

Human Action Recognition based on Human skeleton Hu invariant moments combined with human geometrical characteristics

Qing Ye^{1, a}, Xinran Guo^{2, b} and Yongmei Zhang^{3, c}

¹ College of Electronic Information Engineering, North China University of Technology, Beijing

No.5 Jinyuanzhuang Road, Shijingshan District, Beijing, 100144, China

² College of Electronic Information Engineering, North China University of Technology, Beijing

No.5 Jinyuanzhuang Road, Shijingshan District, Beijing, 100144, China

³ College of Computer, North China University of Technology, Beijing

No.5 Jinyuanzhuang Road, Shijingshan District, Beijing, 100144, China

^a yeqing@ncut.edu.cn, ^b guo_x_r@163.com, ^c zhangym@ncut.edu.cn

Keywords: Human action recognition, Moving object detection, Geometric characteristics of the human body, Human skeleton Hu invariant moments, K-nearest neighbor.

Abstract : Human action recognition is a highly active research topic in the field of video surveillance, human computer interaction and other fields. Due to the huge amount of computation, many existence methods fail in real-time applications. In this paper, we proposed a human action recognition method based on human skeleton Hu invariant moments combined with human geometrical characteristics. Firstly, foreground is extracted by the inter-frame difference and background subtraction. Secondly, human skeleton Hu invariant moments, the minimum bounding rectangle aspect ratio, rectangularity and circularity are calculated. Finally, human actions are recognized by the K-nearest neighbor. Experimental results in Weizmann action datasets show that this method can accurately identify human actions and has good real-time performance. The proposed method can be applied to real-time intelligent video surveillance with high accuracy.

Introduction

Human action recognition has been attracted much attention over the past few decades due to its applicability to many areas, including video surveillance, human computer interaction, etc. Many scholars have concentrated on the accuracy of human action recognition and contributed multiple proven methods. Nguyen et al. [1] build a human body model with shape template matching to detect the moving human. When complex background occurred, it may lead to identification errors. Min et al. [2] use an action region-aware dictionary to identify human actions. Lin et al. [3] propose an effective graph model based on Action Trait Code classification to recognize human actions. The previous methods are high recognition rates which cost complex algorithms and large amount of computations. Considering the running time and validity of the algorithm, we research a human action recognition method based on human skeleton Hu invariant moments combined with human geometrical characteristics. Inter-frame difference and background subtraction are combined to detect moving objects. Feature extraction based on human skeleton Hu invariant moments and human geometrical characteristics. K-nearest neighbor classification is used to identify human actions. This method is real-time and fast in execution time. It can be used in real-time alarm.

Implementation of Human Action Recognition

In this paper, we use human action recognition method which based on human skeleton Hu invariant moments combined with human geometrical characteristics. This method includes moving object detection, human action feature extraction and human action recognition. The diagram of human action recognition algorithm is shown in Fig.1 Firstly, Gaussian filter and the expansion of corrosion closing operation are used to remove the noise of video sequence. Secondly, a combined method which includes inter-frame difference, background subtraction and background updating model [4] is

used to detect moving objects. Then, the body contour of the minimum bounding rectangle aspect ratio, rectangularity and circularity are calculated. Human skeleton Hu invariant moments are calculated when the improved extraction Zhang-Suen thinning algorithm [5] is used to extract human skeleton. Finally, K-nearest neighbor classification is used to action identification.

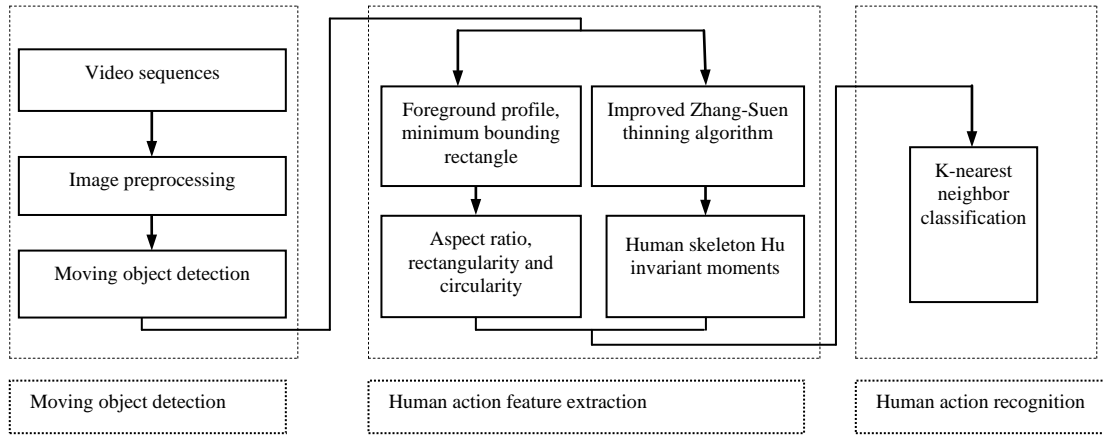


FIG.1 Algorithm Block Diagram

Moving object detection

In moving object detection section, firstly, image preprocessing is executed. Then, the combined method which includes inter-frame difference, background subtraction and background updating model is used to detect moving object.

Image preprocessing

Before moving object detection, Gaussian filter and the expansion of corrosion and closing operation are used to remove noise image so that the image becomes clear, easy to follow-up treatment.

Inter-frame difference and background subtraction

In this paper, inter-frame difference, background subtraction and background update model are used to extract the moving object.

(1) Inter-frame difference method

This method is used to utilize difference between the adjacent frames, then the set threshold is used to extract change areas. Expression of mathematical formula is shown in Eq.1.

$$D_t(x, y) = |I_t(x, y) - I_{t-1}(x, y)| \quad R_t(x, y) = \begin{cases} 1 & D_t(x, y) > T \\ 0 & D_t(x, y) < T \end{cases} \quad (1)$$

Where $I_t(x, y)$ and $I_{t-1}(x, y)$ denote the pixel values in the image at time t and $t-1$ time (x, y) point, $R_t(x, y)$ is the foreground object image (1 as the foreground and 0 as the background), T is the threshold which set by Otsu method.

(2) Background subtraction

The subtraction method is used to calculate absolute difference between the current frame and background image, then motion areas are detected. Expression of mathematical formula is shown in n Eq.2.

$$D(x, y) = |I(x, y) - I_{bk}(x, y)| \quad R(x, y) = \begin{cases} 1 & D(x, y) > T \\ 0 & D(x, y) < T \end{cases} \quad (2)$$

Where $I(x, y)$ and $I_{bk}(x, y)$ denote the pixel values in the image of the current frame and the background image (x, y) point, $R_t(x, y)$ is the foreground object image (1 as the foreground and 0 as the background), T is the threshold which set by Otsu method.

(3) Background update model

Background pixels within the moving region remain unchanged, while the background pixels of the non-moving area are updated with the current background image, which can get a better real-time background. Expression of mathematical formula is shown in Eq.3.

$$I_{bk}(x, y) = \begin{cases} (1-\alpha)I_{bk-1}(x, y) + \alpha I_{bk}(x, y) & \text{if } R(x, y) = 0 \\ I_{bk}(x, y) & \text{if } R(x, y) = 1 \end{cases} \quad (3)$$

Where α denotes the update factor.

The original image is shown in Fig.2(a), when we use background update model ($\alpha = 0.05$), the binary foreground image is showed in Fig.2(c), the update background is shown in Fig.2(c). We can get a clear binary foreground image.



(a) The original image



(b) Update background ($\alpha = 0.05$)



(c) Binary foreground image

FIG.2 Moving object detection

Human action feature extraction

In feature extraction section, because of the differences in the human body, the only body geometric features may affects motion recognition, so human skeleton HU invariant moments combined with human geometric characteristics are used to extract features.

Geometric characteristics of the body include the aspect ratio of the minimum bounding rectangle, rectangularity and circularity. The minimum bounding rectangle high aspect ratio refers to the ratio of the width and height of the minimum circumscribed rectangle. Rectangularity is the ratio between objects area and its minimum bounding rectangle area. Circularity is a ratio which between the foreground area and the square of the foreground outline perimeter.

In the binary image, m_{00} normalized first moment m_{01} and m_{10} are used to get the coordinates of the center of gravity of the object $I_c(x_c, y_c)$. m_{00} is shown in n Eq.4, m_{01} is shown in n Eq.5 and m_{10} is shown in Eq.6.

$$m_{00} = \sum_{i=1}^n \sum_{j=1}^m I(i, j) \quad (4)$$

$$m_{01} = \sum_{i=1}^n \sum_{j=1}^m jI(i, j) \quad (5)$$

$$m_{10} = \sum_{i=1}^n \sum_{j=1}^m iI(i, j) \quad (6)$$

The coordinates of the center of gravity is shown in Eq.7.

$$x_c = \frac{m_{10}}{m_{00}} \quad y_c = \frac{m_{01}}{m_{00}} \quad (7)$$

Finally, the second and third normalized central moment are used to get the image seven Hu moment invariants. The Hu moment invariants is shown in Eq.8.

$$M_1 = \eta_{20} + \eta_{02}$$

$$M_2 = (\eta_{20} + \eta_{02})^2 + 4\eta_{11}^2$$

$$M_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$M_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$M_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$M_6 = (\eta_{20} - \eta_{02})(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$M_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \quad (8)$$

Human action recognition

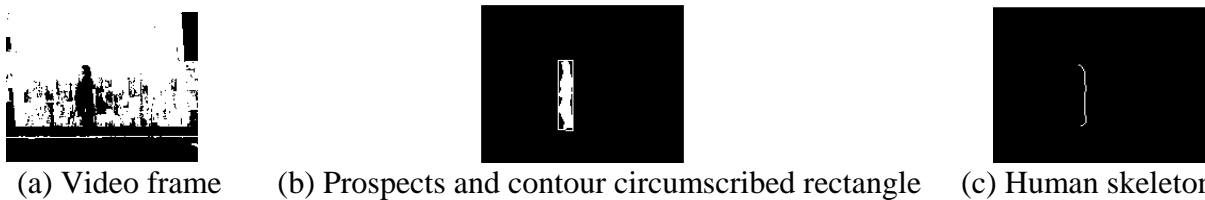
In human action recognition part, K-nearest neighbor classification is used to identify human action. K-nearest neighbor classification algorithm based on learning by analogy, which can achieve higher classification accuracy, robustness, clear concepts. The basic idea of K-nearest neighbor algorithm is that if a sample in the feature space of k most similar, the majority of the sample belongs to a category, then the sample also fall into this category.

We select the action of two people in Weizmann library as training videos, some key frames are extracted from the video sequences. The corresponding human skeleton HU invariant moments and geometric characteristics are calculated. These data are created as a training data set. When a test video is read, the human body motion characteristics are extracted from the video, then, these characteristics are compared with the training data set. Finally, human actions are recognized.

Experimental results and analysis

We complete the recognition algorithm of human action using visual studio 2010 development tools and OpenCV computer vision class open source library, and it is tested in the Weizmann library.

Human motion foreground is shown in Fig.3(a), the minimum bounding rectangle is shown in Fig.3(b) and human skeleton is shown in Fig.3(c). The minimum bounding rectangle of aspect ratio, rectangularity and circularity are calculated by Fig.3(b). The human skeleton Hu invariant moments are calculated by Fig.3(c). Action recognition results are showed in Table1.



(a) Video frame (b) Prospects and contour circumscribed rectangle (c) Human skeleton

FIG.3 The extraction of circumscribed rectangular and human skeleton

Table1. Human action recognition

Result Sample	Recogniti on rate	Error rate			
jump	0.91	0.03(run)	0.01(skip)	0(side)	0.05(walk)
run	0.82	0.09(jump)	0.05(skip)	0.01(side)	0.03(walk)
skip	0.89	0.01(jump)	0.03(run)	0.02(side)	0.05(walk)
side	0.86	0.02(jump)	0.03(run)	0.03(skip)	0.04(walk)
walk	0.94	0(jump)	0.02(run)	0.01(skip)	0.03(side)
average	0.884	—	—	—	—

We select two person action video sequences in Weizmann action datasets, their characteristics are extracted. These data are created as a training data set by K-nearest neighbor classification algorithm.

Then, seven other person as the test data set, their characteristics are extracted and it is compared with the training data set. Finally, human actions are recognized. “walk”、“run”、“jump”、“skip” and “side” in the Weizmann library are tested.

This method of action recognition accuracy rate on average is 88.4%. Actions of “jump” and “walk” recognition rates are 91% and 94% respectively. They are high recognition rates. The action of “Run” recognition rate is 82%. The reason why recognition rate low is that human behavior can be divided into a series of continuous action gesture. Similar posture may cause the error rate.

Conclusion

In this paper, we study an identification method based on human skeleton HU invariant moments combined with human prospects geometrical characteristics is proposed. Inter-frame difference and background subtraction are combined to detect moving objects. In feature extraction section, we proposed human skeleton HU invariant moments combined with human geometric characteristics. These features can reduce the impact of differences in human body, while the method is simple and better real-time. K-nearest neighbor classification is used to action identification. The experimental results show the detection accuracy and execution time can meet the requirements of monitoring system. Intelligent monitoring can be applied to improve the level of intelligence. Future studies will focus on improving the robustness of the detection algorithm in complex

Acknowledgements

This paper is supported by Excellent Talent Training Project of Beijing (2013D005002000002), National Natural Science Foundation of China (Grant No.61371143), Natural Science Foundation of Beijing (Grant No.4132026), National Science and Technology Support Program of China (2012BAH04F00), Technology Project of Beijing Municipal Education Commission (KM201210009007), Excellent Young Teachers Training Project by North China University of Technology.

References

- [1] D T Nguyen, P Ogunbona, W Q Li. Human detection with contour-based local motion binary patterns. IEEE 8th International Conference on Image Processing, 2011, 3609-3612
- [2] Hyun-seok Min, W.De Neve and Yong Man Ro, Sparse Representation-Based Human Action Recognition Using an Action Region-Aware Dictionary, 2013 IEEE International Symposium on Multimedia (ISM), pp.133-139, 2013.
- [3] S Y Lin, C K Shie, S C Chen. Human Action Recognition Using Action Trait Code. International Conference on Pattern Recognition (ICPR). 2012, 3456-3459
- [4] Raharnan, M M Hasan, S M N Hossain. An efficient background updating model for motion detection. 2013 International Conference on Informatics, Electronics & Vision (ICIEV). 2013, 1-5
- [5] Q Ye, J F Dong, Y M Zhang. Real-Time Human Skeleton Extraction Based On Video Sequences. Applied Mechanics and Materials.vol.401-403, 2013, 1410-1414