International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.6 (2016), pp.329-338
http://dx.doi.org/10.14257/ijmue.2016.11.6.29

Reconstruction of Human Behavior and Movement Feature Recognition Based on the Video Image

Yongbo Wang[1,*] and Leiqi Zheng[2]

[1]*Center of Modern Education Technology, Xinxiang Medical University, Xinxiang, Henan 453003, China*
[2]*Center of Modern Education Technology, Xinxiang Medical University, Xinxiang, Henan 453003, China*
E-mail: yongbowang_xmu@126.com

Abstract

Based on the analysis of human motion with the development of computer vision technology gradually become the research focus, while intelligent monitoring and other related real needs also contributed to its rapid development, attracting more and more researchers to join them. It includes both limbs motion tracking, as well as tracking the movement of the whole body, but does not include the subtle body movement, such as facial expressions and gestures of the track. The monocular video motion tracking method in this article are used in a variety of models, including the human body geometry, said camera imaging model and motion model used. Geometric model of the human body by the appearance of the model human skeleton model and composition of each joint skeleton model the relative proportions of the length of a value obtained from anthropometric measurements, and the establishment of a local coordinate system at each joint. Camera imaging using fluoroscopy model and a mathematical model based on this request to the relative three-dimensional coordinates of each joint between the sports model based on kinematics knowledge polynomial representation.

Keywords*: Video segmentation; particle filter; human motion tracking; human movement tracking; dimensional behavior recognition*

1. Introduction

Video analysis of human motion as a hotspot computer vision, has been widely used in intelligent control, a new generation of human-computer interaction, sports analysis and virtual reality research [1-3]. Due to the wide monocular video data sources, all kinds of movies, sports movement and dance are in the form of monocular video memory, and thus based on motion tracking and analysis of monocular video has become an important direction of current research in monocular video, three-dimensional human motion projected onto the target from a two-dimensional depth information is lost cause thus only monocular video resume three-dimensional motion gesture is a very difficult thing, research work challenging, attracting more and more attention from researchers [4-5].

Human behavior refers to the way human action and the reaction to the environment or object, the body through a complex limb movement, describe or express complex human behavior [6]. It can be said that most human behavior is manifested through human limb movement. Therefore, the body's movement to analyze human behavior has become an effective way to understand human behavior. Video analysis of human motion analysis and processing for the main image sequence contains human movement, involving knowledge and technology of image processing and analysis, computer vision, pattern recognition, artificial intelligence, computer graphics, mathematics, kinesiology and other disciplines the video contains its process of human motion segmentation and extraction

ISSN: 1975-0080 IJMUE

[7]. Human dimensional motion model constructed joints dynamic tracking, motion parameters and parameter calculation equation established and video animation reproduction and the like [8-10].

With the substantial increase in computing power, the display chip frequency increase, the widespread use of high-capacity storage devices, so whether it had a significant increase in processing and storage capacity of a single frame image and a sequence of images; a variety of video processing equipment emerging so you can easily record and store large amounts of human continuous motion, video tracking these developments and human behavior recognition algorithm is designed to provide adequate preparation for the hardware.

2. Research Status and Related Theory

2.1. Video-Based Human Motion Tracking Method Classification

A. Classification according to the tracking accuracy

Depending on their human motion tracking applications, human motion to obtain the required accuracy is also different according to the details you want to get it can be divided into several levels: primary precision, intermediate precision and high precision accuracy level of primary human motion tracking of the human body as a major a whole track (not including human limbs), and do not use any high-level information in the tracking process [11].

B. The number of cameras used in classification

Depth information is difficult to recover in human motion tracking based on monocular video, the two-dimensional plane and therefore more suitable for tracking applications. If you need to carry out three-dimensional attitude tracking, you need to meet certain constraints or some prior knowledge. In addition, in the moving target tracking process, often occur moving target is obscured or lost temporarily disappear, *etc.* thereby moving target, this problem is a difficulty tracking moving objects, and based on monocular video of human motion tracking is difficult to solve this problem .

C. Methods are classified according to search

From the perspective of the state space law point of view, video-based human motion tracking problem can be seen in the attitude parameters of the search space, that is a large parameter space (complete human models often have more than 20 degrees of freedom) within found optimal solution process. Different search methods can be divided according to the state of uncertainty and search method random search methods.

2.2. Human Geometric Model

Human geometric model is the shape, size, and represent the topological relations between their respective parts of the body. Typically include human skeleton model and human appearance model. Human skeleton model is mainly indicates the connection relationship between the human body limbs, usually in hours using inverse kinematics solver body rotation. Human appearance model is mainly indicates external texture of the human body limbs, clothing information, usually in the human body model is used to calculate the similarity matching. Generally speaking, the more complex the more accurate the model, the more accurate the results of its track, but needs more parameters and more time to search for the solution space.

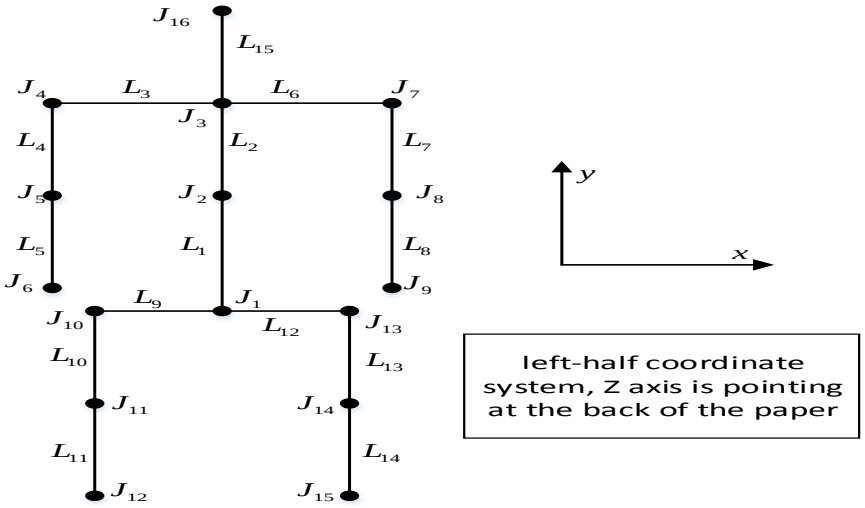

Figure 1. Human Skeleton Model

We will stick body as a tree-like model, as shown in Figure 1, the model consists of 16 bones and 15 joints of the body segments. Wherein, J: is the root of the tree structure, corresponding to the human pelvis joints, as the body's overall position in the body which coordinates the movement of people as a whole to determine the sequence of the three-dimensional trajectory. Model segments (human bone segment) length of the body measured in accordance with, as a relative proportion of the length, in the application should be set based on actual measured values.

3. Video-Based Tracking Mannequin

3.1. Model-Based Human Motion Tracking

Based human motion tracking model is such a process: to first establish the body geometry, in the track of the current frame, according to the last frame of the trace results to predict the attitude of the current frame, the posture of the body in the prediction model is projected onto the image plane, calculate match error, and then optimize posture prediction based on the error, the target function is minimized, resulting in body posture in the current frame, this method is also known as top-down approach.

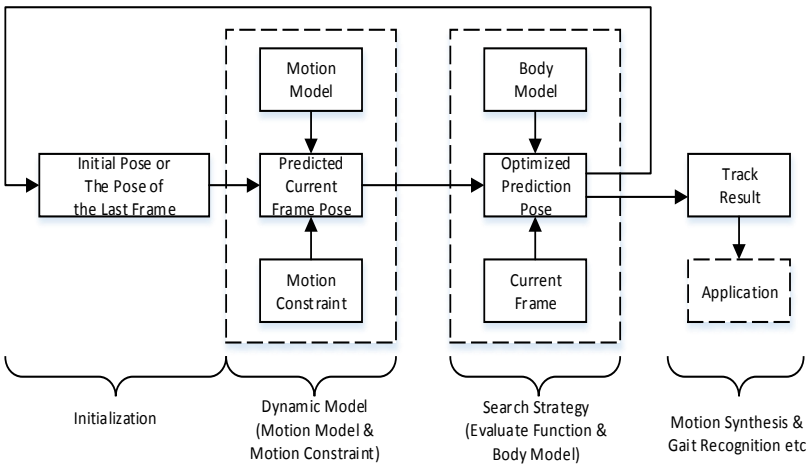

Figure 2. Block Diagram of Human Movement Based on Model

Figure 2 block diagram generally model human motion is based on the model of human motion tracking method due to the use of the actual geometry of the human body as a guide in the tracking process, making tracking more realistic. This method has two obvious the advantages: the use of the model in the tracking process, calculate the similarity can be difficult to block part of the track due to be taken into account due to occlusion, easier to handle when the model matching; the mannequins generally use a skeleton model, according to human movement knowledge of biomechanics or kinematics to eliminate some possible tracking results, so the solution to reduce the search space and improve the accuracy of tracking results.

3.2 Video Object Segmentation

During the period of the video object segmentation, in order to compensate for changes in lighting or background change, the need for background model updated in real time, using the following equation to update the general background of the mean and variance.

$$u_{k+1} = au_k + (1-a)z_{k+1} \tag{1}$$

$$\sigma_{k+1}^2 = a(\sigma_k^2 + (u_{k+1} - u_k)^2) + (1-a)(z_{k+1} - u_{k+1})^2 \tag{2}$$

Where Z_{k+1} is divided into the background of the current frame pixel values, a coefficient for the update, if the background of the rapid changes in the background of the need to update the show, a should be set to a smaller value, whereas a should be set to a large value.

Model commonly used parameters include: Parameter model (corresponding to the camera panning and zooming), parametric models (corresponding to the camera pan, zoom and rotate), parameter affine model, quadratic model parameters and parameter affine model perspective model because the parameters can be expressed. the vast majority of sports type, but also has better balance the computational complexity and estimation accuracy, thus often adopted, affine model parameters as follows:

$$\begin{cases} x = ax' + by' + e \\ y = cx' + dy' + f \end{cases} \tag{3}$$

First select a frame from the N data blocks using a block matching algorithm block the N position corresponding to the block in the next frame, the center coordinates of this group corresponds to block the formation of N points to define the deviation function is as follows:

$$f(a,b,c,d,e,f) = \sum_{i=1}^{n} (x_i - ax_i' - by_i' - \mathrm{e})^2 \tag{4}$$

The establishment of three differential background determines the following criteria:

$$F_k = \begin{cases} 0, |I_k - I_{k-1}| \le T \\ 1, otherwise \end{cases} \tag{5}$$

$$F_{k+1} = \begin{cases} 0, |I_{k+1} - I_k| \le T \\ 1, otherwise \end{cases} \tag{6}$$

$$x = \begin{cases} foreground, if F_k \cap F_{k+1} = 1 \\ background, if F_k \cap F_{k+1} = 0 \end{cases} \tag{7}$$

3.3. Mean Shift Method Applied Kernel

Mean Shift method is the cornerstone of the kernel function, the concept of the use of nuclear function, Mean Shift method can directly calculate the gradient feature points in the feature space, iterative search feature space available to local extreme points.

Given n static points of d-dimension space R^d , $x_i, i = 1, 2, ..., \mathrm{n}$. Multivariate kernel function $K(x)$ and the symmetric positive definite d-dimension band-width matrix H, the probability density estimation of point X can be written as:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_n(x - x_i)^2 \tag{8}$$

The kernel function K d variables (x) to satisfy the following conditions:

$$\int_{R^d} K(\mathrm{x})\,\mathrm{dx} = 1 \tag{9}$$

$$\lim_{|x|\to\infty} \|x\|^d\, k(x) = 0 \tag{10}$$

$$\int_{R^d} xK(\mathrm{x})\,\mathrm{dx} = 0 \tag{11}$$

$$\int_{R^d} xx^T K(\mathrm{x})\,\mathrm{dx} = c_k I \tag{12}$$

Function $G(x) = e^{-|x|^2}$ of the image was shown in Figure 3.

Figure 3. Unit Gaussian Kernel

4. Experiment and Analysis

4.1. Particle Filter Algorithm

Particle filter is a Bayesian estimation method based on the Monte Carlo, the basic idea is to use the corresponding weights in a random sample to describe the probability distribution of these samples are called particles, updated on the basis of a random sample

of observations on the weight and position, to approximate the actual probability distribution. this filtering algorithm recursively, easy to implement on a computer, and that the algorithm can be applied to the observed messages appear when abnormal mutation.

Under normal circumstances a dynamic system models were characterized as follows:

$$x_{k+1} = f_k(x_k, w_k) \tag{13}$$

$$z_k = h_k(\mathrm{x}_k, \mathrm{v}_k) \tag{14}$$

Figure 4 iterative method can be used for the first time the system state value k Xk filtering process.

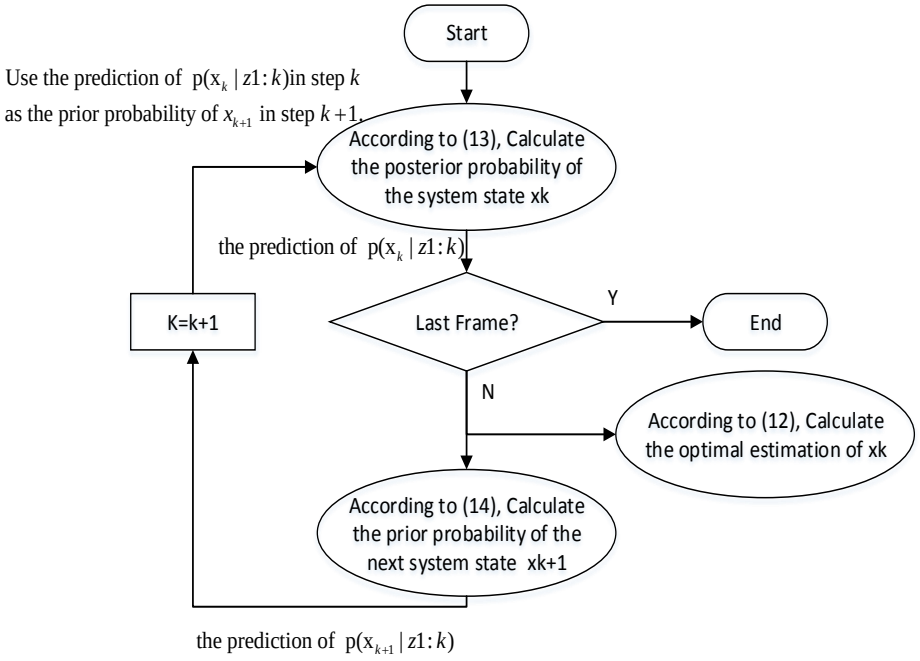

Figure 4. Particle Filter Schematic

Bayesian theory known by the formula (15) is to Xk estimated value and the actual value corresponding to the minimum error value, that is the best estimate of the value of Xk.

$$\overline{x}_k = \int x_k d(p(\mathrm{x}_k \mid z_{i:k})) \tag{15}$$

Tracking system successfully tracked the movement of the body, each frame can be positioned in the human body, while access to the human body in every gesture frame image, which is the data base of human behavior recognition. Human behavior recognition system not only to be used for basic everyday behavior recognition of human movement, such as walking, running, jumping, waving and so on, more importantly, can be used for security, to identify risk behavior such as theft and the like. To accomplish these functions, behavior recognition system should have adequate representation capability, that has a rich enough vocabulary to express human behavior. To make identification system to identify both more complex behavior, with strong communication skills at the same time can quickly identify in real time, using a combination of strategies for basic daily activities represent more complex behavior. As people can walk while waving people explore hands may bend it is also represented in the theft.

4.2. Test Data Processing and Results Analysis

Experiment was detected by edge and calculated by integral image and detection of human limbs two stages. The first stage: in a given video image of each frame, using Haar edge detector detects an image area of a limb or limbs class. Each frame size was 240 × 320 pixels. Figure 5 is a three extracted images in the video.

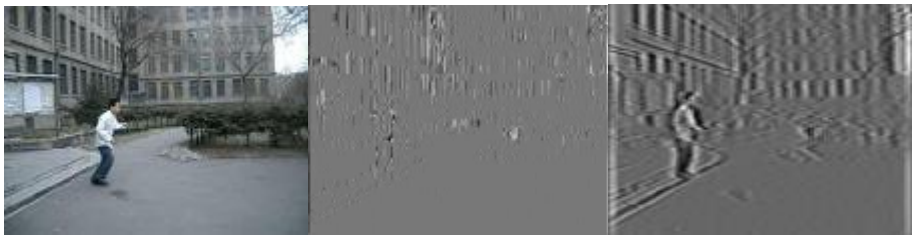

Figure 5. The Use of 3 × 21 Detector Detects an Edge

Taking into account the human body is approximately rectangular, and consider using the integral image detection image rectangle rectangular or class of human limbs, hoping to detect the location of its ideal body shape is rectangular or almost rectangular shape. When the pixels of 3 × 21 classes edge detector for detecting a human body, has a good detection results, shown in Figure 5.

Image feature class Hal edge detector was 3 × 21 pixels more in line with needs. Pixel gray value approximates the torso portion, and with the background there is a clear difference. Thus it can be used to represent the class wherein Hal human body, and calculates the gray value of the human body and in the region of the integral image, and determines the position of the human body based on the calculation result.

Table 1 shows the results of this algorithm applied to video segmentation dynamic scenarios, test video for athletes complete high-dive, whole diving athlete during fast movement changes, split with a certain degree of difficulty. As can be seen from the figure, although relatively complex and athlete background pixels and background pixels change interval change intervals have some overlap, divided by the prospect of this algorithm is quite out of the movement intact.

Table 1. Comparison of Experimental Results in Five Different Backgrounds

Test No	Used algorithm	Total pixels	Background pixels taken as foreground pixels	Ratio of background pixels taken as foreground pixels	Foreground pixels taken as background pixels	Ratio of foreground pixels taken as background pixels
No .1	3 σ	206450000	2419850	1.172%	37798040	18.31%
	proposed	206450000	2409850	1.167%	22515610	10.91%
No .2	3 σ	206450000	4054880	1.964%	38800730	18.79%
	proposed	206450000	4301980	2.084%	21133660	10.24%
No .3	3 σ	206450000	9063330	4.390%	52147360	25.26%
	proposed	206450000	9102730	4.409%	17014360	8.241%
No .4	3 σ	206450000	2828020	1.367%	38662640	18.73%
	proposed	206450000	2964880	1.436%	19837130	9.609%
No .5	3 σ	206450000	2586140	1.253%	44883440	21.74%
	proposed	206450000	2173860	1.053%	15326340	7.424%

Experimental results show that the situation in the larger variance in its entirety herein segmentation algorithm is superior to the rule 3σ segmentation results, but specific to each frame, may introduce division error, which was mainly due to the Gaussian model parameters using approximation the estimated results; experiments also show that with the

increase in the number of frames of segmentation, segmentation better.

4.3. Limb Detection Accuracy

When using a single class Haar feature location torso in the image, inevitably it will be the background or objects in an image similar to the limbs of false detection of human limbs. The detection accuracy is defined as follows: in each frame, as can be correctly detected human body, then that is a correct detection; if not correctly detect the position of the limb, then that is a mistake or missed detection accuracy rate is correct detection frames and the ratio of the total number of frames. Detection rate calculation using the following approach: the 100 detected images into 10 sets, 10 per test, used to calculate the human body detection accuracy of the test results shown in Figure 6.

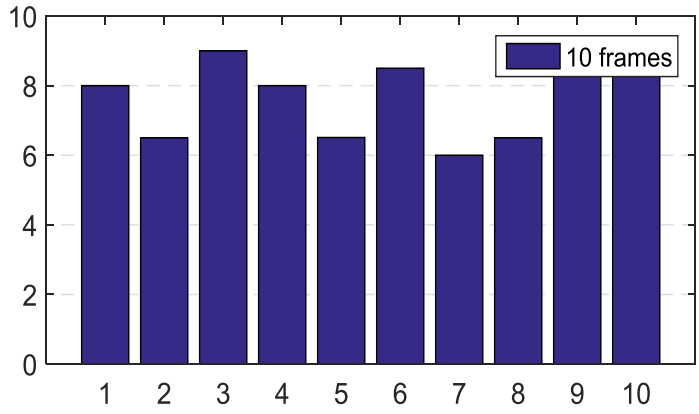

Figure 6. Using a Single Operator to Detect Torso Haar

After calculation, of 100 images to be detected accurately detect torso was 0.78; and the use of class Hal operator detects a combination of video can reduce the false detection rate, shown in Figure 7.

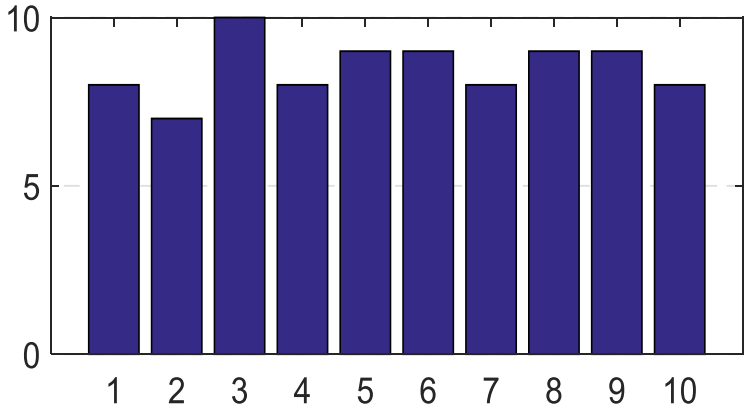

Figure 7. Hal Operator Detects a Combination of Video Can Reduce the False Detection Rate

Haar using three or a combination of three or more operators can also improve the detection accuracy of the limb, but the detection accuracy will not be greatly improved, and the use of class Hal operator of three or more combination detection body will consume more detection time, is not conducive to real-time requirements of intelligent vision system, the deeper reason is that the use of a combination of many types of Hal

operator increased the detection rate for subsequent clustering and learning applications, *etc.* no more substantive help, existing detection accuracy is sufficient for real-time detection.

5. Conclusions

Video-based human motion analysis is one of the important topics in the field of computer vision. It involves pattern recognition, image processing, computer vision, artificial intelligence and other disciplines of knowledge. Video-based human motion analysis in an increasing number of areas has been applied, such as:! Video Surveillance sports analysis, medical image processing, online games, three-dimensional animation. With the development of theoretical research and related applications which will greatly affect people's lives, the establishment of a recovery model human motion tracking method in the related models, including the human skeleton model, human appearance model, the three-dimensional coordinates based on the model and model human motion. Skeleton model tree structure represents the topological body, established two coordinate systems, easy to use inverse kinematics to calculate the rotation angle of the joint.

References

[1] I. Haritaoglu, D. Harwood and L. Davis, "VV: real-time surveillance of people and their activities", IEEE Trans Pattern Analysis and Machine Intelligence, vol.22, no. 8, **(2000)**, pp. 809-830.
[2] M. M. Guo, S. L. Yang, H. Yan, L. F. Kan and B. Yang, "Mobile video alarm system based on cloud computing. Review of Computer Engineering Studies, vol. 1, no. 2, **(2014)**, pp. 5-10.
[3] C. Jian, W. Zhang and Y. Ying, "A micro-gesture recognition on the mobile web client", Review of Computer Engineering Studies, vol. 2, no. 2, **(2015)**, pp. 19-24.
[4] L. Wang, J. C. Fang and Z. Y. Zhao, "Application of backward propagation network for forecasting hardness and porosity of coatings by plasma spraying", Surface & Coatings Technology, vol. 201, no. 9-11, **(2007)**, pp. 5085-5089.
[5] L. C. K. Liau, T. C. K. Yang and M. T. Tsai, "Expert system of a crude oil distillation unit for process optimization using neural networks", Expert System with Applications, vol. 26, no. 2, **(2004)**, pp. 247-255.
[6] L. C. Jun, P. Binay and L. J. Min, "Stochastic nonlinear optimization for robust design of catalysts", Industrial and Engineering Chemistry Research, vol. 50, no. 7, **(2011)**, pp. 38-39.
[7] Y. Zhao, W. Wei, B. S. Hou and L. Wei, "The hardware design of intelligent circuit breaker", Energy Education Science and Technology Part A: Energy Science and Research, vol. 1, **(2014)**, pp. 695-702.
[8] K. Chisachi, "Proceedings of the American Society of Mechanical Engineers Fluids Engineering Division Summer Conference", Part B, **(2015)**, pp. 1273-1280.
[9] J. Dhinakaran, C. M. Ganapathy and T. Kodeswaran, "Quality Improvement of Lubricating Oil Pump Shaft through Statistical Process Control Used in Automobile Industry", Proceeding Engineering, vol. 2012, no. 38, **(2012)**, pp. 2053-2062.
[10] P. P. Fua, "Automatic Detection and Tracking of Human Motion with a View-Based Representation", Proceedings of the 7th European Conference on Computer Vision, vol. 2002, **(2002)**, pp. 476-491.
[11] A. F. Bobick and J. W. Davis, "The recognition of human movement using temporal templates", Pattern Analysis and Machine Intelligence, vol. 23, no. 3, **(2001)**, pp. 257-267.

Authors

Yongbo Wang, He received his Bachelor Degree in Art and Design from Department of Fine Arts, Henan Normal University, Henan, China in 2003, and his Master Degree in Arts in 2013 from the same university. He is a lecturer working in the Center of Modern Education Technology at Xinxiang Medical University. His current research interests include computer application technology, modern education technology, video and image processing, virtual reality and 3D animation. He has published more than ten academic papers, composed one academic monograph as deputy editor, and won 6 awards of all levels in his academic field.

Zheng Leiqi, Male, born in Sept.1978. He received his Bachelor Degree in Education Technology from Henan Normal University, Henan, China in 2003. He is working at the Center of Modern Education Technology at Xinxiang Medical University. His main research interests are video processing and program editing.