ATLANTIS PRESS

Advances in Intelligent Systems Research, volume 133

2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE2016)

An Approach for Detecting Human Posture by Using Depth Image

Xianshan Li, Maoyuan Sun and Xiuxiu Fang

College of Information Science and Engineering, Yanshan University, Qinhuangdao, HeBei Province, 066004, China

***Abstract*—This paper introduces a method that can detect human posture by using depth image. The method uses head model to locate the human position which includes edge extraction, template matching and human detection. Then we extract the HOG feature from the depth images to get the characteristic vector of the original image. At last, a generalized regression neural network is processed to classify and identify the human posture. Experiments show that our method is able to identify the human posture from a depth image with a satisfactory recognition rate.**

Keywords-kinect; depth image; human posture; regional growth

I. INTRODUCTION

The advent of low-cost video devices has led to great interest in service robot that takes advantage of depth information gotten from Kinect about the service object. Our approach is based on depth image from data collected by Kinect.

Conventional approach of human posture recognition is mainly based on RGB image. It is difficult to extract the foreground from complex background by the method for light and environment affected the recognition results very seriously.

Human posture recognition is mainly based on two kinds of methods, one is model-free, the other is model-based that is the main method in recent years. Ferrari [1], in the presence of occlusion problem, presents a solution to the upper body two-dimensional pose recognition, and has achieved some good results. On the basis of Ferrari, by adding more constraints, Marcin[2] simplify the requirements for initialization to further improve the accuracy of the identification. Jiang[3] et al. coverage the segmentation of the target image with a rectangular to the human body parts to transform the pose identification problem into a rectangular coverage area and the original image segmentation consistency problem.

Comparing with traditional methods, the depth image based posture recognition add the skeleton information to extracting the local feature from original image in recent years. D.Grest [4] firstly definite the skeleton model and the starting position, then use the ICP iterative algorithm to track the human. R.Girshic[5] proposed an offset regression algorithm, which can solve the problem of estimating the position of the body part connection point due to occlusion or sensor limitation. Keskin C [6] divide the 3D hand model into 21 bone connection points by using real-time 3D skeleton fitting algorithm, and then train the hand model through the random forest. Deng Rui[7] convert depth map to 3D point cloud, filter the depth information, then train the model by using support vector machine (SVM). Mao Ye [8] process the depth map of the three-dimensional point cloud through the depth map and the motion graph matching extract previously and ultimately improve the accuracy.

The method use an Ω model to match the head that is represented as the person's location. Then the Hog character is extracted from the region contains the person that is portioned by region growth algorithm, and the GRNN training method is efficiently performed.

II. HEAD MATCHING AND CHARACTER DETECTION

There are holes in the depth image output from Kinect, that is, the depth of the part of the value is 0. The first step is using nearest neighbor interpolation algorithm that choose the nearest neighbor distance value of four adjacent points. The second step is using the median filtering algorithm to filter and smooth data.

A. *Edge Extraction*

The valuable information contains in edge of the image can be used for image analysis and target recognition. There are many kinds of edge detection methods, such as Robert operator, Sobel operator, Prewitt operator, Laplacian operator, Canny operator and so on. In this paper, Canny operator is used to extract edge by setting threshold which can remove partial non coherent edge information. The image of edge extraction is shown in FIGURE I.

Due to the holes in the edge information which will cause great influence to the following figures detection, we repair the edge extracted from depth image. We search the detected edge, and assign the surrounding value to 1 if the pixel is 1. By this step, a more rough edge is obtain for the edge of the original 1 points is changed to the edge of an unit 9 points. The discontinuous edges are changed into continuous edges as is shown in FIGURE II.

B. *Template Matching*

Image matching technology is based on a searching process, that is, searching a similar module of a known image module in another image. Several classical templates matching method includes AD, MAD, SD, MSD, Prod and so on. For the great noise in the depth map got from kinect, MAD algorithm is used to matching the head.

FIGURE I. EDGE EXTRACTION

FIGURE II. EDGE REPAIR

We denote $S(i, j)$ as the original image covered by templates which size is $N_x \times N_y$ as is shown in figure 3(a), $\mathrm{T}(i, j)$ as the templates image which size is $M_x \times M_y$ （$N > M$） as is shown in the following figure 3(b). The sub graph is search graph covered by the template; the reference point is the pixel in the upper left corner of sub graph which coordinates is i, j with the range of $1 < i, j < N - M + 1$.

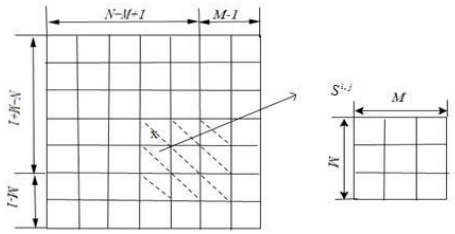


FIGURE III. SCHEMATIC DIAGRAM OF TEMPLATE MATCHING

It is shown in FIGURE III that if T and $S^{i,j}$ are agreement, that is the difference between them is zero, the similarity between T and $S^{i,j}$ can be measured by the formula(1).

$$D(\mathrm{i}, \mathrm{j}) = \sum_{m=1}^{M} \sum_{n=1}^{M} [S^{i,j}(m,n) - T(m,n)]^2 \qquad (1)$$

The formula for calculating the correlation of MAD algorithm is:

$$D(x, y) = \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} | S(\mathrm{i}+\mathrm{x}, \mathrm{j}+y) - T(\mathrm{i}, \mathrm{j}) | \qquad (2)$$

The range of the reference image registration is $(N - M + 1)^2$. If the overlap part of reference image $T(\mathrm{i}, \mathrm{j})$ and the image $S(i, j)$ is zero, the result of (2) must be zero. According to the formula (2), the value $D(x, y)$ of the points within the matching scope in the $S(i, j)$ can confirm the matching points of $T(\mathrm{i}, \mathrm{j})$ and $S(i, j)$ that is determined as the zero or minimum point. The head model and the final positioning effect of the head selected in our works are shown in FIGURE IV.

C. Human Detection

If the human foot and the floor in the same depth, it is very difficult to extract the contour of the entire body from the depth image using conventional edge detectors. And likewise, if person and some objects are stuck together at the same depth, the detection is also very difficult. Because the person's foot is usually straight and erect, we use response filter, denoted as $F = [1,1,1,-1,-1,-1]^T$, to extract the contour of the human foot and the floor.

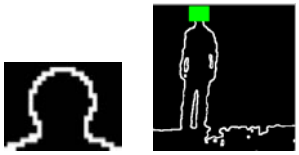


FIGURE IV. HEAD MATCHING

Then we extract the entire human character by the regional growth algorithm. Assuming a depth of the body surface value is continuous and in a certain range, we set the position of the head located in the template matching step as the seed region, and define the similarity between x and y as:

$$\mathrm{S}(x, y) = | \mathrm{depth}(x) - \mathrm{depth}(y) | \qquad (3)$$

where S is the similarity, $\mathrm{depth}()$ is the depth of pixels that is the average of the pixels' depths in a region defined as:

$$\mathrm{depth}(\mathrm{R}) = \frac{1}{N} \sum_{i \in R} (\mathrm{depth}(i)) \qquad (4)$$

The regional growth algorithm based on depth information is as follow:

Algorithm : Regional Growth based on Depth Information

Input: Depth image I1

Output: Depth image I2 containing only human characters

BEGIN

1. initialize: region=seed

2. Calculate the average depth of head region by using the formula(4)

3. Find adjacent pixels in a region

3.Calculate the similarity of the pixels in the region by using the formula (3)

4. if (S < threshold) {

5. add the pixel with the highest similarity of the region

6. calculate the new average depth of the region }

7. else

8. goto(3)

END

The threshold in the algorithm is set as 0.5 which is selected according to the experimental results, for the result is not

satisfactory if the threshold is too small and the other regions may be overwritten causing by overgrowth if it is too large. The seed selection and extraction result are shown in FIGURE V and FIGURE VI.

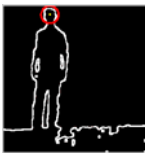


FIGURE V. SEED SELECTION POINT

FIGURE VI. EXTRACTION RESULT

III. FEATURE EXTRACTION

HOG is a feature descriptor for object detection in the field of computer vision and image processing. The mainly application of this method is in the static image of pedestrian detection, later, this method is also used in the pedestrian detection, vehicles and animal detection in videos.

HOG has many similarities with edge orientation histogram, scale invariant feature transform and shape context method, but HOG is calculated in a grid of dense and uniform size unit.

In this paper, HOG feature is used to extract the feature of depth map because of the independence of the illumination and depth image.

As usual using methods, we firstly calculate the gradient, then construct gradient direction histogram, at last, connect all the features of the block to get the characteristic vector of the original image. Detailed practices can refer to the literature [9].

IV. CLASSIFICATION METHOD

GRNN (Generalized Regression Neural Network) was first proposed by Specht that was a branch of RBF neural network. GRNN has the advantages of approximation ability, classification ability and learning speed, and has better effect when the sample size is small or the data is not stable. We use the GRNN to classify and identify the human posture. The structure of GRNN is as follows.

Determining the radial basis function center of the hidden layer neuron is the first step of GRNN learning algorithm. Set the training set sample input matrix P and the output matrix T respectively as:

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1Q} \\ p_{21} & p_{22} & \cdots & p_{2Q} \\ \vdots & \vdots & & \vdots \\ p_{R1} & p_{R2} & \cdots & p_{RQ} \end{bmatrix},\quad \boldsymbol{T}=\begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1Q} \\ t_{21} & t_{22} & \cdots & t_{2Q} \\ \vdots & \vdots & & \vdots \\ t_{S1} & t_{S2} & \cdots & t_{SQ} \end{bmatrix} \quad (5)$$

where $p_{i,j}$ is the *i*th input variables of the *j*th training samples, $t_{i,j}$ is the *i*th output variables of the *j*th training samples, R is the dimension of input variables, Q is the number of training set samples.

A training sample corresponds to a neuron. Determining the threshold of the hidden layer neurons is the second step of GRNN. The corresponding threshold value of Q neurons can be express as:

$$\boldsymbol{b}_1=\begin{bmatrix} b_{11} & ,b_{12} & ,\cdots, & b_{1Q} \end{bmatrix}' \quad (6)$$

where $b_{11}=b_{12}=\cdots=b_{1Q}=\dfrac{0.8326}{\text{spread}}$, spread is the expansion velocity of radial basis function.

Then we determine the weights between the hidden layer and the output layer by:

$$\boldsymbol{a}^i=\exp(-\|\boldsymbol{C}-\boldsymbol{p}_i\|^2\boldsymbol{b}_1),\quad i=1,2,\cdots,\mathrm{Q} \quad (7)$$

where, $\boldsymbol{p}_i=[p_{i1},p_{i2},\cdots,p_{iR}]'$ is the *i*th training sample vector. The connection weights of hidden layer and output layer can be set to the training sample matrix, that is $\boldsymbol{W}=t$.

At last, after the weights are determined, we calculate the output of the output layer neuron according to the following formula:

$$\boldsymbol{n}^i=\frac{\mathbf{LW}_{2,1}\mathbf{a}^i}{\sum_{j=1}^{Q}a_j^i},\quad i=1,2,\cdots,Q \quad (8)$$

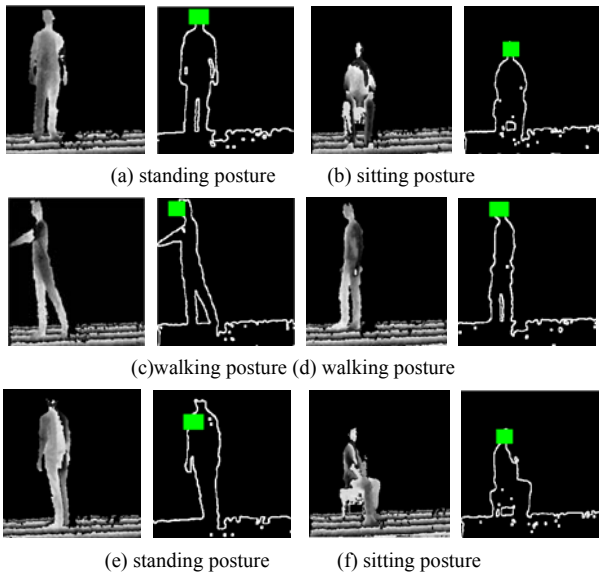
$$\boldsymbol{y}^i=\text{purelin}(\boldsymbol{n}^i)=\boldsymbol{n}^i,\quad i=1,2,\cdots,\mathrm{Q} \quad (9)$$

V. EXPERIMENTAL RESULTS

The whole experiment include three parts, part I is locating the characters in image, part II is extraction the character based on the depth of region, part III is feature extraction and classification recognition.

A. Locating the Characters In image

Because the head and shoulders of human form a curve similar to the shape Ω, we use the Ω template and MAD algorithm to search and match the image. The following figures show the matching result of three different postures of head. Left figure shows the original image, right figure shows the matching region of the head.



(a) standing posture (b) sitting posture

(c)walking posture (d) walking posture

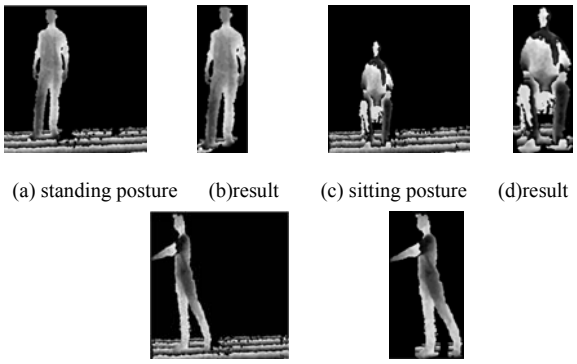
(e) standing posture (f) sitting posture

FIGURE VII. HEAD LOCATION EXPERIMENT RESULTS

Sub graph (a), (b), (d), (f) are the correct results, the curves of figure (a) and (b) are more standard. Although the curves of figure (d) and (E) have a certain deviation, the location results are still acceptable for the algorithm allows the existence of the error. But some matching results are not very ideal, such as (c) and (E). The sub graph (c) can be ignored, because the most of the head region has been located and the Ω curve is not standard. Due to the poor image depth effect, figure (E) has not full access to the head region which causes inaccurate head matching. But the inaccurate head location will affect the selection of the next seed zone.

B. Extraction the Character Based on the Depth of Region

The main purpose of the character extraction experiment is to determine the feasibility of the depth based region growth algorithm proposed in this paper. Region extraction results for several different postures are shown in FIGURE VIII.



(a) standing posture (b)result (c) sitting posture (d)result

(e) walking posture (f) result

FIGURE VIII. CHARACTER EXTRACTION RESULTS

It can be seen from the experimental results that the regional growth algorithm is feasible. For the extraction of the characters, some irrelevant areas such as the ground are removed. Since the ground is at the same depth as the human foot, the ground will be extracted as the human body similar region when the region is grown.

C. Feature Extraction and Classification Recognition

In this section, we classify the different postures by feature extraction, feature training and construct GRNN classifier. Our experiment are carried out to identify the three kinds of posture respectively namely standing posture, walking posture and sitting posture. We use ten images as a training sample and fifteen images as a test sample to test each posture. But there will be some error classifications in the processing, and table 1 shows the correct rate for each posture.

The experimental results show that the sitting posture has the highest false positive for the sitting posture is relatively complex, and has the interference of the stool. In standing posture and walking posture error recognition, the error usually occurs for the two kinds of postures are high similarity in some cases causing miscarriage of justice.

TABLE I. POSTURE RECOGNITION ACCURACY

posture / correct rate /time	sitting	standing
correct rate (%)	27.27%	73.34%
false positive (%)	72.73%	26.66%
running time(s)	12.72	11.61

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we address the problem of detecting human posture from a depth image. Firstly, we use a head model with the shape Ω to locate the human position in the image. The human detecting includes three sections that are edge extraction, template matching and human detection. Then the HOG feature is calculated to get the characteristic vector. We use a GRNN model to classify the human posture in a depth image. Our experiments prove that our method is feasible. The further research is how to improve the correct rate of recognition and reduce processing time.

ACKNOWLEDGMENT

This work was supported by the Research Program of Hebei Educational Committee Grant No.QN2015109 and Research of Yanshan University for Youths Grant No. 15LGA009.

REFERENCES

[1] G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] Ferrari V, Marin-Jimenez M, Zisserman A, “ Pose search: retrieving people using their pose”, Computer Vision and Pattern Recognition, 2009,pp.1-8.

[3] Marchin E, Ferrari V. “Better appearance models for pictorial structures”, British Machine Vision Conference, 2009,23(1), pp.345-353.

[4] Jiang H.” Human pose estimation using consistent max covering”, Pattern Analysis and Machine Intelligence, 2011, 33(9), pp. 1911-1918.

[5] Shotton J, Sharp T, Kipman A, et al. “Real-time human pose recognition in parts from single depth images”, Communications of the ACM, 2013, 56(1), pp. 116-124.

[6] Ross Girshick, Jamie Shotton, et al. “Efficient Regression of General-Activity Human Pose from Depth Images” , Computer Vision(ICCV), 2011, pp.18-24.

[7] Keskin C, Kıraç F, Kara Y E, et al. “Real time hand pose estimation using depth sensors”, Consumer Depth Cameras for Computer Vision. Springer London, 2013, pp. 119-137.

[8] DENG Rui,ZHOU Ling-ling,YING Ren-dong, “Gesture extraction and recognition research based on Kinect depth data”, Application Research of Computers, 2013, 30(4), pp.1263-1266.

[9] Ye M, Wang X, Yang R, et al. “Accurate 3D pose estimation from a single depth image”, Computer Vision (ICCV), 2011, pp. 731-738.