is shown in Figure 1a, where S is the number of samples and P is the maximum number of rounds of training.In the feature selection stage, all candidate feature information is obtained by querying the statistical table. Each additional feature can quickly find the classification error of the overall strong classifier.[5] The total time complexity of the algorithm is O (SMPlogS), as shown in Figure 1a. The time complexity of the general Adaboost algorithm is O (SMP + SMLogS).This shows that the time complexity required to select a feature and train a weak classifier is $\frac{1}{S}+\frac{1}{\log S}$ of the original Adaboost algorithm. In monocular video surveillance scenes or still images, most areas are background images, and human targets occupy only a few areas.In the cascade structure of Cascade classifier, as long as the detection area is judged to be non-human by a certain classifier, the detection process of the target is ended, and the output is non-human. After passing through all stages of detection, it is the human body, and its structure is shown in Figure 1b.

Each strong classifier is represented by Equation 1, and is a linear combination of selected features.

$$M(x)=\operatorname{sgn}\left[\sum_{p=1}^{P}\alpha_i m_i(x)-\beta\right] \quad (1)$$

Among them, α_i represents the weak classifier weight, β is the strong classifier threshold, and its initial definition is $\beta=\frac{1}{2}\sum_{p=1}^{P}\alpha_i$ .The strong classifier M_r contains different numbers of weak classifiers $m_i(x)$, which are composed of a feature v_j , threshold β, and directions M_j indicating inequality signs.The final output of the algorithm is a cascaded strong classifier $M=\{M_1,M_2,\cdots M_r\}$ as a human detector.

Fast target human detection

When the detection window slides on the detection image, only 2r number of pixel regions change.Correspondingly, the bins of 2sr histograms change, where s is a factor between 0 and 1, and r is the detection window size. The pseudo code of the histogram search method based on the "block" update is:

Form=1: scaledo

(A) Initializing $q_{i,j}, i=1,L,n$,and $j=1,L,m$

(B) For $i=1:n$ do (1) For $j=2:m$ do, $q_{i,j-1}\leftarrow q_{i,j}$

(2) endform, $q_{i,m}\leftarrow q_{new}$

(C) Endforn

Endforq.

Suppose a given test window M is used to calculate its CT value characteristic histogram. Let f_M be the corresponding histogram and $|M|$ be the number of all pixels in M.[6] At each pixel position (x, y), the value of the k-th bin of histogram h_M is represented as

$$f_k=\sum_{(x_i,y_i)\in M}1\{b(x_i,y_i)=k\} \quad (2)$$

Where $1\{g\}$ is an index function that maps a pixel (x_i,y_i) to the corresponding bin.When the detection window slides, only the leftmost column C_L and rightmost column C_R need to be re-counted, which is denoted as

$$f_k=f_k-\sum_{(x_i,y_i)\in C_L}1\{b(x_i,y_i)=k\}+\sum_{(x_i,y_i)\in C_R}1\{b(x_i,y_i)=k\} \quad (3)$$

The number of changing pixels is $2r(2r=|M|)$.
Among them, MissRate and FPPW are defined as

$$MissRate=\frac{FalseNegatives}{TruePositives+FalseNegatives} \quad (4)$$

$$FPPW=\frac{FalsePositives}{TrueNegitives+FalsePositives} \quad (5)$$

The background image is tested to obtain the DET curve shown in Figure 2. Figure 2 shows the comparison of the detection performance DET curve on the INRIA database using a single HOG feature and a HOG-CT hybrid feature. It can be seen that the detection performance of the Adaboost algorithm constructed in this study has been significantly improved.

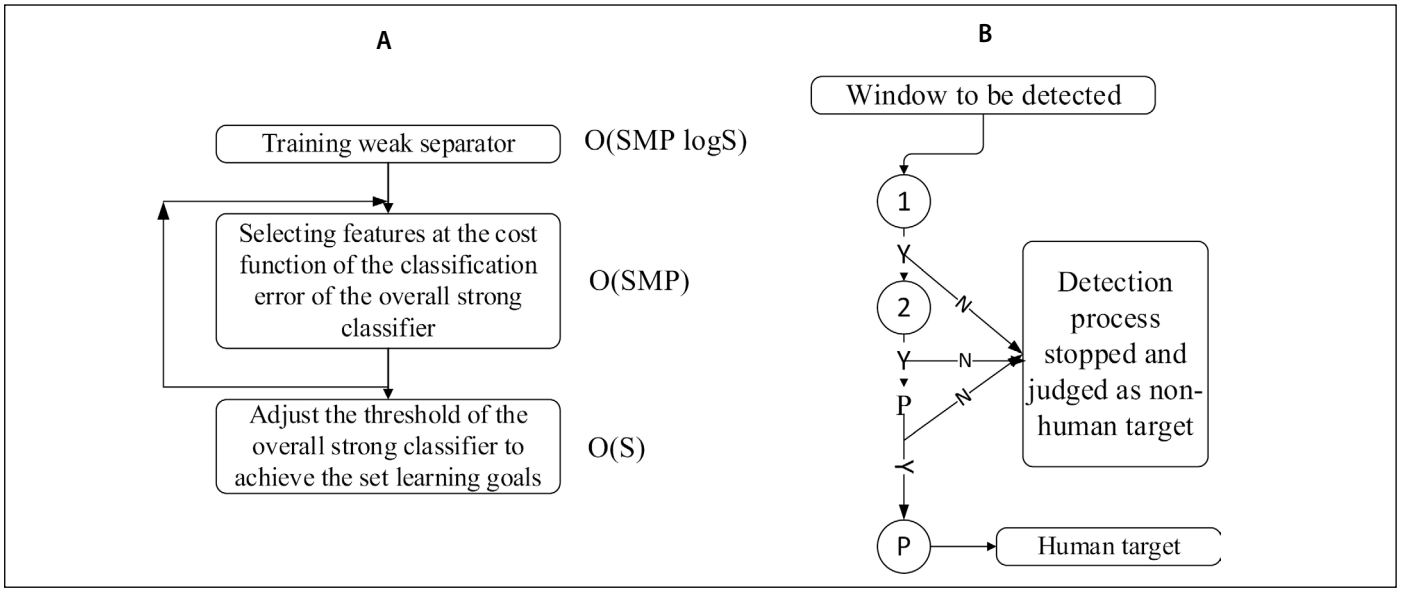

Figure 1. The process of fast feature selection to improve the AdaBoosting algorithm (A) and hierarchical cascade classifier (B).

RESULTS

Human pulse information and blood pressure waveform measurement methods under monocular vision

The characteristic parameters of the pulse that have a greater correlation with the blood pressure value are mainly including: Main wave amplitude h1, weighted wave amplitude h4, weighted wave relative height h4 / h1, waveform coefficient K value,h1 (1 + t3 / t4) reflecting cardiac output per stroke,ascending branch average slope h1 / t1, systolic relative area s1 / (s1 + s2). Figure 3 shows the pulse waveform extracted using the algorithm of this study and one of its pulse cycles.

Based on the pulse wave at the center of the pulse (Figure 3). The pulse parameters and calculated SBP and DBP are shown in Table 1.

The minimum value of the blood pressure waveform p (t) obtained initially is expressed as p (t) min and the maximum value is p (t) max. Using mathematical formula 6 of data standardization, p (t) can be mapped between [SBP DBP], and a reasonable blood pressure waveform P (t) can be obtained. (Figure 4)

$$P(t) = DBP + \frac{P(\mathrm{t}) - P(t)_{\min}}{P(\mathrm{t})_{\max} - P(t)_{\min}} \quad (6)$$

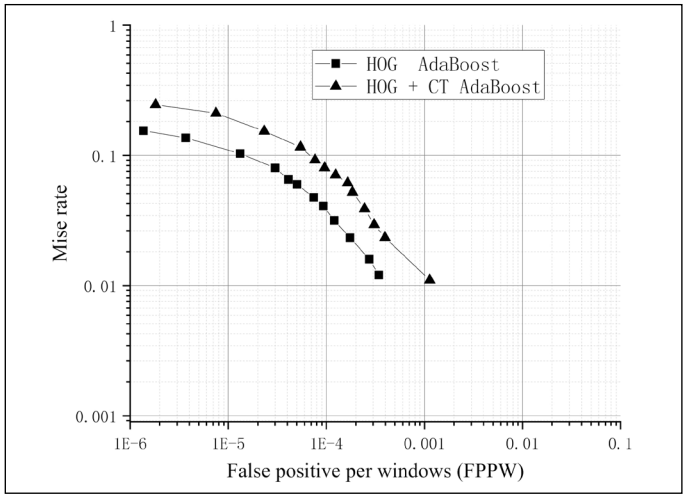

Figure 2. Comparison of detection performance using HOG and HOG-CT features in human detection.

Human Motion Recovery Based on Monocular Vision

This study proposes a generative 3D human body motion restoration method based on barrage vision. Firstly, analyze the contour of the human body to obtain the position information of the trunk and end nodes, and then optimize the 3D pose.[7] Experimental analysis was performed using a gait rehabilitation training machine. Let the human body posture be X, usually a high-dimensional space such as a joint point, and the video image observation is Z. If the two-dimensional human body contour is extracted, the recovery process is $Z \to X$ .It is very easy for human vision to judge the pose of a human in a monocular video image.While $Z \to X$ is a serious morbid problem compared to computers.But in the video sequence, the introduction of time-domain information transforms the reasoning problem into a dynamic process, which can be described by $\{Z_i | i = 1, L, t\}, \{X_i | i = 1, L, t-1\} \to X$.

Modeling of camera imaging model

It is considered that the coordinate value (x, y, z) of a point in three--dimensional space and the coordinate (u, v) of that point on the two--dimensional projection plane satisfy equation:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{1}{s} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (7)$$

Among them, the parameter s is a scale factor, which is obtained from $s = z / f$ .z is the z coordinate value of the point in three-dimensional space, and f is the focal length of the camera.[8] From equation (7), we can know that when z changes, the value of s will change linearly.The change amount ds of s relative to the z value dz is expressed as $T(dz) = ds$.[9-11] Figure 5 shows the imaging diagram of the three end-to-end bone segments under the perspective projection model.

Table 1. Pulse characteristics and calculated blood pressure values.

Pulse characteristic parameters	K	h1/h2(mm/)	H1 (1+t3/t4)(mm)	W(mm)
	0.315	2.319	0.416	3.761
Calculated blood pressure	SBP(mmHg)		DBP(mmHg)	
	119.54		74.65	

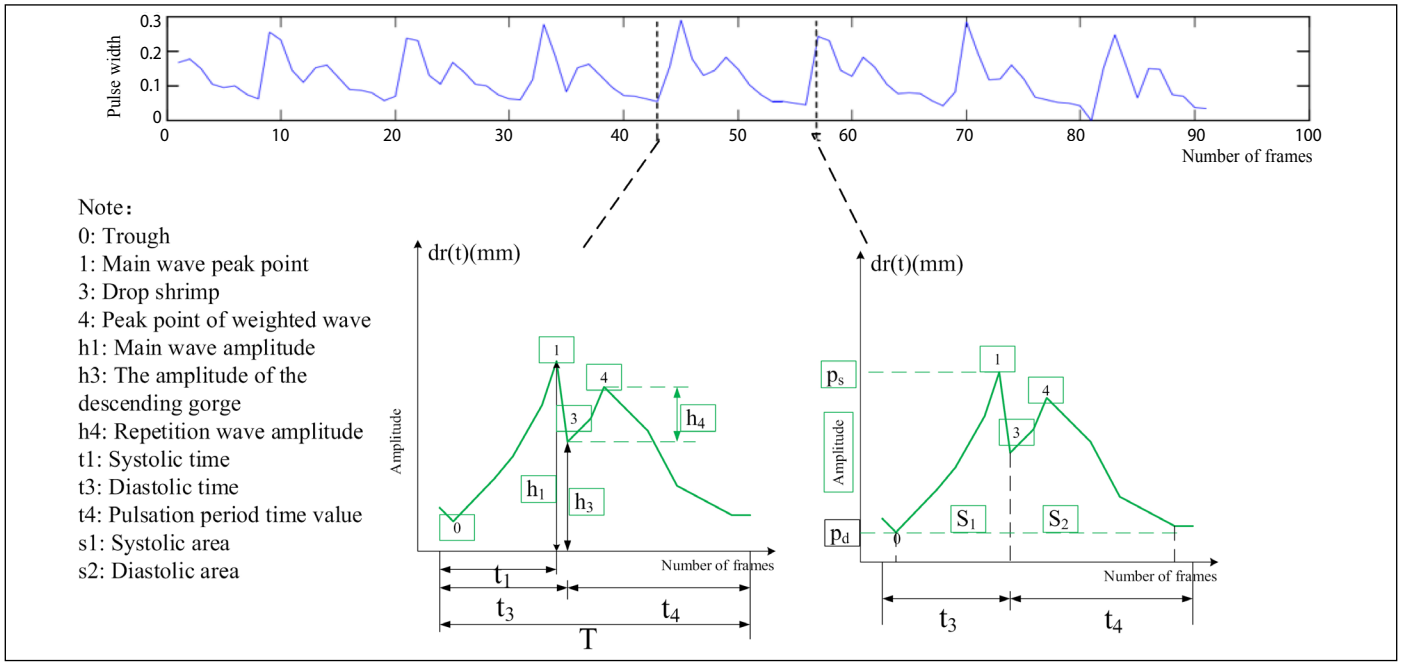

Figure 3. Pulse waveform extracted using the monocular vision Adaboosting algorithm and one of its pulse cycles.

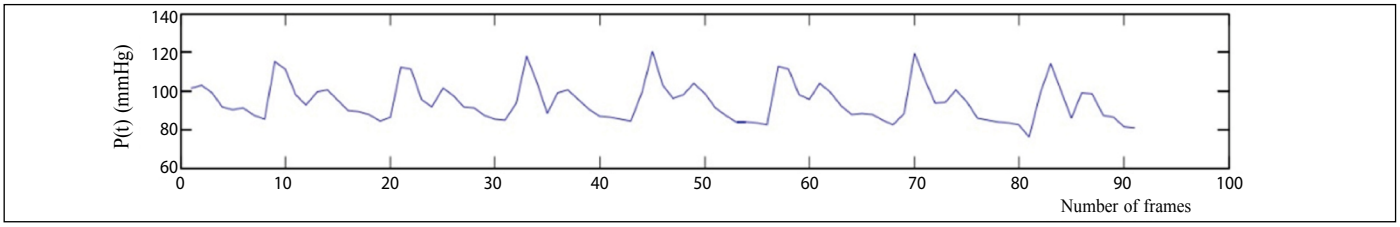

Figure 4. Blood pressure waveform after the monocular vision algorithm is revised.

In Figure 5, the skeletal segments ab and cd are parallel to the projection plane, and their imaging on the projection plane are a'b 'and c'd', respectively.The skeletal segment bc is not parallel to the projection plane, and its imaging on the projection plane is b 'c '.Assuming that the lengths of ab, bc, and cd are L_{ab}, L_{bc}, L_{cd} respectively, their values can be obtained from the human skeleton model.The projection lengths of a'b ', b'c' and c'd 'are $L_{a'b'}, L_{b'c'}, L_{c'd'}$.

Since bone ab is parallel to the projection plane, its projection on the z-axis intersects at point M.The variable factor q corresponding to the point M is calculated according to $q_{ab} = OM / g = L_{ab} / L_{a'b'}$.Similarly, we can get the variable factor q corresponding to the projection point N of the bone cd on the z axis, $q_{ab} = L_{ab} / L_{a'b'}$. Since the bones ab and cd are parallel to the projection plane, the distance between the point M and the point N satisfies $d_z = c_z - b_z$ .According to the knowledge of space geometry, L_{bc} satisfies :

$$\begin{aligned} L_{bc} &= \sqrt{(c_x - b_x)^2 + (c_y - b_y)^2 - (c_z - b_z)^2} \\ \Rightarrow L_{bc} &= \sqrt{(c_x - b_x)^2 + (c_y - b_y)^2 + dz^2} \\ \Rightarrow L_{bc} &= \sqrt{(s_{cd} c_x - s_{ab} c'_x)^2 + (s_{cd} c'_x - s_{ab} b'_x)^2 + dz^2} \end{aligned} \quad (8)$$

For $s_{cd}, s_{ab}, c'_x, b'_x, b'_y$ and L_{bc} are known, dz can be calculated by Equation (8).Relative to the absolute change of z (| dz |), the corresponding absolute change of s is $|ds| = abs(s_{cd} - s_{ab})$.

Modeling of human bone model during exercise recovery

We see the human body as a tree-like stick model, as shown in Figure 6.The skeletal model consists of 16 joint points and 5 body segments.Among them, M1 is the root node of the tree structure, which corresponds to the pelvic joint of the human body;The length of the line segment (human bone segment) in the model is obtained from anthropometrics.That is, 1) local coordinate system. It is fixed to each body segment, and the origin of the coordinate system is the attachment joint of the body segment; 2) Global coordinate system. The origin of the coordinate system is the M1 joint.

The composition of the objective function

This can be formalized as

$$\hat{h} = \arg\min \{E(h; M, R)\} \quad (9)$$

Where h is the three-dimensional pose vector. M is the camera model (transformation matrix from world coordinates to image coordinates),R is the analyzed contour. E () is an objective function that calculates the degree of matching between S and P (after C transformation).The objective function proposed in this chapter contains five parts, which correspond to the five skeleton segments on the human skeleton model:

$$\begin{aligned} &E(h; M, R) = E_{Torso}(h, M, R) + \Pi_{lp} E_{LUpper}(h; M, R) \\ &+\Pi_{rh} E_{RUpper}(h; M, R) + \Pi_{lf} E_{Llower}(h; M, R) + \Pi_{rf} E_{RLower}(h; M, R) \end{aligned} \quad (10)$$

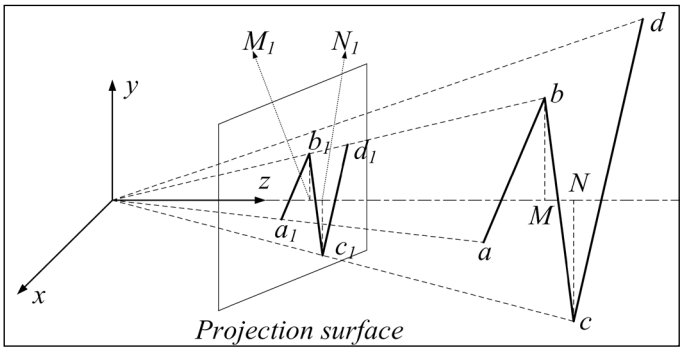

Figure 5. Schematic diagram of the lower bone section of the perspective projection model during motion recovery.

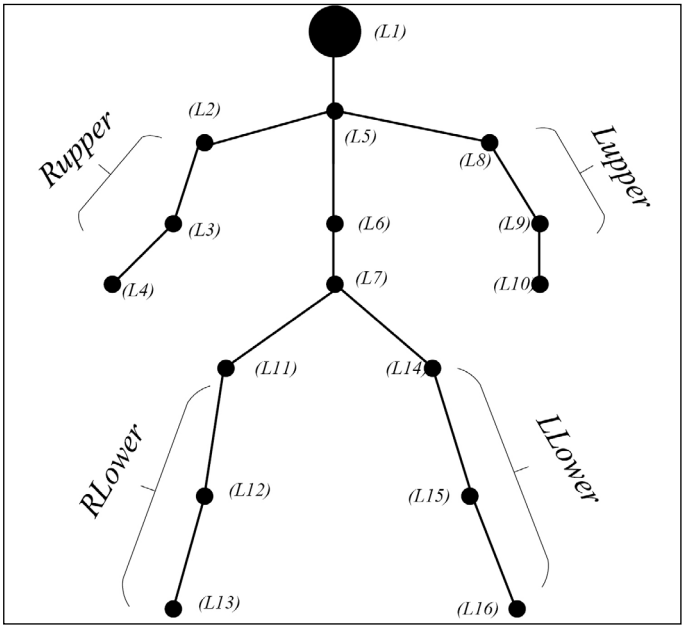

Figure 6. Human skeleton model during exercise recovery.

The definitions of the four ll values for both hands and feet are as follows: That is, when a limb (hand, foot) is blocked, the corresponding ll = 0, otherwise ll = 1.It can be seen that if a hand or foot is not positioned during the contour analysis step, its corresponding skeleton segment does not contribute to the objective function. Each term to the right of the equal sign in Equation 8 is further composed of three sub-terms: core region term $E^{core-area}$, end coverage term $E^{coverage}$, and timing smoothing term $E^{smoothness}$ .Take $E_{LUpper}(h; M, R)$ as an example:

$$E_{LUpper}(h; M, R) = a_1 E^{core-area}_{LUpper}(h; M, R) + a_2 E^{coverage}_{LUpper}(h; M, R) + a_3 E^{smoothness}_{LUpper}(h; M, R) \quad (11)$$

Note that the right 5 terms of Equation 9 are actually independent of each other. For example, when calculating the $E_{LUpper}(h; M, R)$ term, only the left upper limb LUpper skeleton segment is involved. Therefore, the optimization of the objective function can be divided into five independent sub-optimization processes. Figure 7a is a specific

optimization process. The objective function as a whole is optimized through continuous loop optimization. When the objective function value stops decreasing or reaches the number of loop iterations, the optimization ends. The sub-optimization process uses a simulated annealing algorithm. Simulated annealing can be seen as an improvement on the gradient descent method (Figure 7b). As the temperature T decreases, the probability $\exp(-\Delta E / T)$ decreases. It can be proven that when the initial temperature is high enough and the annealing speed is low enough, the output of the simulated annealing algorithm approaches the global optimal value asymptotically with probability 1.By setting the appropriate initial temperature and annealing coefficient, the extent to which the simulated annealing can overcome the local minimum can be easily controlled.

DISCUSSION

Human gait analysis and sports rehabilitation based on monocular vision

The scope of Figures 8A and 8B show the relationship between the hips and knee joints during the gait and treadmill in the gait cycle. The angular relationship curve of the gait machine is smooth and closed. This suggests that its lower extremity is in line with coordinated requirements. The angular relationship curve on the treadmill has different characteristics from the gait machine. This may be because people are not limited by the running machine, and the steps and frequencies will change to various factors, which cannot be periodically as motion on the gait machine.

From the experimental results of Figures 8A and 8B, it can be seen that the feasibility and accuracy of the lower extremity step motion detection and analysis system based on single-eye visual development. This study is a useful attempt to analyze and restore human movements based on single-eye visual tags, and have obtained some satisfactory preliminary research results.

CONCLUSION

This article is based on the feasibility and accuracy of the analysis and analysis system of human low limbs based on monocular vision. The next step is to use a dynamic programming algorithm to select a related feature from a relational feature library, and these features correspond to the weighting distance between the weight, the distance between the three-dimensional posture.

The author declare no potential conflict of interest related to this article

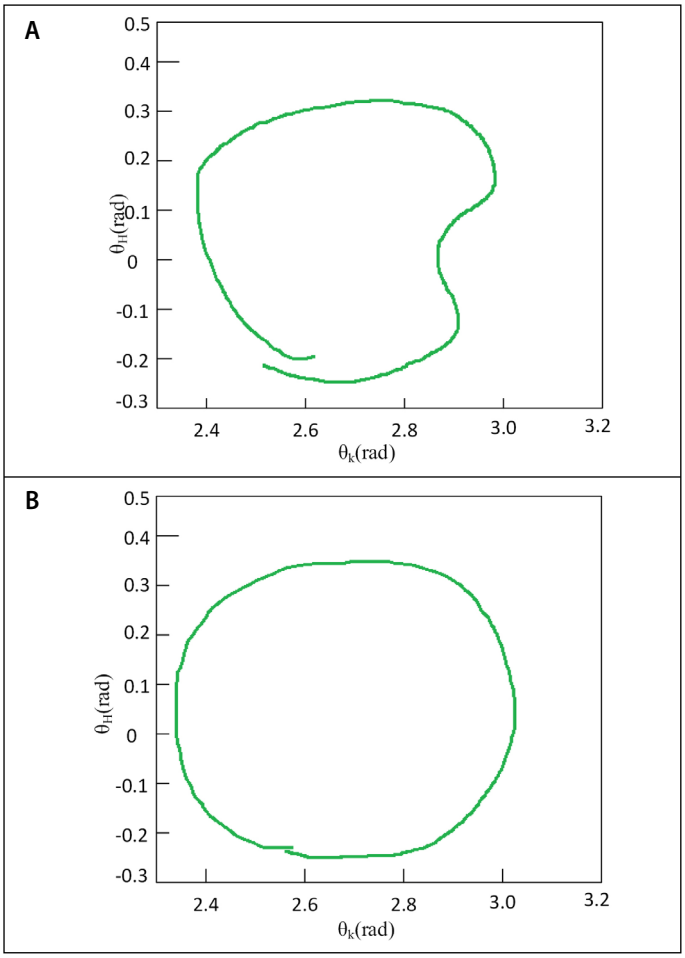

Figure 8. Joint-angle relationship on treadmill (a) and gait machine (b) in monocular vision.

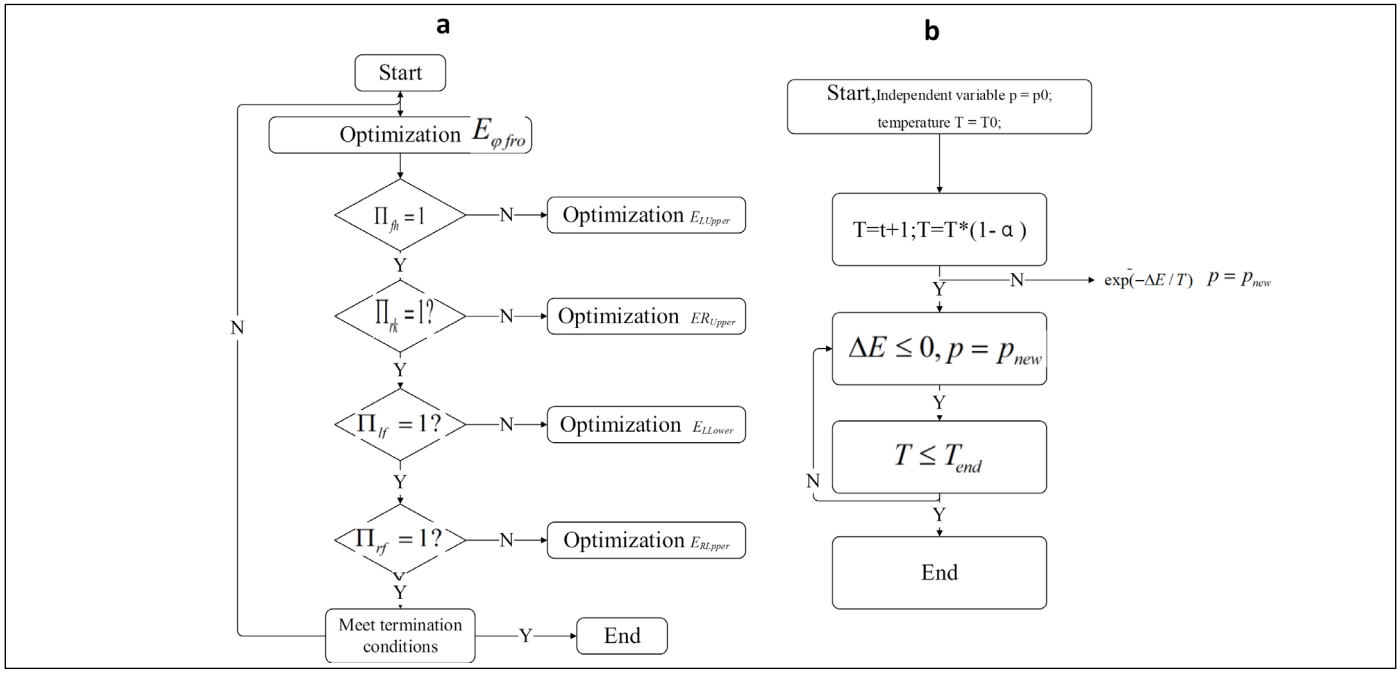

Figure 7. Iterative optimization process (a) and simulated annealing algorithm flow (b) during motion recovery.

AUTHORS' CONTRIBUTIONS: Dongbo Liu analyzed and explained that interactive games and human animation provide a wider range of realistic motion data and technology options. The completion of posture detection and motion recovery under monocular vision has become an important research direction.

REFERENCES

1. Chen HT, Wu YC, Hsu CC. Daytime preceding vehicle brake light detection using monocular vision. IEEE Sensors Journal. 2015;16(1);120-31.
2. Lee TJ, Yi DH, Cho DI. A Monocular Vision Sensor-Based Obstacle Detection Algorithm for Autonomous Robots. Sensors (Basel). 2016;16(3):311.
3. Lin S, Garratt MA, Lambert AJ. Monocular vision-based real-time target recognition and tracking for autonomously landing an UAV in a cluttered shipboard environment. Autonomous Robots. 2017;41(4):881-901.
4. Yuxi F, Guotian H, Qizhou WA. New Motion Obstacle Detection Based Monocular-Vision Algorithm. In 2016 International Conference on Computational Intelligence and Applications (ICCIA). 2016;31-5.
5. Jia B, Liu R, Zhu M. Real-time obstacle detection with motion features using monocular vision. The Visual Computer. 2016;31(3);281-93.
6. Su S, Zhou Y, Wang Z, Chen H. Monocular Vision- and IMU-Based System for Prosthesis Pose Estimation During Total Hip Replacement Surgery. IEEE Trans Biomed Circuits Syst. 2017;11(3):661-670.
7. Huang XY, Gao F, Xu GY, Ding NG, Xing LL. Depth information extraction of on-board monocular vision based on a single vertical target image. Journal of Beijing University of Aeronautics and Astronautics. 2016;41(4);649-55.
8. Chen Z, Zhang Z, Dai F, Bu Y, Wang H. Monocular vision-based underwater object detection. Sensors. 2017;17(8);1784-6.
9. Xu LY, Cao ZQ, Zhao P, Zhou C. A new monocular vision measurement method to estimate 3D positions of objects on floor. International Journal of Automation and Computing. 2017;14(2);159-68.
10. Zhang G, Liu J, Li H. Joint Human Detection and Head Pose Estimation via Multi-Stream Networks for RGB-D Videos[J]. IEEE Signal Processing Letters. 2017;3(13):19-32.
11. Yu XG, Li YQ, Zhu WB. Wearable strain sensor based on carbonized nano-sponge/silicone composite for human motion detection. Nanoscale. 2017;9(20):6680-92.