Optimized Moving Body Behavior Recognition Model Based on Multi Texture Gradient Feature

Mingxi Zhang
Langfang Teachers University, Langfang 065000, Hebei, China

Abstract
In order to guarantee the validity of the feature extraction, and more accurately identify moving body, an optimized moving body behavior recognition model is put forward based on multi texture gradient feature. Firstly, a similarity transformation between target model and candidate model is defined. In data processing, background weighting based on histogram in initial target model is introduced to optimize the background gradient. Then some particles are sampled in the iteration optimization of Mean shift algorithm and the gradient features are optimized by the multiplicative fusion of color and LBP texture features. The simulation experiments show that this improved Mean shift algorithm has better convergence and higher accuracy in moving human behavior recognition.

Key words: Gradient Feature Optimization, Mean Shift Algorithm, Moving Body, Behavior Recognition, Multiplicative Fusion.

1. INTRODUCTION

Human motion analysis refers to the detection and tracking of human body collected in the video, and the understanding and description of human behavior, which belongs to the category of image analysis and image understanding (Aggarwal and Ryoo, 2013). The purpose of the analysis and understanding of human behavior is to make computer system learn the knowledge and experience from the human operator, so that it can automatically identify the human behavior or abnormal detection, thereby reducing the burden of human. A sound solution to human behavior analysis can be applied to many aspects, including the intelligent video surveillance, human-computer interaction, sports analysis, sports training, medical diagnosis, video recovery and so on (Weinland et al., 2012). Human motion analysis based on video is to research and develop the ability of a computer to understand human movement, which has a broad application prospect. In practice, there is still a lot of technology imperfection. Therefore, the research of this technology has great practical significance with enormous challenges, which will lead to the development of other related fields.

Human behavior recognition is essentially a classification problem. There are two methods to solve this problem: one is based on template matching, the other is based on the state model. The template matching method is to transform the motion image sequences into one or a set of templates, matching the sample template and known template by calculating the distance between them. And the template which has the smallest distance to known template is the belonging behavior classification of sample (Gritai et al., 2014). Bobick and Davis represented the human behavior in the image sequences by the motion energy image and motion history image, and then used the Markov distance classifier to identify the human behavior (Hu and Zhao, 2013). Wang used the average motion energy (AME) and mean motion shape (MMS) templates to describe human behavior, identifying human behavior by using the nearest classifier (Bui et al., 2012). Yilmaz et al. proposed the space-time convolution (STV) template method to describe the behavior, fusing the human behavior as a point set of 3D temporal volumes, and using geometric features to measure the similarity between templates (Lin et al., 2013). Weinland adopted the moving history volumes (MHV) method to describe the process of human behavior, extracted the feature vector in the MHV model, and then used the Fisher classifier to identify the behavior (Huang et al., 2013).

In addition, Veeraraghavan et al. also used the dynamic time warping (DTW) technology to perform human behavior matching. The purpose of dynamic time warping is to map the time axis of the candidate behavior template to the time axis of the known behavior template in a nonlinear way, which makes least the cumulative loss of them (Gall et al., 2014). Gritai et al. also used the DTW to identify the behavior. The method based on template matching method has low computational complexity and is easy to implement, but the robustness is not good enough, and it is sensitive to the behavior time interval and noise (Hu, 2013). The method based on the state model describes the behavior of a certain kind of behavior consisting of a series of states which represent the static state or dynamic state and connect to each other with some probability. Yamato et al first carried out the work of using HMM to identify human behavior. They trained the HMM with standard learning algorithm, and finally recognized the different actions of the tennis players in the database. Their work proved that HMM can effectively model the characteristics change in human behavior cycle (Park and Byun, 2013). Brand et al.
proposed CHMM to model two random but connected processes in behavior, and applied the model for sign language recognition. Ren and Xu combined the method based on the characteristics and CHMM to identify a variety of behavior in the smart classroom (Wagner et al., 2012). Fine et al proposed hierarchical hidden Markov model (HHMM) which can express more clearly the behavior details of the human motion in different chromatographic level (Feris et al., 2012). The method based on the state model can be more accurate for the model of human body movement, but it needs a large quantity of training and complex calculation.

In light of the defects of Mean shift algorithm in practical application, this paper put forward a moving body recognition model based on Mean shift algorithm optimized by gradient features and conducted simulation experiment. The experimental results show that the proposed algorithm is effective.

2. PERFORMANCE ANALYSIS OF MEANSHIFT ALGORITHM

Current moving body feature extraction algorithms mainly include SIFT algorithm and Mean shift algorithm. This paper firstly introduces the performance analysis of these two algorithms.

SIFT algorithm extracts the key points of object form image. These feature points are invariant with the scaling and rotation of image, varying illumination intensity and camera angle.

The first step for key point detection is to detect the pixel points which are insensitive to the image size at different scales. The scale space of image is defined as the function $L(x, y, \sigma)$ convoluted by input image $I(x, y)$ and Gaussian kernel function $G(x, y, \sigma)$ with different scales.

$$L(x, y, \sigma) = G(x, y, \sigma) \cdot I(x, y) \quad (1)$$

Here, $\cdot$ is the convolution operation about $x$ and $y$.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

To effectively detect the position of key point in image scale space, the differential Gaussian function $D(x, y, \sigma)$ is used and calculated by the subtraction of image convolution result at two neighboring scale.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, (k-1)\sigma)) \cdot I(x, y)$$

$$= L(x, y, k\sigma) - L(x, y, (k-1)\sigma) \quad (3)$$

Then the obtained extreme point is set as candidate key point. Taylor expansion is utilized for this scale space function $D(x, y, \sigma)$.

$$D(X) = D + \frac{\partial D}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \quad (4)$$

Where $D(X)$ and its expanded item is calculated at the sampling point and $X = (x, y, \sigma)^T$ is the compensation. When the expanded item is zero, the position of corresponding extreme point $X$ is determined by following equation.

$$\hat{X} = \frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \quad (5)$$

In concrete calculation, the difference between neighboring sampling points is approximated as the scale space function $D(x, y, \sigma)$ and its expansion. If the compensation $\hat{X}$ is larger than 0.5 in each dimension, then the distance between current sampling point and extreme point is not the shortest. Then current sample point is removed and another point is used for calculation in same steps. Finally, $\hat{X}$ adds to the corresponding sample point to obtain the approximation around the extreme point.

$D(\hat{X})$ of scale space function in extreme point is used to exclude those unstable extreme points with low contrast ratio. Equation (5) is added to equation (4) to get the following expression.

$$D(\hat{X}) = D + \frac{1}{2} \frac{\partial D}{\partial X} \hat{X} \quad (6)$$

Through the setup of threshold value, the extreme point with $|D(\hat{X})|$ lower than threshold value is abandoned. Traditional threshold value is set as 0.05.
Mean shift algorithm is a smooth and null-parameter density estimation method. Whether it is convergent is the core of whether the tracking of object is accurate. The convergence of Mean shift and the increase of value equals to the proof question: if core function is defined by a descending concave function \( k(x) \), then series \( \{y_j\}_{j=1}^\infty \) and \( \{\hat{f}_{h,k}(j)\} \) converge by descending order.

Because the \( n \) is finite, \( \{\hat{f}_{h,k}(j)\} \) is bounded, according to the definition of \( \{\hat{f}_{h,k}(j)\} \),

\[
\{\hat{f}_{h,k}(j)\} - \hat{f}_{h,k}(j) = c_{k,d} \frac{1}{nh^{d/2}} \sum_{i=1}^n \frac{y_{j+1} - y_j}{h} \left[ \frac{y_j}{h} - x_i \right]^2 - k \left[ y_j - x_i \right]^2 \]
\]

(7)

According to the definition of concave function, for all \( x_i, x_j \in [0, \infty) \) and \( x_i \neq x_j \), there is

\[
k(x_j) \geq k(x_i) + k'(x_i)(x_j - x_i)
\]

(8)

Due to \( g(x) = -k'(x) \), equation (8) can be transferred to,

\[
k(x_j) - k(x_i) \geq +g(x_i)(x_j - x_i)
\]

(9)

Adding it to equation (7),

\[
\hat{f}_{h,k}(j+1) - \hat{f}_{h,k}(j) \geq c_{k,d} \frac{1}{nh^{d/2}} \sum_{i=1}^n \frac{y_{j+1} - y_j}{h} \left[ \frac{y_j}{h} - x_i \right]^2 - k \left[ y_j - x_i \right]^2
\]

(10)

There is,

\[
\hat{f}_{h,k}(j+1) - \hat{f}_{h,k}(j) \geq c_{k,d} \frac{1}{nh^{d/2}} \left[ y_{j+1} - y_j \right] \left[ \frac{\sum_{i=1}^n \frac{y_j - x_i}{h}}{\frac{\sum_{i=1}^n \frac{y_j - x_i}{h}}{h}} \right] \]

(11)

For \( k(x) \) is monotonically decreasing when \( x \geq 0 \), \( \sum_{i=1}^n \frac{y_j - x_i}{h} \) is a positive real number. Therefore, if \( y_{j+1} \neq y_j \), \( \frac{c_{k,d}}{nh^{d/2}} \left[ y_{j+1} - y_j \right] \left[ \frac{\sum_{i=1}^n \frac{y_j - x_i}{h}}{\frac{\sum_{i=1}^n \frac{y_j - x_i}{h}}{h}} \right] \) is non-negative. \( \{\hat{f}_{h,k}(j)\} \) is proved to be a increasing series. It has boundary so it is a convergent series.

Then, continuous \( m \) items of equation (11) are accumulated to obtain,

\[
\hat{f}_{h,k}(j+m) - \hat{f}_{h,k}(j) \geq c_{k,d} \frac{1}{nh^{d/2}} \left[ y_{j+m} - y_j \right] M
\]

(12)

\( M \) means the minimum of \( m \) items \( \sum_{i=1}^n \frac{y_j - x_i}{h} \). It is seen that \( \{y_j\}_{j=1}^\infty \) is a Cauchy series and convergent too. 10 different image series are adopted test the feature extraction of moving human body. 100 observing series are selected from each image sequence to see the accuracy of feature extraction. The statistic results are plotted in following image.

The performance comparison results show that Mean shift algorithm has better performance. Therefore, this paper chooses the Mean shift algorithm for the identification of human motion behavior and improves it in light of its defects.

3. HUMAN MOTION BEHAVIOR RECOGNITION MODEL BASED ON GRADIENT FEATURE OPTIMIZATION

3.1. Background Gradient Weight Based on Histogram

This paper chooses \( \{\hat{O}_a\}_{a=1}^\infty \) to represent the background model which calculates the histogram information of target area. Here, background area is set three times as large as target area. In \( \{\hat{O}_a\}_{a=1}^\infty \), \( \sigma^* \) represents the minimum non-zero value with its limiting condition.
\[
\begin{align*}
\nu_e &= \min \left( \frac{\hat{\sigma}_h}{\hat{\sigma}_x}, 1 \right) \\
\left\{ v_e = \min \left( \frac{\hat{\sigma}_h}{\hat{\sigma}_x}, 1 \right) \right\}_{u=1, \ldots, m} 
\end{align*}
\]  

This limit condition is used to define a similarity transformation between target model and candidate target model. The weight \( \nu_e \) is used to decrease the background part similar to target feature.

\[
\begin{align*}
\hat{p}_u(y) &= C_k v_u \sum_{i=1}^{n} k \left( \left\| y - x_i \right\| / h \right) \delta[b(x_i) - u] \\
C_k &= \frac{1}{\sum_{i=1}^{n} k \left( \left\| y - x_i \right\| / h \right) \sum_{u=1}^{m} v_u \delta[b(x_i) - u]} \\
\end{align*}
\]

Figure 1. Performance comparison results of the two algorithms

At the same time, the new target model is updated to,

\[
q_u = C_v \sum_{i=1}^{n} k \left( \left\| x_i \right\| \right) \delta[b(x_i) - u] 
\]

The normalized constant \( C \) is,

\[
C = \frac{1}{\sum_{i=1}^{n} k \left( \left\| x_i \right\| \right) \sum_{u=1}^{m} v_u \delta[b(x_i) - u]} 
\]

The target candidate model is,

\[
\hat{p}_u(y) = C_k v_u \sum_{i=1}^{n} k \left( \left\| \frac{y - x_i}{h} \right\| \right) \delta[b(x_i) - u] 
\]

Normalized constant \( C_k \) is,

\[
C_k = \frac{1}{\sum_{i=1}^{n} k \left( \left\| \frac{y - x_i}{h} \right\| \right) \sum_{u=1}^{m} v_u \delta[b(x_i) - u]} 
\]

In the process of background transformation, the target is to decrease the effect of major background features in the target area.

Through the iteration of Mean shift algorithm, the weight of candidate target area determines the convergence time of tracking algorithm. Only when the feature expression in background declines, the corresponding information of target area decreases.

In the background weighting method, the meaning of each weight parameter is analyzed. The weight \( w_i \) of \( x_i \) in candidate target area is calculated, making \( u \) become the bin value of feature space, then \( \delta[b(x_i) - u] = 1 \) is obtained corresponding to \( x_i \). It can be simplified into,

\[
w_i = \sqrt{w_{i-u}} / \hat{p}_u(y) 
\]

Adding equation (14) and (16) into this expression, there is,

\[
w_i = \sqrt{\frac{C_v \sum_{i=1}^{n} k \left( \left\| x_i \right\| \right) \delta[b(x_i) - u]}{C_k v_u \sum_{i=1}^{n} k \left( \left\| \frac{y - x_i}{h} \right\| \right) \delta[b(x_i) - u]}} 
\]

By removing shared parameter \( v_u \), and adding \( C \) and \( C_k \) into the denominator and numerator, there is,
\[ w_i = \sqrt{\frac{C'C_h}{C'C}}w_i \]  

Equation (20) is the improvement of weight \( w_i \) followed by the iteration at the predicted position by this algorithm. It is found that,

\[ y_i = \frac{\sum_{j=1}^{n_i} {y_j g_{i,j} w_i}}{\sum_{j=1}^{n_i} g_{i,j} w_i} \]  

From above equation, the iteration number of Mean shift doesn’t change very much. It is also seen that the purpose of background weighting method is to abate the effect from background information, but in Mean shift algorithm, the iteration number doesn’t decrease. It is to say that this method has little improvement on the algorithm. To ensure the application of background weighting method in target model, this paper chooses the batch treatment, namely adding background weighting method in data processing and initial target model while abandoning it in candidate target. After small adjustment, a new weighting equation is defined,

\[ w_i = \sqrt{\frac{{\C'}{C}}{C}} \times \sqrt{v_i} \times \sqrt{\frac{q_i}{p_i}} \]  

Then by removing the influence from constant \( \sqrt{\frac{C'}{C}} \), the weight can be simplified,

\[ w_i' = \sqrt{v_i} w_i \]  

3.2. Gradient Feature Optimization Based on Multiple Textures

To effectively avoid the particle degeneracy, this paper adopts some particles in the iteration optimization of Mean shift algorithm, the weight \( \lambda_i \) here is realized by the combination of weights of two features. Considering the contribution of color and LBP texture feature in target description, different adjustment is required for different tracking environment. In the improved algorithm, the weight fusion way of Mean shift algorithm is,

\[ \lambda_i = \alpha \lambda_i^c + (1-\alpha) \lambda_i^l \]  

Here,

\[ \lambda_i^c = \sum_{x_i} \delta(C(x_i) - \mu_c) \frac{q_i}{p_i(y)} \]  

\[ \lambda_i^l = \sum_{x_i} \delta[LBP(x_i) - \mu_l] \frac{q_i}{p_i(y)} \]  

These are the weights of Mean shift algorithm about color and texture features. \( \alpha \) is determined according to different tracking environment. This paper set the \( \alpha \) as 0.5.

In smoothing process, the calculation of each particle is realized through multiplicative fusion of color weight and LBP texture weight. The fusion expression is written as follows,

\[ w_i = w_i' w_i \]  

\[ \lambda_i = \frac{1}{2\pi\sigma^2} \exp\left( -\frac{(d_i')^2}{2\sigma^2} \right) \]  

\[ w_i' = \frac{1}{\delta\sqrt{2\pi}} \exp\left( -\frac{(d_i')^2}{2\sigma'^2} \right) \]
\[ w'_j = \frac{1}{\delta \sqrt{2\pi}} \exp \left( -\frac{(d'_j)^2}{2\sigma^2} \right) \]  

(29)

Those are the weights based on color and texture features for each particle in the particle filtering process. \( \sigma \) represents the variance of Gaussian distribution.

4. SIMULATION EXPERIMENT

To prove the effectiveness of the algorithm, this paper conducted the simulation experiment. Firstly, the convergence of improved Mean shift algorithm is analyzed and compared with that of standard Mean shift algorithm, as shown in figure 2.

![Figure 2. Convergence analysis results of improved algorithm](image)

Then, a moving human body video is used to conduct the human behavior recognition of Mean shift algorithm and compare it with that of SIFT algorithm. Figure 3 is the human image frame, figure 4 the recognition result of SIFT algorithm and figure 5 the recognition result of improved Mean shift algorithm.

![Figure 3. Human motion image sequence](image)

![Figure 4. SIFT algorithm recognition results](image)
From the simulation results, it is seen that the proposed Mean Shift algorithm has better convergence and higher accuracy in moving body behavior recognition.

5. CONCLUSIONS

The final purpose of human action recognition is to make robotic system learn and understand the behavior model in video information like human being. This technology involves in many subjects like the artificial intelligence, pattern recognition, computer vision etc. With the development of whole society, there is growing need of mainly intelligent video supervision and human action recognition are becoming a hot research area. In view of the defects of Mean shift algorithm in real situation, this paper proposed a moving human action recognition model based on Mean shift algorithm optimized by gradient features. The simulation results show that this algorithm has higher accuracy and better convergence performance than standard Mean shift algorithm.

REFERENCES