www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

RESEARCH ON CAMERA-BASED HUMAN BODY TRACKING USING IMPROVED CAM-SHIFT ALGORITHM

Jiude Li
Linyi University, College of Physical Education, Linyi, Shandong, China

Submitted: Jan. 31, 2015 ***Accepted: Apr. 10, 2015*** ***Published: June 1, 2015***

Abstract- Camera-based human body detection and tracking is an important research subject of computer vision, which has a widely used in the field of military and civil. In this paper, we focus on the technology of human body tracking based on improved camshaft algorithm. Firstly, we introduce some common image noise reduction algorithms. By combination the frame difference and background subtraction methods, an improved moving target detection algorithm is proposed, by which the whole region of target can be detected. Then, with the analysis of particle filtering and traditional Cam-shift algorithm, we introduce a new human body tracking method that is able to choose the target automatically due to the detection result. On the basis of the detection and tracking results, the algorithm of motion parameter estimation is analyzed. Finally, a set of human body detection and tracking experiments are designed to demonstrate the effectiveness of the proposed algorithms.

Index terms*:* Human body tracking; motion detection; Cam-shift tracking; motion parameter estimation.

I. INTRODUCTION

Motion tracking and detection based on computer vision is an important issue in the field of computer vision [1-4]. Testing, tracking of moving object in video sequences, obtain the target motion parameters such as position, speed, and is the basis of more advanced movement target behavior analysis. Rapid real-time detection of moving targets, the moving target tracking, and achieve the goal of sports data analysis, can be widely used in the field of sports. For example, on a track and field athletes training video processing, extract the instantaneous velocity, arms swinging amplitude, frequency and other information, how much per second image due to the video sequences, contains a lot of human eyes can't capture valuable information, can be used for the analysis of technical shortage athletes, improve sports skills.

All the information that the human can perceive, more than 80% can be achieved though visual. Let the computer have the ability as human is the goal of people struggle. Although the computer cannot perform efficiently, flexibly and intelligently as human being, people reach their goal step by step with the efforts of the researchers at home and abroad. Computer vision, expert system and natural language are the most active questions in artificial intelligence technology. With the progress of modern science and technology and the development of economic level, technologies based on computer vision blend in every corner of our life gradually. The computer vision technology can be used in many occasions, such as intelligent monitoring system, perceptual interface, motion analysis, and so on.

At present, scholars at home and abroad have made great researches for moving target detection, tracking and target behavior understanding based on computer vision. VASM [5] is a video monitoring system developed by several research institutions including Carnegie Mellon University, which can be used for behavior understanding automatically according to received video and can be used for reduction the monitoring costs in battlefield and avoiding the Carnegie Mellon university casualties in dangerous environment. Real-time video monitoring system W4 [6] at the University of Maryland can recognize and make location of human, and make extraction and segmentation operations of the various parts of body. By modeling the outline of human body, it can realize multiple human tracking and make semantic understanding of human behavior.

Frequency information is an important media, such as the video including video image and sound that are combined synchronously. In this paper, we focus on detection, tracking and analysis of motion target in video. Thus we only are interested in video images. According to the different application and purpose, the image processing can be divided into two cases, i.e., online processing and offline processing. Online processing usually is applied to higher real-time demand occasion, which calls for higher requirements of hardware. For example, the swimming pool monitoring system needs extraction, tracking, analysis, and making decision of the motion information in video to estimate the status of the swimmer. If danger drowning appears, it sends the information to monitor, in order to deal with the danger timely.

According to the numbers of cameras, video image processing can also be divided into monocular camera processing and multiple cameras processing. In monocular camera processing, it can be divided into two cases, i.e., static and dynamic. At present, most of the applications are done in the static situation. The reported researches always focus on the motion detection and tracking which are also the main contents in this paper.

Target detection, is a very key and basic link in the visual analysis, the purpose of which is extracting the foreground image form the video sequences. The performance of target detection plays an important role in the progress of video analysis. The accuracy and effectiveness of detection are the premise for effective target recognition and behavior analysis. As one of the key issues of video processing, lots of excellent algorithms for it have been proposed after decades of development. Meanwhile, there still have lots of difficult problem needs researching. The difficulties of motion detection are the background extraction and updating, the changes of light, shade disturbances, the changes of the background, and so on. So far, there has not a generic detection algorithm. The common used methods have optical flow method, background difference, frame difference, statistical method, and the mixing algorithms of above. Target tracking is the progress to determine the motion parameters of target, such as position, velocity, gesture, etc., according to the object in video sequence, which is the basis for behavior understanding. In general, the methods of target tracking can be divided into two types, deterministic algorithms and random algorithms. In random tracking algorithm, it tracks the target by estimating the state of target, such as particle filtering. The deterministic one is done by solving the optimal matching problems, such as mean shift algorithm.

In this paper, we will focus on the human body detection and tracking with the improved Cam-shift algorithms. According to the basic process of computer vision detection, the main work of this article can be divided into several aspects of image preprocessing, including target detection, image post-processing, target tracking and target motion measurements.

II. OVERVIEW OF HUMAN BODY DETECTION AND TRACKING

A. Video Image Processing

In general, the first important problem in video image processing is dealing with the image noise. Despite the current image acquisition device is more and more advanced, but the noise in the image acquisition process in inevitable. According to he causes, image noise can be divided into two types, i.e., external noise and internal noise [7, 14]. The external noise is always produced by the outer electromagnetic wave or power source of acquisition system. Internal noise can be divided into three types according to the noise source, i.e., Gaussian noise that caused by resistive element, Poisson noise (also called impulse noise) caused by photovoltaic conversion and grain noise in photographic progress.

In general, noise is random, which can be described by mathematical statistics theory and method. In the sense of statistical science, the noise can be divided into stationary and non-stationary noise. In realistic application, without considering its strict mathematical definition, noise can be seen as stationary if the statistical characteristics do not change over time. Otherwise, if the statistical characteristics change, we can regard it as non-stationary noise. If the amplitude of all noise in an image is most the same, however with random locations, the kind of this noise is generally referred as the impulse noise, as shown in figure 1 (a); If the amplitudes of noise in image distribute randomly as Gaussian, and the noise locates almost every point in the image, it can be called Gaussian noise, as shown in figure 1 (b).

(a).

(b).

Figure 1. Image with Gaussian and impulse noise

Usually we use mathematical statistical characteristics to describe the noise. The most important indicators are noise correlation function, the mean, variance, etc. A grayscale image can be treated as a signal with two dimensions, i.e., $f(x, y)$, the amplitude of which represents the brightness. Here, noise can be regarded as a distraction on brightness of image, denoted as $u(x, y)$. The overall strength of noise can be described by the mean of the noise, as shown in (1).

$$\begin{aligned}\bar{u} &= E\left(u\left(x,y\right)\right)\\ &= \frac{1}{M\times N}\Sigma_{x=1}^{M}\Sigma_{y=1}^{N}u(x,y)\end{aligned} \tag{1}$$

In (1), M and N represent the number of line and row respectively. And the variance σ^2 reflects the distribution of noise, as shown in (2).

$$\begin{aligned}\sigma^2 &= E\left(\left(u\left(x,y\right)-\bar{u}\right)^2\right)\\ &= \frac{1}{M\times N}\Sigma_{x=1}^{M}\Sigma_{y=1}^{N}\left(u(x,y)-\bar{u}\right)^2\end{aligned} \tag{2}$$

According to the influence of noise to image, the model of noise can be divided as additive noise and multiplicative noise. Define the output under noise is $g(x, y)$, and the additive noise model and multiplicative noise model can be described as in (3) and (4).

$$g(x,y) = f(x,y) + u(x,y) \tag{3}$$

$$\begin{aligned}g(x,y) &= f(x,y)\left[1+u(x,y)\right]\\ &= f(x,y) + f(x,y)n(x,y)\end{aligned} \tag{4}$$

From (4), the output image with multiplicative noise suffers from the image signal $f(x, y)$. The influence will be larger with larger amplitude of $f(x, y)$. For example, the grain noise on the negative will increase with larger $f(x, y)$.

The calculation of multiplicative noise is very complex, usually if with little change of signal $f(x, y)$, the second item in the formula (4) can be treated as constant, and the noise can be seen as additive noise. Additive noise often exists in the form of impulse noise or Gaussian noise. According to the characteristics of these two types of noise, the median filter or its improved form is often used for filtering impulse noise, and mean filter is used for treating Gaussian noise.

Median filtering is a nonlinear signal smooth processing technology based on statistical theory, which is a kind of neighborhood operation and is similar to the convolution operation in image processing [8-9]. Median filter can be used to eliminate the high frequency component in Fourier space, but also has an affect on the low frequency component. In the realization process of the median filter, a window template with 2 dimensions is often used, such as template with 3×3, or 3×3 regions. This window template slides over the whole image, the center position of which represents the pixel that is processing, and the size of which stands for the selection neighborhood. This filtering algorithm is used widely, but when the image contains the details such as dots, lines, median filter will makes these derail information blurred. In order to solve this problem, the median filer is improved by many scholars. For example, the weighted median filter treats this contradiction of filtering noise and retraining details of image by assignment different weights for pixels in the window. Some scholar put forward a kind of adaptive filter that handles impulse noise with greater probability. In this adaptive filter, the size of neighborhood region can be changed according to certain given conditions. This kind of filter retains the advantages for filtering impulse noise of traditional filter. For the non-smooth impulse noise, the details can be remained well with this adaptive algorithm, so as to have the ability to reduce the signal distortion of target on the borders of image.

The mean filter is a kind of linear filter, and is based on the local statistical information of image, the basic ideal of which is replacing the original value of pixel with the average value of pixels in the neighborhood region. The same as the median filter, the mean filter first choose a window template for the current pixel, and then assign the average value of the pixels in window to the center pixel (x, y). Suppose the gray value of pixel after processing is $g(x, y)$, then the principle of mean filter can be expressed as follows:

$$g(x,y) = \frac{1}{M} \Sigma_{f \in S} f(x,y) \tag{5}$$

B. Human Body Motion Detection and Tracking

✧ Motion Detection

Motion detection is the foundation of target tracking, target classification, behavior understanding. The main purpose of motion detection is removing the static background from the video sequence, detecting the moving target and its motion information, and finally extracting the moving target efficiently. However, in the most application environment, the background is not immutable. For example, according to the light changing, shadow would appear in the image, the branches sway caused by wind, and so on. These factors make motion detection much complex. As a result, in the practical application, we must eliminate the interference of background, so as to extract the motion area fully in the video sequence. The motion diction problem mainly can be divided into two categories, i.e., detection with fixed camera and detection with moving detection. For the motion detection problem with moving camera, the famous solution is the optical flow algorithm, which predicts the state of motion of camera through obtaining the optical flow field in image sequence by solving a partial differential equation. For the motion detection problem with fixed camera, the basic methods have background difference, time difference (also called frame difference), Gaussian background model, and so on. For the different applications, the improved motion detection methods based on the above ideas have been proposed one after another. Next, we will give an overview of the main detection method and the algorithm used this paper.

Optical Flow: This method is about the concept of motion detection, which is usually used for describing movement of targets, the surface or the edge, which are caused by the motion between target and observer. It has been widely used in pattern recognition, computer vision and other filed of image processing. The optical flow is founded on the following assumptions: the movement of target or movement of background is the only factor that causes the changes of image gray, i.e., the gray of target of background is constant while time goes on. The basic idea of motion detection based on optical flow is estimate the optical flow corresponding to motion according to a certain constraint condition, by the characteristics of the velocity field in image sequence caused by the changes of moving target. The advantage of optical flow is that it can deal with the detection problem when the displacement between frames is large. However, the calculation in optical flow requires repeated iteration computation, which causes a large time complexity. It is unable to be applied in the occasions with a requirement of real-time processing. So in many detection systems with optical flow needs supports of special hardware.

Frame Difference: The basic idea of this method is obtaining the contour of target by doing a difference operation of two adjacent frames in video sequences. It can deal with the situation of multiple moving targets and moving camera. When moving target appears in monitoring scenario, the image of different frame will appear relatively obvious difference. According to this phenomenon, obtain the brightness difference absolute value of two frame images through doing subtraction operation of two adjacent images, and judge there is a moving target or nor according to the difference of brightness is greater than a set threshold value or not. Denote the difference image between frames as D_k. In order to extract the moving target, do binarization processing at first, then do mathematical morphology filtering operation, and at last segment the moving target. The flowchart of motion detection based on frame difference is shown in figure 2.

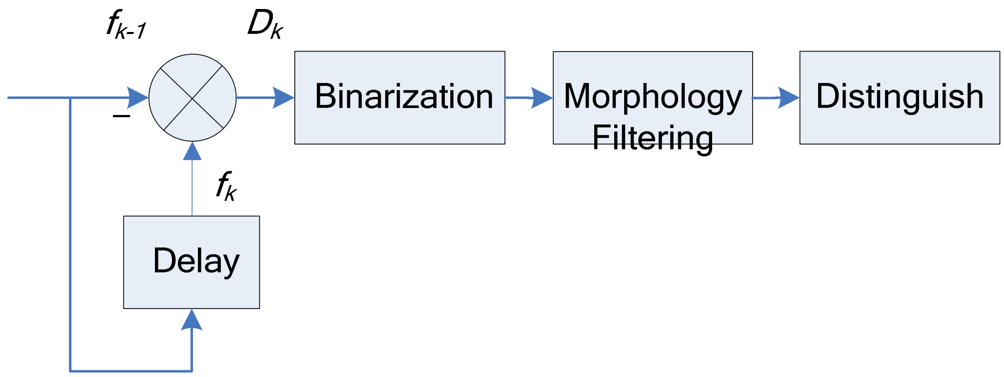


Figure 2. Frame difference method

Background Difference: It is a simple and effective target extraction algorithm, which has a good performance for static scene, but is unable to deal with situation when camera moves. Like frame difference method, it is doing subtraction of images, the difference is that it do the operation between the current frame and background, the mathematical expression of which is shown in (6).

$$R(i,j) = F(i,j) - G(i,j) \tag{6}$$

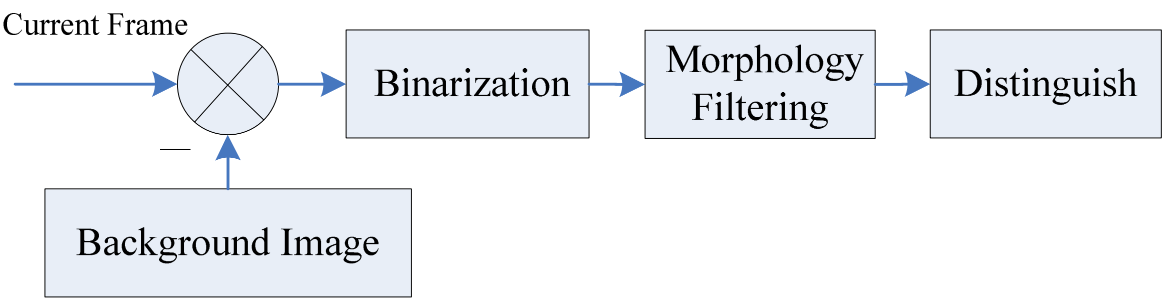


Figure 3Background difference methohd

In above formula, $F(i, j)$ is the current frame image, $G(i, j)$ stands for the background image and $R(i, j)$ represents the difference image. And then the binarization processing is needed, which can be expressed as in follows:

$$R(i,j) = \begin{cases} 1, & |F(i,j) - G(i,j)| > T \\ 0, & others \end{cases} \tag{7}$$

T in (7) stands for a gray threshold, which can control the sensitivity of target detection. After this operation, we can get binarization image of $R(i, j)$, in which the pixels with value 1 represent the moving target, the pixels with value 0 stands for the background image. The flowchart of this algorithm is shown in figure 3.

Statistical background model: In the static camera, the most important step of detection is the description of background modeling. Background modeling is the most fundamental problems in computer vision. Many scholars at home and abroad have tried variety mathematical models to describe the background, the most common used of which at present is the Gaussian distributed parameter model. In most practical application scenario, the background in the camera image sequence acquisition is changes over time. But in a short period, the change of background generally will not be obvious. So the Gaussian background model of every pixel in video image sequence can be set up, in which each pixel in image is treated to obey Gaussian distribution with mean μ and standard variance σ. Moreover, the distribution the pixel is independent of each other, which produces the modeling method of single Gaussian background and its improved forms.

✧ Target Tracking

The purpose of target tracking in computer vision is recording the interested positions of moving target in each frame in the continuous video image sequence, in order to obtain the trajectory of target for a period of time. Furthermore, information of direction, acceleration, velocity, and size of target can also be available. An ideal of target tracking algorithm needs to have a few characteristics, i.e., speed ability, robustness, simplicity. Now, we will give an overview of particle filtering and Cam-shift algorithm.

Particle filtering is a kind of Bayesian Filter through non-parameterized Monte Carlo algorithm, which can almost be used for nonlinear system described by state space model, as well as the nonlinear non-Gaussian system that can not be filtering with Kalman filter. The particle filter is flexible, and can be realized using parallel computing. Therefore, in the application of particle filter, the practical value of it is better than Bayesian filter (Kalman filter, mesh filter, etc.), which is often used to solve the problem of recognition, communication, economic and others. The progress of particle filter is shown as follows:

(1) Bayesian importance sampling:

$$\begin{cases} E\left(g\left(x_{1:t}\right)\right) = \dfrac{\Sigma_{i=1}^{N} g(x_{1:t}^{i}) w_t(x_{1:t}^{i}) / N}{w_t(x_{1:t}^{i}) / N} = \Sigma_{i=1}^{N}(x_{1:t}^{i})\tilde{w}_t(x_{1:t}^{i}) \\ \tilde{w}_t(x_{1:t}^{i}) = \dfrac{w_t(x_{1:t}^{i})}{\Sigma_{i=1}^{N} w_t(x_{1:t}^{i})} \end{cases} \tag{8}$$

(2) Sequence importance sampling:

$$w_t(x_{1:t}) = w_{t-1}(x_{1:t-1}) \frac{p(z_t \mid x_t) p(z_t \mid x_{t-1})}{q(x_t \mid x_{0:t-1}, z_{1:t})} \tag{9}$$

(3) Degradation: Usually, we use the effective sampling scale N_{eff} to describe the degradation of particle coefficients:

$$N_{eff} = \frac{1}{\Sigma_{i=1}^{N}(w_t^i)^2} \tag{10}$$

(4) Importance density function:

$$w_t^i \propto w_{t-1}^i p(z_t \mid x_t^i) \tag{11}$$

(5) Importance sampling:

The process of importance sampling is to keep the particles with larger weights for replicating, and weed out the particles with smaller weights. The process of importance sampling can be described as follows: first, recalculating the weights of particles according to the new observation and assign different value for weights according to the difference between observations and practical values. Then, the particles with large weights will have more and more "offspring" particles, and particles with smaller weights will have few "offspring" particles the weights of which will be assigned with $1/N$. In the third step, it is the process of state transition, in which add the particles of random prediction at time t. In the last step, it is the process of observation, in which the final state of particles can be obtained by the weight sum of all particles which is similar to step 1.

III. HUMAN BODY TRACKING AND MOTION PARAMETER ESTIMATION WITH IMPROVED CAM-SHIFT ALGORITHM

A. Vision Tracking with Mean-shift and Particle Filtering

✧ Introduction of Mean-shift Algorithm

Mean-shift is a common used algorithm in target tracking, which is a space analysis method based on non-parameterized core density estimation [10-11]. Under the situation that the probability density function is known, the position of maximum value can be determined by sampling data based on probability density function. Suppose that the sampling set obtained is x_i, i=1, 2, …, n and $x_i \in R^d$, where d stands for the dimension of sampling space. The core density estimation with core function $K(x)$ and width h can be expressed in (12).

$$f(x) = \frac{1}{nh^d}\Sigma_{i=1}^{n} K(\frac{x-x_i}{h}) \tag{11}$$

If using the contour function k(x) of K(x), the equation in (11) can be rewritten as in (12):

$$\hat{f}_{h,k}(x) = \frac{1}{nh^d}\Sigma_{i=1}^{n} k\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \tag{12}$$

Then the core density gradient function can be expressed as:

$$\begin{cases} \hat{\nabla} f_{h,k}(x) = \dfrac{2}{Ch^2} M_{h,G}(x) \hat{f}_{h,G}(x) \\ M_{h,G}(x) = \dfrac{\Sigma_{i=1}^{n} x_i g\left(\left\| \dfrac{x - x_i}{h} \right\|^2\right)}{\Sigma_{i=1}^{n} g\left(\left\| \dfrac{x - x_i}{h} \right\|^2\right)} \\ g(x) = -k(x),\, G(x) = Cg\left(\|x\|^2\right) \end{cases} \tag{13}$$

In (13), C is a given constant.

(13) Indicates that the Mean-shift vector is proportion to the normalized density gradient estimation obtained by core function, i.e., the direction of the core vector is the same as the direction of the increasing of largest degree. Based on Mean-shift algorithm, through iteration computation to make the searching window move to the position of real target, it can reach the purpose of target tracking.

Next, we will show the flowchart of vision tracking based on Mean-shift algorithm.

1. Form the motion model of target. First, select the target window in the first frame and determine the initial searching position. And then, calculate the histogram of the image in searching window, after which the modeling of target motion is completed.

2. Form the motion model of target to be selected. In this step, obtain the model to be selected in the region that the target may lie in by particle filter from the video image sequence.

3. Calculate the Mean-shift of target. Compare the similar degree between motion model in initial frame and motion model in current model using the likelihood function. The principle of calculating the Mean-shift is make the likelihood function maximum (Bhattacharyya coefficient is adopted in this paper), by which the vector obtained is the one from initial position to current position.

4. Similar degree threshold judgment. If the similar degree above is smaller than the threshold stetted, turn to step 5, or else go to step 6.

5. Generate particles randomly using particle filtering and calculate the weight of each particle. And then update particles to calculate the average estimation of target dynamic state.

6. Make the current tracking window as the target region in the next frame image and return to step 2.

B. Improved Cam-shift Algorithm with Selection Target Automatically

The core idea of Cam-shift is similar to the Mean-shift algorithm, which is a kind of non-parameter probability estimation algorithm. It is widely used in target tracking and has the advantage of real-time and framework. This algorithm works well for human body tracking.

Cam-shift is founded on the information of colors, which determine the position and size of target in video image sequence using the color information and initialized the searching window of next frame image using the current position and size of target [12-13,15-16]. It tracks the target through redoing the above operations. As the initial searching position is that of previous frame, and the time interval between two adjacent images is short, the distance that target moves is also short, this method works well. This progress has a advantage in real-time than Mean-shift. As the color information of target changes little when moving, it is more robust. The traditional Cam-shift algorithm is shown in figure 4.

The traditional Cam-shift is a semi-automatic tracking algorithm, in which the initialized searching window that would lead to a bad practicability. In this paper, we will improve the traditional algorithm by an automatic tracking region setting.

The key problem of automatic tracking is extracting inscribed rectangular. In this paper, we use a kind of filling ratio algorithm for the extraction operation.

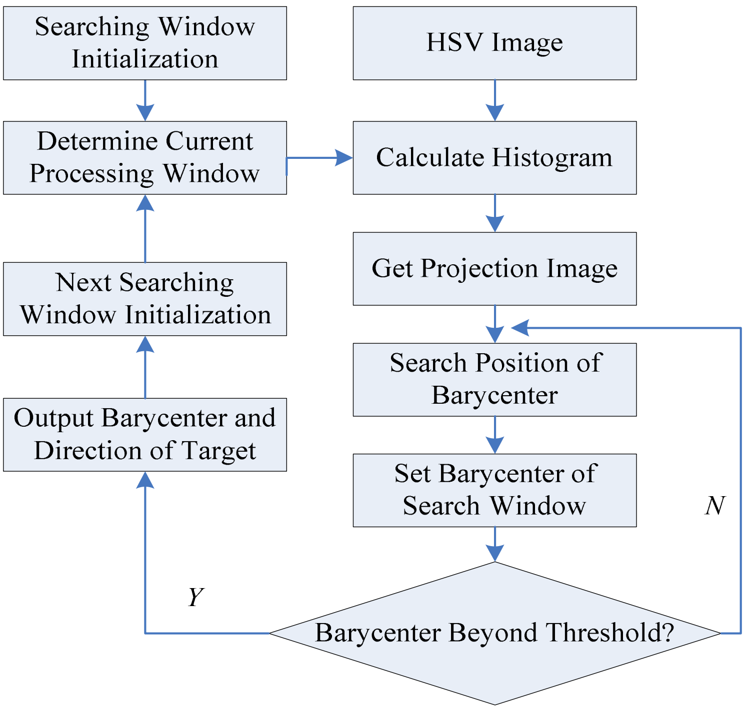


Figure 4. The flowchart of Cam-shfit algorithm

In a practical algorithm, it is hard to extract an ideal inscribed rectangular, especially when needs a high real-time performance. Thus, approximate inscribed rectangular can be considered. The filling ratio algorithm changes the size of rectangle according to the area ratio that the target. In the calculation progress, the circumscribed rectangle should be determined at first, and then shrink this rectangle until it is close to the inscribed rectangular of target. Now we will give the expression of filling ratio algorithm. Denote the circumscribed rectangle area with size $M \times N$ by matrix D. As the target detection algorithm has done binary processing of moving target, thus the pixel value of target is 1, and the pixel value of background is 0. The matrix D can be expressed as follows:

$$\begin{cases} D = \begin{bmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & \ddots & \vdots \\ a_{M1} & \cdots & a_{MN} \end{bmatrix} \\ a \in \{0,1\} \end{cases} \tag{14}$$

Then the filling ratio of row i, column j, and area in this rectangle can be defined as:

$$\begin{cases} \eta_r^i = \dfrac{\Sigma_{j=1}^{N} a_{ij}}{N} \\ \eta_c^j = \dfrac{\Sigma_{i=1}^{M} a_{ij}}{N} \\ \eta = \dfrac{\Sigma_{i=1}^{M} \Sigma_{j=1}^{N} a_{ij}}{M \times N} \end{cases} \tag{15}$$

Then the flowchart of filling ratio algorithm can be expressed as follows:

1. Determine the circumscribed rectangle of target;

2. Calculate the filling ratio of rectangle area η, and let η_{last} equals to 0. η_{last} represents the filling ratio of the whole area after last calculation. And the initial value ratio can be assigned to 0;

3. Set a threshold ρ of filling ratio. If the rectangle closer to target is obtained, then go to step 4, or do the following circulation:

(1) Calculate the filling ratio of each row and each column;

(2) Decrease the number of row and column according to the filling ratio of row and coverage of column respectively;

(3) Let η_{last} equals to η, and calculate the filling ration of new area.

4. Output the rectangle area, and obtain the area of automatic target tracking.

This progress is shown in figure 5.

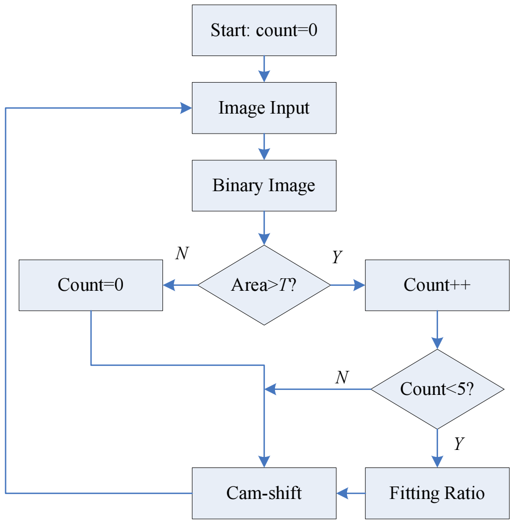


Figure 5. The flowchart of Cam-shfit algorithm

C. Location and Parameter Estimation of Human Body

The key problem for motion parameter detecion is lacation. The goal of target location is mark the position information of target in the video image sequence, so as to make a sufficient prepare for motion parameter calculation. The position of target can be expressed as particle filtering or Cam-shift algorithm. Furthermore, in this paper, we try to use binary image for location with barycentre method and maximum projection method.

✧ Velocity Calculation

When programming the procedure of velocity calculation, a global variable can be set for expressing time t. Record time t of each frame image, and calculate the lag Δ_t of two images between frame k and frame k+1. Then the instantaneous velocity of human body can be approximately expressed as follows:

$$v = \frac{ds}{dt} \approx \frac{\Delta s}{\Delta t} \tag{16}$$

The velocity expressed in (16) is a kind of discrete instantaneous velocity. In this paper, the full expression is given with least square algorithm, which is shown in (17).

$$\Sigma_{i=0}^{m}\left(\varphi^*(x_i)-y_i\right)^2 = \min_{\varphi\in\phi} \Sigma_{i=0}^{m}\left(\varphi(x_i)-y_i\right)^2 \tag{17}$$

In (17), φ* is the expression of fitting curve with least square algorithm, and Φ is the set of polynomial. In order to reduce the computation, second order polynomial is used in this paper. (x_i, y_i) is the sampling data.

✧ Magnitude Calculation

The magnitude in this paper means the swing magnitude of arms when human moves. Through analysis of human's motion state when walking, the histogram in vertical direction of binary image can be obtained with target location. As the noise exists, there are often several non-zero curves in the histogram.

As the pixel area of noise is far less than that of target, and the mean of noise projection is less than that of target, so the width of line with high mean value can be seemed as the width of target. That is to say, it almost equals to the swing magnitude of human arms.

✧ Frequency Calculation

Frequency in this section means the frequency of arm swing, i.e., the reciprocal of the time interval between two continuous swings. In practical image processing application, first record the swing magnitude and the corresponding time of each frame image, and save them in a structure chained list. Then, find out the maximum magnitude swing with local extreme method and the corresponding time interval.

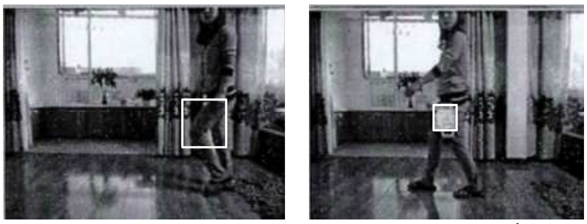
✧ Projection Between Pixels and Real-distance

For the position information obtained above is expressed in pixel form, the relationship of it to real distance need to be determined. In this paper, we treat the stature of human as the prior information. Use the maximum of histogram in vertical direction projection by maximum projection method to represent the pixel number of human stature. According to the imaging principle of camera, targets with same size may have different pixel number in video image when the position or angle changes. So, when calculating the real distance expressed by pixel, each frame needs a computation.

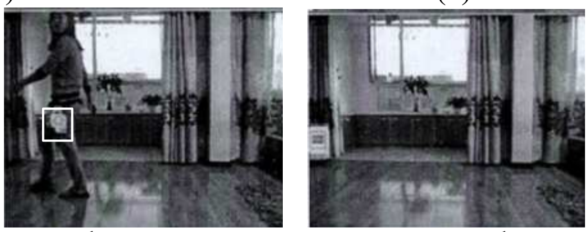
IV. EXPERIMENTS AND ANALYSIS

Next, we selet a video sequence of human walking in AVI form. In oder to test the performance of improved Cam-shift algorithm proposed in section III, we choose the region of leg with fast velocity and high frequence for tracking. The tracking reslut is shown in figure 6.

From the tracking experiment results in figure 6(a) to figure 6(d), we can see that it performs a good tracking with the improved Cam-shift algorithm. Then we will test the motion parameter estimation method proposed in the previous section. A part of experiment data is shown in table 1.



(a) Initial frame (b) 80th frame



(a) 165th frame (b) 220th frame

Figure 6.Tracking performance with improved Cam-shift algorithm

Tabel 1. A part of experiment data

Frame	72	73	74	75	76	77
Position	53	97	139	179	230	268
Time(s)	4.057	4.113	4.168	4.220	4.278	4.335

The displacement curve to time by least square fitting method is:

$$s^* = 153.4829t^2 - 490.0708t - 482.8427 \qquad (18)$$

Then the velocity expression can be obtained by difference operation of formula in (18):

$$v = 306.9658t - 490.0708 \qquad (19)$$

As the velocity is calculated by difference of position, so the velocity error can be expressed by position error:

$$\begin{cases} \Delta s(4.057) = s^*(4.057) - s(4.057) = 0.6434 \\ \Delta s(4.113) = s^*(4.113) - s(4.113) = -1.3852 \\ \Delta s(4.168) = s^*(4.168) - s(4.168) = -0.4852 \\ \Delta s(4.220) = s^*(4.220) - s(4.220) = 3.3434 \\ \Delta s(4.278) = s^*(4.278) - s(4.278) = -2.8994 \\ \Delta s(4.335) = s^*(4.335) - s(4.335) = 4.4786 \end{cases} \qquad (20)$$

From the error displayed in (20), we can see that the velocity calculated by least square algorithm is less than 3 percent.

V. CONCLUSION

The purpose of human body detection is detecting out the target, extracting the entire target and doing binary operation for motion parameter estimation. The basic calculation requirement of motion parameter estimation in this paper is the good performance of target extraction.

In this paper, we first analyzed the advantages and shortages of background difference, frame difference and optic flow. For the human body tracking is an important problem for behavior understanding, we proposed a new tracking algorithm based on improved Cam-shift algorithm, which can set the tracking region automatically. Using this algorithm, it can converge to the inscribed rectangle closer to the real target much faster with rectangle expression for target. And with the full use of target information, it performs a good robustness. Then based on least square method, the motion parameter estimation algorithm was also put forward using the information of detection and tracking.

Although the basic function can be realized based on the algorithms proposed in this paper, however, there are some problem needs further research. For example, until now, there is not still a general algorithm for practical application, and the universality and reliability need to be improved. Moreover, the algorithm accuracy of motion parameters estimation needs to be improved in the future. To obtain a more accurate projection between information in video image sequence and in real world, the imaging characteristic should also be considered. The algorithms proposed in this paper involve a large amount of data processing, how to improve the real-time capability on hardware platform needs more and more researches.

.

VI. REFERENCES

[1] A. Kyme, S. Se, S. Meikle, G. Angelis, W. Ryder, K. Popovic, D. Yatigammana and R. Fulton. “Markerless Motion Tracking of Awake Animals in Positron Emission Tomography”, IEEE Transactions on Medical Imaging, 2014, 33(11): 2180 - 2190, doi: 10.1109/TMI.2014. 2332821.

[2] P. J. Noonan, J. M. Anton-Rodriguez, T. F. Cootes, W. A. Hallett and R. Hinz. “Multiple target marker tracking for real-time, accurate, and robust rigid body motion tracking of the

head for brain PET", 2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2013: 1 - 6, doi: 10.1109/NSSMIC.2013.6829268.

[3] R. Krishnan and B. Natarajan. "Two-Stage Approach for Detection and Reduction of Motion Artifacts in Photoplethysmographic Data", IEEE Transactions on Biomedical Engineering, 2010, 57(8): 1867 - 1876, doi: 10.1109/TBME.2009.2039568.

[4] A. Heinrich, G. Di, D. Znamenskiy, J. P. Vink and G. de Haan. "Robust and Sensitive Video Motion Detection for Sleep Analysis", IEEE Journal of Biomedical and Health Informatics, 2014, 18(3): 790 - 798, doi: 10.1109/JBHI.2013.2282829.

[5] R. Col. A. Lipton and T. Kanade. "A System for video surveillance and monitoring: VSAM final report", Carnegie Melon University, Pittsburgh, America, 2000.

[6] D. S. Lee. "Effective Gaussian mixture learning for video background subtraction", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 827- 832, doi: 10.1109/TPAMI.2005.102.

[7] Joaquim de Mira Jr., Hugo Vieira Neto, Eduardo B. Neves, Fábio K. Schneider,Biometric-oriented Iris Identification Based on Mathematical Morphology, Journal of Signal Processing Systems, 80(2):181-195, 2015..

[8] K. Xiangui, M. C. Stamm, P. Anjie and K. J. R. Liu. "Robust Median Filtering Forensics Using an Autoregressive Model", IEEE Transactions on Information Forensics and Security, 2013, 8(9): 1456 -1468, doi: 10.1109/TIFS.2013.2273394.

[9] C. S. Regazzoni and A. Teschioni. "A new approach to vector median filtering based on space filling curves", IEEE Transactions on Image Processing, 1997, 6(7): 1025 -1037: doi: 10.1109/83.597277.

[10] D. Barash and D. Comaniciu. "Meanshift clustering for DNA microarray analysis", IEEE Computational Systems Bioinformatics Conference, 2004: 578 - 579, doi: 10.1109/CSB.2004.1332503.

[11] S. Yamada, K. Chomsuwan, S.C. Mukhopadhyay, M. Iwahara, M. Kakikawa and I. Nagano, "Detection of Magnetic Fluid Volume Density with a GMR Sensor", Journal of Magnetics Society of Japan, Vol. 31, No. 2, pp. 44-47, 2007.

[12] D. Kai, J. Yongfeng, J. Yinli, L. Gang, L. Yanyan Li and Q. Shenglong. "Object tracking based on improved Mean-Shift and SIFT", International Conference on Consumer Electronics, Communications and Networks, 2012: 2716 - 2719, doi: 10.1109/ CECNet.2012.6201691.

[13] P. Hidayatullah and H. Konik. "CAMSHIFT improvement on multi-hue and multi-object tracking", 2011 International Conference on Electrical Engineering and Informatics, 2011: 1-6, doi: 10.1109/ ICEEI.2011.6021825.

[14] S.C. Mukhopadhyay, K. Chomsuwan, C. Gooneratne and S. Yamada, "A Novel Needle-Type SV-GMR Sensor for Biomedical Applications", IEEE Sensors Journal, Vol. 7, No. 3, pp. 401-408, March 2007.

[15] Y. Shujun, C. Xiaodong, W. Sen, H. Zhihai , W. Yi and Y. Daoyin. "Cam-shift algorithm -based moving target recognition and tracking system", 2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS), 2012: 181 - 185, doi: 10.1109/VECIMS. 2012.6273211.

[16] Huang-Nan Huang, Shuo-Tsung Chen, Muh-Shi Lin, Woon-Man Kung, Chih-Yu Hsu,Optimization-Based Embedding for Wavelet-Domain Audio Watermarking, Journal of Signal Processing Systems,80(2): 197-208, 2015.

[17] N. K. Suryadevara and S. C. Mukhopadhyay, "Determining Wellness Through An Ambient Assisted Living Environment", IEEE Intelligent Systems, May/June 2014, pp. 30-37.

[18] N.K. Suryadevara, S.C. Mukhopadhyay, R. Wang, R.K. Rayudu, Forecasting the behavior of an elderly using wireless sensors data in a smart home, Engineering Applications of Artificial Intelligence, Volume 26, Issue 10, November 2013, Pages 2641-2652, ISSN 0952-1976, http://dx.doi.org/10.1016/j.engappai.2013.08.004.

[19] Pan Feng, Wang Xiaojun and Wang Weihong,research and restoration technology of video motion target detection based on kernel method, International Journal on Smart Sensing and Intelligent Systems, vol.7, no.4, pp.1516-1534, 2014.

[20] Shaoping Zhu and Yongliang Xiao, Intelligent detection of facial expression based on image, International Journal on Smart Sensing and Intelligent Systems, vol.8, no.1, pp. 581-601, 2015.