

Human Actions Recognition Using Improved MHI and 2-D Gabor Filter Based on Energy Blocks

Junfeng Sun, Hongji Xu*, Yingming Zhou, Lingling Pan, Feifei Li and Min Chen

School of Information Science and Engineering, Shandong University, Jinan, China

*Corresponding author

Abstract—The concept of motion-history image (MHI) is widely adopted by many researchers to solve problems of human actions recognition. An improved MHI with only one parameter is proposed in this paper, it is easier to be implemented and can retain more effective movement information compared with the original method. Furthermore, two-dimensional (2-D) Gabor feature based on energy blocks (EB-Gabor) is proposed to encode the texture information of MHI. The 2-D Gabor feature with high dimension is divided into multiple energy blocks and then the energy features of these energy blocks can be obtained. The energy features are served as the input of Adaboost. Experimental results on public benchmark KTH video database demonstrate the superiority of the proposed method over the state-of-the-art action recognition approaches.

Keywords—human actions recognition; motion-history image (MHI); 2-D Gabor feature; adaboost

I. INTRODUCTION

Human actions recognition has been widely applied in a number of real-world applications, such as video analysis, human-computer interaction and smart surveillance. Though significant progress has been made in past decades [1], recognizing actions is still a quite challenging task due to the inherent limitations of traditional data source, such as the variations in the lighting conditions, self-occlusions and cluttered backgrounds. In order to improve the accuracy of actions recognition, researchers have proposed some effective algorithms in motion segmentation and feature extraction, such as motion-history image (MHI) and two-dimensional (2-D) Gabor.

MHI [2] is the common method to describe how the motion is moving in the video sequence. It uses intensity to express the recent motion, where bright pixels represent the recent movement and dark pixels represent the past movement. Recently, a new approach to model and recognize actions was proposed in [3], it extracted 7 Hu moments from the MHI at each frame and used these features to encode the patterns captured from different actions. Murayama et al. [4] described the low-level features with the histogram of the oriented gradients of MHI, and calculated the similarity between the sequences by continuous dynamic programming for unusual human actions detection. Moreover, Ahad et al. [5] extended 2-D MHI to three-dimensional (3-D) and employed Hu moments to calculate the feature vectors. Most of above works have achieved a good recognition effect, but they all need to do a lot of parameters estimation before forming the MHI.

2-D Gabor [6] could extract the features of both time (spatial) and frequency domain. It is used in many applications, such as fingerprint recognition, image analysis and face detection, etc. For example, by using 2-D Gabor, Liu et al. [7] characterized the local appearance and shape on adaptive hierarchical depth motion maps (AH-DMMs) for human actions recognition. In [8], the 2-D Gabor filter was decomposed into one-dimensional (1-D) filters, and each filter was implemented on graphics processing units (GPUs). More recently, Kim et al. [9] proposed a novel approach to compute the 2-D Gabor by reducing the computational redundancy at multiple orientations and frequencies. However, the 2-D Gabor feature vector has a high dimension which can negatively influence the performance efficiency and occupy more memory resources.

To reduce the complexity of parameters estimation, an improved MHI is proposed in this paper. Firstly, we extract key frames from the original video, and then update the intensity value of these key frames, finally, combine the processed images to MHI. In addition, we propose a novel approach to reduce the dimension of 2-D Gabor by dividing the feature into multiple energy blocks and recoding these energy blocks. In this way, the texture information can be effectively acquired from MHI.

II. THE PROPOSED METHOD

A. Improved MHI

The MHI is a view-based method that shows motion trajectories from image brightness. MHI is widely used in the field of human actions recognition. The original MHI which is expressed as $H_\sigma(x, y, t)$ can be computed by

$$H_\sigma(x, y, t) = \begin{cases} \sigma & \text{if } Diff(x, y, t) \geq \xi \\ \max(0, H_\sigma(x, y, t-1) - \delta) & \text{otherwise} \end{cases} \quad (1)$$

$$Diff(x, y, t) = |I(x, y, t) - I(x, y, t \pm 1)| \quad (2)$$

where x , y and t denote the position and time, σ decides the temporal duration of MHI, δ is the decay parameter, and the threshold ξ is given. $I(x, y, t)$ is the intensity value of pixel location with coordinate (x, y) at the t^{th} frame of the

video. The original MHI has three parameters, which requires a large number of calculations to estimate the best combination of parameters to avoid the loss of useful information. To solve the problem, an improved MHI is proposed in this paper, which is expressed as

$$H(x, y, t) = \begin{cases} \bar{I}(x, y, t) & \text{if } H(x, y, t-1) \neq 0 \text{ and } \bar{I}(x, y, t) \neq 0 \\ H(x, y, t-1) + \bar{I}(x, y, t) & \text{otherwise} \end{cases} \quad (3)$$

where $\bar{I}(x, y, t)$ is a series of gray-value update images which can be calculated by

$$\bar{I}(x, y, t) = \begin{cases} \frac{255t}{N-1} + (255 \bmod (N-1)) & \text{if } Diff(x, y, t) \geq thd \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$Diff(x, y, t) = |I(x, y, t) - I(x, y, t \pm 1)| \quad (5)$$

where the threshold thd is given. N is the total number of frames in a video. Therefore, there is only one parameter thd in our calculation method, which greatly reduces the computational complexity. Figure 1 shows the process of parameters estimation. From Figure 1, we can see that the noise in the image is gradually disappearing with the threshold thd of improved MHI increases. However, if thd is too large, the movement foreground will disappear. So we can get that the optimal threshold is 45. In the same way, an optimal combination of parameters $\{\sigma = 250, \delta = 15, \xi = 40\}$ of original MHI can be obtained. As shown in Figure 1, the improved method can get a clearer trajectory with less computation of parameters estimation compared with the original method.

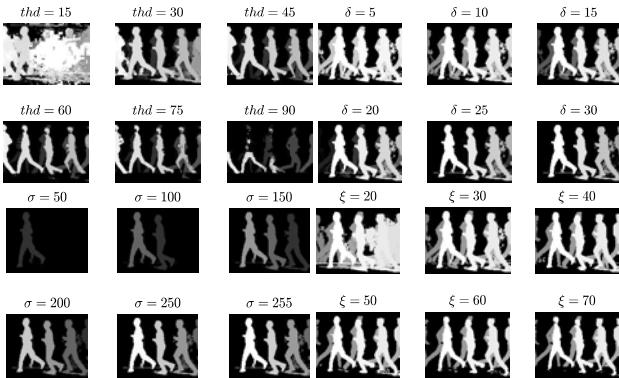


FIGURE I. THE PROCESS OF PARAMETERS ESTIMATION

Figure 2 shows an example of a complete calculation process of improved MHI: Firstly, segment the video and extract key frames. Next, calculate gray-value update images by (4) to remove the noise and get a clear movement foreground. Finally, calculate the MHI by (3).

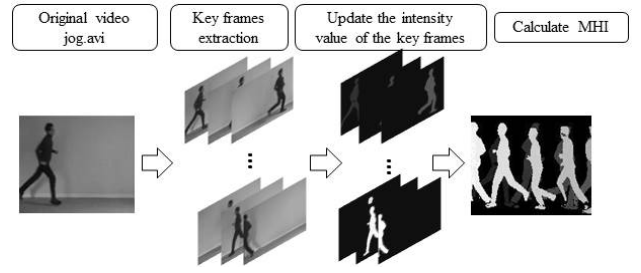


FIGURE II. THE CALCULATION PROCESS OF IMPROVED MHI

B. 2-D Gabor Based on Energy Blocks

2-D Gabor feature has good spatial locality and orientation selectivity, so it is widely used in image processing field. The basic function of 2-D Gabor filter can be computed by

$$\Psi_{\omega, \theta}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \exp[j\omega x'] \quad (6)$$

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta$$

where (x, y) is the pixel position in the spatial domain, θ is the orientation of Gabor filter, ω is the radial center frequency, and σ is the standard deviation of the Gaussian function along the x-axis and y-axis. For feature extraction of MHI, the images are convolved with the Gabor filter as

$$G_{\omega, \theta}(x, y) = MHI(x, y) * \Psi_{\omega, \theta}(x, y) \quad (7)$$

Equation (7) can be expressed as the sum of the real part $\text{Re}\{G_{\omega, \theta}(x, y)\}$ and the imaginary part $\text{Im}\{G_{\omega, \theta}(x, y)\}$ as

$$G_{\omega, \theta}(x, y) = \text{Re}\{G_{\omega, \theta}(x, y)\} + \text{Im}\{G_{\omega, \theta}(x, y)\} \quad (8)$$

$\text{Re}\{G_{\omega, \theta}(x, y)\}$ is used to characterize the local appearance and shape on MHI. In general, 40 filters [10] are used to calculate Gabor Real features.

Gabor Real feature vector with 40 filters is $120 \times 160 \times 40$ (the size of MHI is 120×160) which has a high dimension. Usually, principle component analysis (PCA) is adopted to reduce dimension, and this method is defined as Gabor-PCA. In this paper, a novel Gabor feature based on energy blocks is proposed which is defined as EB-Gabor. The EB-Gabor can be calculated by

$$EB_j = \sum_{k=(j-1)i+1}^{j \times i} |R_k|^2 \quad (9)$$

$$EB\text{-Gabor} = [EB_1, EB_2, \dots, EB_n]$$

Figure 3 shows the process of EB-Gabor segmentation. Figure 3 (a) expresses the N-dimensional EB-Gabor and it is evenly divided into n energy blocks as shown in Figure 3 (b).

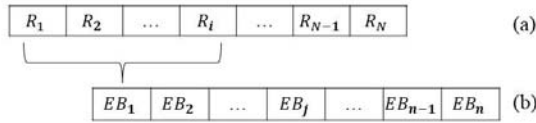


FIGURE III. EB-GABOR SEGMENTATION

After feature extraction, Adaboost [11] is adopted for classification which is widely used in human actions recognition. The kernel of Adaboost is to train different learning algorithms for the same training set, and then combine these algorithms to construct a stronger final learning algorithm.

III. EXPERIMENTS AND ANALYSIS

In this section, the proposed method is evaluated on KTH video database containing six types of human actions (walk, jog, run, box, hand wave and clap) performed several times by 25 subjects.

A. The Performance of Improved MHI

In order to show the performance of improved MHI, 45 videos of every action are chosen for training and the other 45 videos are chosen for testing. Firstly, we calculate the MHI with the original method ($\sigma = 250, \delta = 15, \xi = 40$) and the improved method ($thd = 45$) respectively, and then extract EB-Gabor and use Adaboost for classification.

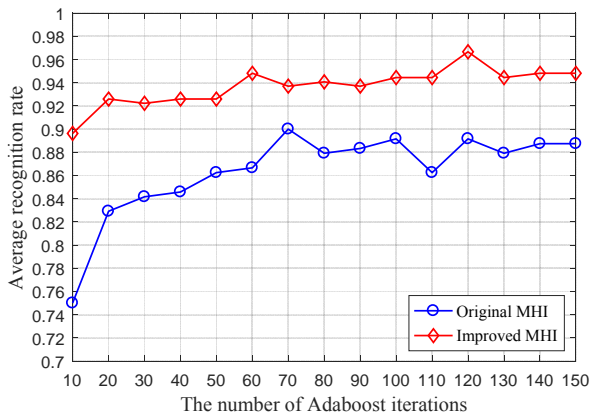


FIGURE IV. THE AVERAGE RECOGNITION RATE OF IMPROVED MHI AND ORIGINAL MHI

Figure 4 shows the average recognition rate of the improved MHI and the original MHI with the Adaboost of $\{10, 20, 30, \dots, 150\}$ times iterations. The recognition rate of the improved MHI has reached 96.5% with the Adaboost of 120 times iterations, which is 6.5% higher than that of the original MHI whose best recognition rate is 90%. What's more, the recognition result of the proposed method is always better than the original method in any Adaboost iterations, which fully states that the proposed method has higher validity and stronger robustness than the old method.

B. The Performance of EB-Gabor

In order to show the performance of the proposed EB-Gabor, Gabor-PCA is used for comparison in this paper. The energy reserved of original data after PCA algorithm is shown in Figure 5. From Figure 5, we can see that the higher the dimension of Gabor-PCA is, the more the energy reserved of original data is. Therefore, we adopt 260 dimensional Gabor-PCA which can save 99% energy.

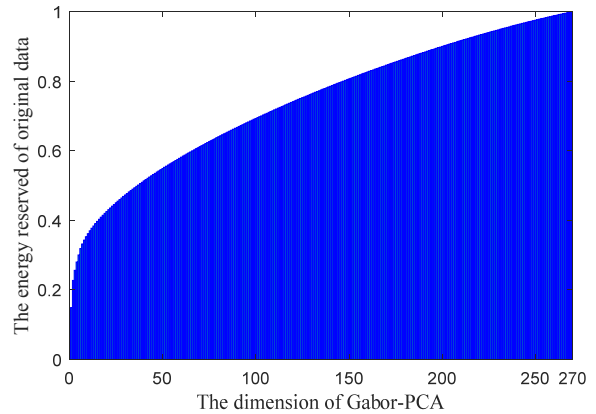


FIGURE V. THE ENERGY RESERVED OF ORIGINAL DATA

Figure 6 shows the average recognition rate of EB-Gabor and Gabor-PCA with the Adaboost of $\{10, 20, 30, \dots, 150\}$ times iterations. The recognition rate of EB-Gabor has reached 96.5% with the Adaboost of 120 times iterations, which is 8.6% higher than that of Gabor-PCA whose best recognition rate is 87.9%. It is obvious that the EB-Gabor always behaves better than Gabor-PCA in any Adaboost iterations. By comparing, we have confirmed the effectiveness of the proposed method. The confusion matrix is shown in Table 1. It can be seen that most of six actions are classified very well with an accuracy of around 97.8%.

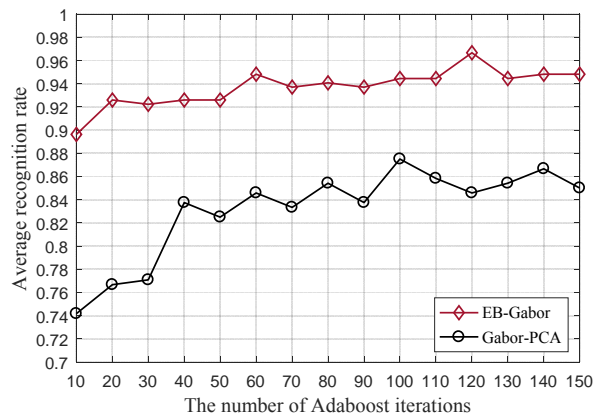


FIGURE VI. THE AVERAGE RECOGNITION RATE OF EB-GABOR AND GABOR-PCA

TABLE I. CONFUSION MATRIX OF OUR METHOD FOR KTH VIDEO DATABASE

Box	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Clap	0.0222	0.9778	0.0000	0.0000	0.0000	0.0000
Hand Wave	0.0222	0.0000	0.9778	0.0000	0.0000	0.0000
Jog	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
Run	0.0667	0.0000	0.0000	0.0000	0.9333	0.0000
Walk	0.0889	0.0000	0.0000	0.0000	0.0000	0.9111
	Box	Clap	Hand Wave	Jog	Run	Walk

C. Comparison with the State-of-the-arts

Furthermore, we compare the performance of our method with several state-of-the-art methods on the KTH video database and report the results in Table 2. It can be seen that our method outperforms the mentioned methods.

TABLE II. COMPARISON WITH THE STATE-OF-THE-ARTS ON KTH VIDEO DATABASE

Method	Year	Accuracy
Kamiński et al. [12]	2017	81.8%
Li et al. [13]	2017	91.6%
Zhao et al. [14]	2017	93.9%
Wang et al. [15]	2016	94.5%
Cho et al. [16]	2016	94.8%
Our method	2017	96.5%

IV. CONCLUSIONS

This paper presents an effective method for human action recognition using improved MHI and EB-Gabor. The improved MHI can capture more details of motion and shape clues by calculating gray-value update images. Furthermore, EB-Gabor is proposed to encode the texture information of MHI, it extracts energy features on the basis of 2-D Gabor. The experimental results on the benchmark KTH video database show that the proposed method outperforms the state-of-the-art approaches and achieves a highest accuracy of 96.5% with 6-class human actions.

ACKNOWLEDGMENT

This work was financially supported by the National Natural Science Foundation of China (61771292), the National Key Research and Development Program of China (2017YFC0803403), and the Natural Science Foundation of Shandong Province of China (ZR2016FM29).

REFERENCES

- [1] M. Liu, H. Liu and Q. Sun, "Action classification by exploring directional co-occurrence of weighted stips," Proceedings of IEEE International Conference on Image Processing, 2014, pp. 1460-1464.
- [2] H. Liu, M. Liu and Q. Sun, "Learning directional co-occurrence for human action classification," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2014, pp. 1235-1239.
- [3] E. C. Alp and H. Y. Keles, "Action recognition using MHI based Hu moments with HMMs," Proceedings of International Conference on Smart Technologies, 2017, pp. 212-216.
- [4] H. Murayama and K. Yamada, "Detection of unusual human activity based on sequence of actions with MHI and CDP," Proceedings of TENCON 2010 - 2010 IEEE Region 10 Conference, 2011, pp.1663-1667.
- [5] M. A. R. Ahad, T. Ogata, J. K. Tan, H. S. Kim and S. Ishikawa, "Comparative analysis between two view-based methods: MHI and DMHI," Proceedings of International Conference on Computer and Information Technology, 2008, pp. 1-6.
- [6] R. R. Isnanto, A. Hidayatno, A. A. Zahra, Eskanesiari, A. I. Bagaskara and R. Septiana, "Herb leaves recognition using combinations of Hu's moment variants — Backpropagation neural network and 2-D Gabor filter — Learning vector quantization (LVQ)," Proceedings of International Conference on Information Technology, Computer, and Electrical Engineering, 2017, pp. 219-224.
- [7] H. Liu, Q. He and M. Liu, "Human action recognition using adaptive hierarchical depth motion maps and Gabor filter," Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2017, pp. 1432-1436.
- [8] X. Wang and B. E. Shi, "GPU implementation of fast Gabor filters," Proceedings of IEEE International Symposium on Circuits and Systems, 2010, pp. 373-376.
- [9] J. Kim, S. Um and D. Min, "Fast 2D complex Gabor filter with kernel decomposition," IEEE Transactions on Image Processing, vol. 27, no. 4, 2018, pp. 1713-1722.
- [10] Q. Zhang, R. Wang, Y. Cai and X. Xu, "Null space local tangent space alignment discriminant mapping with feature fusion for face recognition," Journal of Computational Information Systems, vol. 8, no. 2, 2012, pp. 633-641.
- [11] K. B. Ge, J. Wen and B. Fang, "Adaboost algorithm based on MB-LBP features with skin color segmentation for face detection," Proceedings of International Conference on Wavelet Analysis and Pattern Recognition, 2011, pp. 40-43.
- [12] L. Kamiński, S. Maćkowiak and M. Domański, "Human activity recognition using standard descriptors of MPEG CDVS," Proceedings of IEEE International Conference on Multimedia & Expo Workshops, 2017, pp. 121-126.
- [13] J. Li, T. Wang, Y. Zhou, Z. Wang and H. Snoussi, "Using Gabor filter in 3D convolutional neural networks for human action recognition," Proceedings of Chinese Control Conference, 2017, pp. 11139-11144.
- [14] R. Zhao, Y. Zhao, R. Deng and F. Li, "Hierarchical random forest for senior action recognition in videos," Proceedings of International Conference on Wavelet Analysis and Pattern Recognition, 2017, pp. 171-176.
- [15] M. Wang, J. Sun and J. Yu, "Human action recognition based on feature level fusion and random projection," Proceedings of International Conference on Computer Science and Network Technology, 2016, pp. 767-770.
- [16] W. Cho, S. Kim and S. Park, "Human action recognition using variational Bayesian hidden Markov model with Gaussian-Wishart emission mixture model," Proceedings of International Conference on Machine Learning and Cybernetics, 2016, pp. 201-206.