Journal of Artificial Intelligence Practice (2016) 1: 14-19
Clausius Scientific Press, Canada

Body Pose Estimation Based on Half - body Mixed Model

Jiuhua Tao[1,a], Xinhua Wu[1,b] and Gang Liu[1,c]

[1]Computer Science Technology, Wuhan University of Technology, Wuhan, Hubei 430063, China
[a]744401856@qq.com, [b]444083710@qq.com,[c]liu_gang@whut.edu.cn

Keywords: Human pose estimation; Object detection; HOG feature extraction.

Abstract: In order to improve the effect and speed of human pose estimation from the static image, this paper proposes a method based on the prior knowledge of HOG eigenvalue and face detection to establish the human body bust mixed model for human pose estimation. First, assume that the bust human model contains K components, the static image is divided into M * N cells, each cell may be one of the components, according to the fractional calculation formula to calculate the root component scores, and ultimately determine the human body. The bodily mixed model can be used to calculate the position and direction of human limb accurately.

1. Introduction

Human pose estimation is to detect the position distribution of each limb from the static image and calculate its direction and scale information. The results of general pose estimation are divided into two-dimensional and three-dimensional, The method of attitude estimation is divided into model-based and model-free. In this paper, the human body feature extraction, model training, target detection and attitude estimation are analyzed and discussed in detail, and the relative theoretical methods are compared and analyzed from the practical application.

Recognition of human behavior requires first to detect the position of the body part from the image and calculate its direction and scale information, that is, to achieve human pose estimation. The human pose in image and video sequence has always been a key problem in computer vision research. It is the basis of behavioral understanding, behavior monitoring and human-computer interaction, and has wide application prospect.

The methods of human pose estimation are mainly model-based[1,2,3] and model-free[4,5,6]. Model-free methods often require more training samples, more training time and more stable training algorithms, so the current mainstream model is based on the attitude estimation method. In this paper, the hybrid model is used to estimate the posture of the upper body.

Aiming at the problem of slow detection[7,8] of human body, this paper proposes a bodily limb model to replace the original model to optimize the detection process. The experimental results can greatly improve the detection speed without loss of detection accuracy.

2. Human Feature Extraction

Because HOG features can show strong robustness under illumination, morphological changes

and other complex environments, this paper chooses HOG feature as the feature of describing human body target. After extracting the human body feature information, LSVM[9,10,11] is used to classify it. When the deformable part model is established, different weights of different parts are set according to the different contribution of the different regions of human body to the detection effect. The components with larger response score are more important to the testing process. Based on previous research, this design enriches the marker information in the training samples and improves the detection model to improve the detection performance.

Detecting the human body from the image needs to match the component model with the image. The simple method is to use image color information to image segmentation and then to match with the model. But this method is ineffective, because the human dress and lighting, background changes, resulting in more difficult to get the image segmentation threshold. Therefore, this paper chooses the HOG feature which obtains the static image according to the gradient direction so as to better solve the problem.

To compute the HOG feature, first calculate the gradient as follows:

$$G_x(x,y) = H(x+1,y) - H(x-1,y) \quad G_y(x,y) = H(x,y+1) - H(x,y-1) \tag{1}$$

H (x, y) represents the gray value of the image at pixel (x, y), Where $G_x(x,y)$ and $G_y(x,y)$ represent the horizontal and vertical gradients at (x, y), respectively. The gradient size and gradient direction at the pixel (x, y) in the image are:

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \qquad \alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{2}$$

3. Half - body hybrid part model

Simple use of rectangular or oval parts that represent the body parts, often require location, orientation, scale and other information in order to accurately represent the status of a component, correct estimation of the human body posture is faced with many problems such as many parameters and large computation. The use of the Pictorial model lacks prior knowledge of the appearance of the human body, and therefore requires the use of hybrid models to combine the two to exploit their advantages.

Component Representation. The hybrid model still uses components to make up the human body, but instead of using rectangles or ellipses to represent the human part more accurately, the alternative is to use only square components. The rectangular or elliptical representation indicates the direction in which the component is estimated when the component is estimated, and the square representation uses only four possible component states to replace the orientation, as shown in Figure 1. At the same time the component state does not need scale information, so only a few parameters are needed to describe the human attitude.

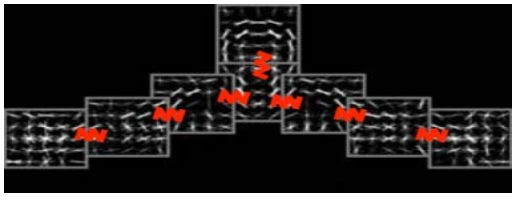

Figure 1. Body bust mixed model

In this paper, a hybrid model of human body is proposed for target detection. The method divides

the human body into several interrelated parts, uses the graph model to model the human body, and utilizes the graph reasoning method to optimize the human posture. Experimental results show that the model in the human body detection can also achieve good results, the following focuses on the human body blending model.

4. Experimental procedure

The human body half-body hybrid model consists of two layers of filters: a root filter and part filters. The root filter is used to capture the overall contour feature of the object, while the component filter is used to capture the details of the features in the target, such as eyes, nose, and mouth. The filter is actually a rectangular template, and each element of the matrix is a d-dimensional weight vector.

A model for detecting a bust of a human body having n component targets can be described as a (n + 2) tuple $(F_0, P_1, \cdots, P_n, \mathrm{b})$. Where F_0 is the root filter, P_{i} is the model of the i-th component, and b is the prior probability. Each part model can be defined as a triple $(F_i, \mathrm{v}_{\mathrm{i}}, \mathrm{d}_{\mathrm{i}})$, F_{i} is a local filter of P_{i}, v_i is a two-dimensional vector that records the relative position of P_{i} and the root part, and d_i is a four-dimensional vector representing the deformation cost coefficient of each possible position of P_{i}.

The HOG feature extracted in the image can be represented as a feature map G. A feature bitmap is a two-dimensional matrix in which each element is a d-dimensional vector that corresponds to a cell in the image. The upper left corner of the filter F is placed at the coordinate (x, y) of the feature bitmap G, and the two rectangles are calculated as "dot product". The score is defined as:

$$\sum_{x',y'} F[x', y'] \cdot G[x + x', y + y'] \tag{3}$$

The higher the score, the greater the likelihood that a cell becomes the component represented by the filter.

When making an assumption about an object, we need to specify the position of each filter in the feature pyramid in the model, $\mathrm{z} = (p_0, \cdots, p_n)$, Where $\mathrm{p}_i = (x_i, y_i, l_i)$ indicates the location and hierarchy of the i-th filter. A hypothetical score is composed of three parts: 1. the fraction of all the component filters in their respective positions; 2. the distortion of the position of each component relative to the root; and 3. the prior probability. Using the mathematical formula:

$$\mathrm{score}(p_0, \cdots, p_n) = \sum_{i=0}^{n} F_i' \cdot \phi(H, p_i) - \sum_{i=1}^{n} d_i \cdot \phi_d(dx_i, dy_i) + b \tag{4}$$

Where $(dx_i, dy_i) = (x_i, y_i) - (2(x_0, y_0) + v_i)$ gives the offset position of the i-th component and $\phi_d(dx, dy) = (dx, dy, dx^2, dy^2)$ is the deformation feature.

In order to detect an object in the image, a global score is calculated from the position of the component, that is:

$$\mathrm{scope}(p_0) = \max_{p_1, \cdots, p_n} scope(p_1, \cdots, p_n) \tag{5}$$

The larger the $\mathrm{scope}(p_0)$, the higher the probability of an object in the image.

In the training process of the deformable part model, a hidden support vector machine (LSVM) designed for weak supervised learning is introduced. Considering a classifier, the score for each sample in LSVM can be expressed in the following form:

$$f_{\beta}(x) = \max_{seZ(x)} \beta \cdot \Phi(x, z) \tag{6}$$

Where β is a vector of model parameters, z is invisible variables, set Z (x) defines the possible hidden variable value sample x, that is the position of each member. Similar to the classical SVM algorithm, we need to train β using training set $D = (< x_1, y_1 >, \cdots, < x_n, y_n >)$, where -1 <y <1, training The procedure is to minimize the objective function:

$$L_D(\beta) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n} \max(0, 1 - y_i f_{\beta}(x_i)) \tag{7}$$

Where $\max(0, 1 - y_i f_{\beta}(x_i))$ is the standard connection loss and C controls the weight of the regularization term. If there is only one possible hidden variable per sample x_i, $|Z(x_i)| = 1$, then f_{β} is linearly related to β. Linear SVM can be considered as a special case of hidden SVM.

In order to train the object detection model, a large number of negative samples are often required. But it is unrealistic to use all negative samples at the same time. Generally consider training data include positive samples and "negative" negative samples. Bootstrapping can be used to train an initial model based on the initial subset of negative samples and subsequently collect negative samples that can not be correctly classified in the initial model to form a "difficult" negative set of samples. A new model is then trained using the negative negative samples. This process is repeated several times until the model is trained.

After reasoning, distance translation, and message passing, we need to go back, and if the body gesture reasoning algorithm determines that a cell is a human model root, it needs to look back to find other parts of the body. Depending on the number of rows, columns, and types of cells in which the root part resides, you can know the cell location and part type in which the message is passed in the child, which is the child of the root part. Subsequent searches for the number of rows, columns, and types in the subassembly reveal that the next subcomponent can be repeated by repeating the process to find the final model leaf node.

All of the above components together is a complete human body, according to parts of the body part of the information, location and direction and scale.

5. Experimental conclusions

Figure 2 is a simulation of driver behavior of the half-body posture estimation results, we can see from the figure, the use of this model based on the human body half-body hybrid model can be completed body bodily gesture estimation.

Figure 2. Estimation of body half - body posture

Figure 3 shows that this method can be applied to the INRIA human database

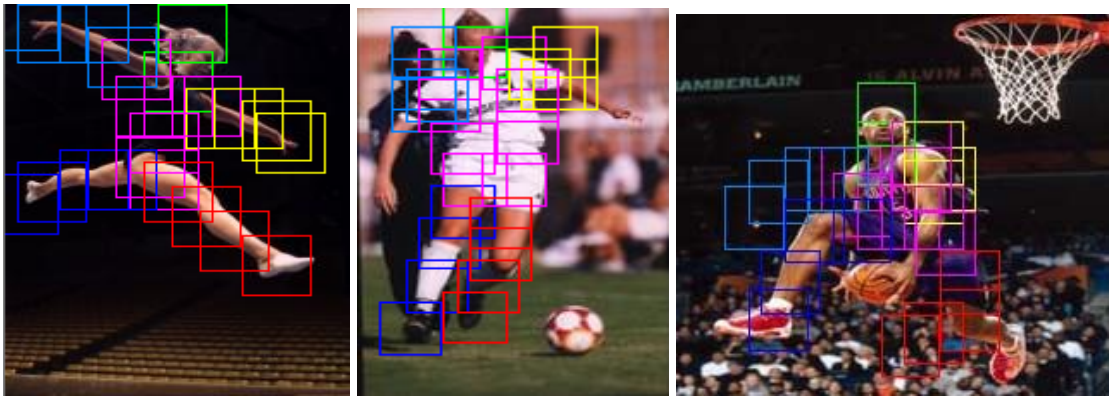

Figure 3. Results of attitude estimation in INRIA dataset

In this paper, the hybrid model is used to estimate the attitude of the driver. First of all, the driver's body is modeled, and the body parts are represented by a simple Part model. Spring links are used between the components to determine the human body based on the root component scores. The detailed steps of estimating the human pose principle by using the hybrid model, calculating HOG features, distance conversion, message passing, backtracking and non-maximum suppression are introduced in detail. Experiments show that the hybrid model can be used to calculate the position and direction of the driver 's upper limb accurately.

Acknowledgements

National Natural Science Foundation of China (51179146, China)

References

[1] V. Ferrari, M. J. Marln-Jimnez, A. Zisserman. Pose Search: Retrieving People Using Their Pose. Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, 1-8, 2009.

[2] E. Marcin, F. Vittorio. Better Appearance Models for Pictorial Structure. Proceedings of British Machine Vision Conference, 2009.

[3] B. Daubney, D. Gibson, N. Campbell. Monocular 3D Human Pose Estimation Using Sparse Motion Features. Proceedings of International Conference on Computer Vision, 1050-1057,2009.

[4] Yi Wang, Gang Qian. Robust Human Pose Recognition Using Unabled Markers. Proceedings of Workshop on Applications of Computer Vision, 1-7, 2008.

[5] P. Kohli, J. Rihan, M. Bray. Simultaneous Segmentation and Pose Estimation of Humans using Dynamic Graph Cuts. International Journals of Computer Vision, 79(3):285-298, 2008.

[6] R. Okada, S. Soatto. Relevant Feature Selection for Human Pose Estimation and Localization in Cluttered Images. Proceedings of European Conference on Computer Vision, 434-445,2008.

[7] nlpr-web.ia.ac.cn/course/object-recognition.pdf

[8] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake. Real-time Human Pose Recognition in Parts from Single Depth Images. Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, 1297-1304, 2011.

[9] P. Felzenszwalb, R. Girshick, D. McAllester, D. Ramanan. Object Detection with Discriminatively Trained Part Based Models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(9): 1627-1645, 2010.

[10] P. Felzenszwalb, D. McAllester, D. Ramanan. A Discriminatively Trained, Multiscale, Deformable Part Model. Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, 1-8, 2008.119

[11] P. Felzenszwalb, D. P. Huttenlocher. Distance Transforms of Sampled Functions. Theory of Computing, Vol. 8, 415-428, 2012.