A Novel Multiple Person Pose Estimation Optimization Model Utilizing Genetic Algorithm

Qing Zhang1, Lei Ding1, Kai Qing Zhou1 and Jian Feng Li1
1College of Information Science and Engineering, Jishou University, Jishou, 416000, Hunan, China.
Email: yylxdinglei@126.com

Abstract. For traditional human pose estimation models rely on a large amount of human body feature information, this paper proposes an optimization model using genetic algorithm to solve the problem of multiple person body part assembly. Different from other human body parts assembly method. The method proposed in this paper depends on the joints position information, namely the sum of the connection distances between the joints as the objective function, and finds the optimal value to obtain the best human pose assembly information. The simulation results show that compared with the traditional OpenPose model, the model proposed in this paper can obtain the same human skeleton using less position information.

1. Introduction
Human pose estimation technology has developed rapidly and widely used in the research field of computer vision, and study how to accurately recognize the target human body from a given image and obtain the target human body pose estimation [1-2]. Compared with single-person pose estimation, multi-person pose estimation is more complicated and requires more information.

In the multi-person method, there are two detection methods, namely the top-down method [3-5] and the bottom-up method [7-8]. Cao et al. achieve a fast connection to the trunk body (Part Affinity Fields, PAFs) joints to match different postures [8]. The advantage of this method is that it runs fast and the average detection accuracy is well. However, the method needs to change the articulation joints coordinates in the skeleton PAFS assembly stage, so in the actual test, the case due to missing PAFs human skeleton caused the connection error occurs. Since the human body structure has fixed connection rules, multiple joints will produce connection modes. In this case, the solution to body part assembly can be converted to intelligent optimization algorithms. The intelligent optimization algorithm can quickly find the optimal solution that meets the requirements from a large number of possible results, and at the same time, which can reduce the degree of dependence of the algorithm on prior information, thereby achieving the same effect.

The main innovation of this paper is to build a model using genetic algorithm to obtain the multiple person body part assembly with less position information.

2. The Traditional Human 2D Pose Estimation Model
The OpenPose model is essentially a dual-parallel convolutional network model. It uses two convolutional networks to complete body part assembly in the image.
As shown in Figure 1, the VGG-19 network is used to extract the input image’s bottom-level features. One network uses a non-maximum value suppression algorithm to generate a confidence map for human joints’ coordinate positioning. The other network uses the local area affinity field algorithm to connect key parts to essential limbs. After that, the two convolutional networks’ processing results are summarized, and the Hungarian algorithm is used for body part assembly. The final output of the human pose in the image to be tested.

3. Problem description and Genetic Algorithm Application

In the body part assembly process of the OpenPose model, the Hungarian algorithm is used to solve the problem of maximum bipartite graph matching. However, in the assembly process, we need to rely on the joint confidence map, and the local affinity field information between the limbs. The local affinity field information between the limbs is susceptible to background interference and errors, which leads to mistakes in body part assembly.

3.1 Problem description

In the image of two-person interaction, there are 36 joints and two pose sequences. According to this feature, the assembly process is transformed into sequentially selecting 18 human body joints and selecting an optimal point from each part’s candidate points for body part assembly. However, if we try all the joint connections, there will be $2^{18}$ connection methods, and this calculation method will consume a lot of computing resources. Therefore, we propose a connection method based on distance constraints for combining human joints in two-person images. The inspiration for the method comes from solving the TSP problem by selecting the joints with a shorter distance to connect to minimize the total joint connection distance. Among them, the distance constraint is divided into two types of distance constraints, as follows:

1) The overall goal is the shortest total length of the connection distance between the joints, as shown in equation (1), where $n$ is the number of related nodes in the two-person image, and $j$ is the currently detected joint.

$$f(p_1, p_2, \cdots, p_n) = \min\left(\sum_{j=1}^{n} d(p_j, p_{j+1})\right)$$

2) The connection distance between each group of candidate’s joints should be less than the length threshold. As shown in Figure 2, $c_1$, $c_2$ are the point cluster centers, $p_1$ to $p_6$ are joints, $L_{12}$ to $L_{46}$ are the connections between the joints, and TH value is the set threshold.

$$\forall n \in \{L_{12}, L_{13}, L_{15}, L_{16}, L_{42}, L_{43}, L_{45}, L_{46}\}, TH \geq n$$

3.2 The implementation process of the genetic algorithm

This paper proposes a model using a genetic algorithm to solve the matching problem with no reference limb confidence information during body part assembly under dual constraints. The comparison flowchart between the traditional model and the improved model is shown in Figure 3. First, random
coding is performed on different parts of different people to create chromosomes and initial populations. Then through a series of operations such as selection, crossover, and mutation, the objective function is the fitness function value. Then select and mutate operations. The optimal individuals are selected from the iterative population to form an elite population, and a solution that satisfies the dual constraints is selected. The joint sequence of a person in the image is obtained.

![Figure 2. Schematic diagram of double joint connection](image)

There are N connection modes in the whole connection process, and each connection mode is a chromosome. There are 18 genes in each chromosome, and each gene corresponds to the corresponding joints. The joints number of each part is [0,1].

There are 18 human joints and their distribution and serial numbers are shown in Figure 4. In order to avoid unreasonable joint connection methods, for example, the arm nodes are connected to the leg nodes, the corresponding distances are specified according to the connection sequence of the human joint pose, and the adjacency matrix is shown in Figure 5, where the red area on the way is Correspondence that can be connected. The corresponding connection’s distance value is calculated according to the adjacency matrix, and the distance matrix corresponding to the adjacency matrix is obtained. Then, according to each chromosome’s different connection sequence, the sum of the accumulated distance is calculated, and this value is the target fitness value to be solved.

![Figure 3. Comparison flowchart two models](image)
Figure 4. Human joints and connection sequence.

Figure 5. Adjacency matrix between nodes

Due to constraint (2), it is necessary to set a threshold and perform distance comparison analysis with all possible connections to distinguish the relationships that meet the constraint. However, due to the
limitation of different image resolutions, the threshold value needs to be changed according to the image’s size to conform to the usual posture connection method.

First, we use the k-means algorithm to perform cluster analysis on all joints to solve the problem. Since the number of people is known, the two cluster centers positions can be obtained, and the Euclidean distance $c_{dis}$ between the cluster centers can also be calculated. Next, based on the computed distance matrix between the related nodes of the reaction. The distance value from the center point to the left and right hip joints is extracted, the minimum value and the second minimum value are selected from them, and the average value $p_{xmin}$ is calculated. Since the position of the character in the image is uncontrollable, it is necessary to judge $c_{dis}$ and $p_{xmin}$. When $c_{dis} > p_{xmin}$, there is a distance threshold $th = c_{dis}$, when $c_{dis} < p_{xmin}$, there is a distance threshold $th = p_{xmin}$.

The mutation operation in the genetic algorithm swaps two genes’ positions on a parent chromosome to form a new offspring individual. According to the joints’ characteristics in the two-person image, constraint (2) is first introduced here so that the distance between each pair of genes meets it. Thereby, a binary matrix of the current posture connection of constraint (2) can be obtained. In a binary matrix, the true value and the false value mean to exceed the constraint and satisfy the constraint condition, respectively. Find the true value index in the matrix, which also means the position of the gene in the chromosome that requires mutation operation. Then replace the candidate points corresponding to the gene, as shown in Figure 6.

\[
\begin{array}{cccccccccccc}
\text{False} & \text{False} & \text{False} & \text{True} & \text{True} & \text{False} & \text{False} & \text{False} & \text{False} & \text{False} & \text{False} & \text{False} \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0
\end{array}
\]

Figure 6. Local mutation operation process

4. Experimental

The test data comes from 8 videos in TikTok, each video picture taken, and the picture contains two persons. We send 8 test pictures into the OpenPose model for prediction and compare the predicted human skeleton with the algorithm predicted by this article. The result is as follows.

As shown in Figure 7, the first image is the original image. The second image is the human 2D pose predicted by the OpenPose model. The third image is the human 2D pose obtained by the predicted joints and using the optimization algorithm. The basic parameters of GA (No crossover rate), such as the population number, the number of evolutions, and the mutation rate are assigned as 90, 10, 0.5.

As shown in Table 1, each connection optimal solution corresponding to a minimum error and average error. The value of the connection solution represents the total connection distance of a person's joints. Observing the minimum error and the average error, our model generally finds the optimal solution with a longer distance more times when searching for the optimal solution (Standard optimal solution second column). In pictures with complex characters, the optimal solution with a shorter distance is difficult to find. Since there are only two people in the picture, as long as an optimal solution is found, the remaining joints are all the joints of the other person.
Table 1. All test data results

| Test image | Standard optimal connection solution | Experimental optimal connection solution | Experimental worst connection solution | Average | Minimum error | Average error |
|------------|--------------------------------------|----------------------------------------|---------------------------------------|---------|---------------|--------------|
| 1          | 1149                                 | 1149                                   | 1583                                  | 1414    | 0             | 765          | 95          |
| 2          | 1046                                 | 1128                                   | 1397                                  | 1193    | 82            | 0            | 147         | 177        |
| 3          | 902                                  | 1062                                   | 1241                                  | 1120    | 0             | 160          | 213         | 86         |
| 4          | 1667                                 | 1762                                   | 2063                                  | 1887    | 95            | 0            | 220         | 125        |
| 5          | 1455                                 | 1603                                   | 1455                                  | 1710    | 0             | 148          | 153         | 5          |
| 6          | 1413                                 | 1588                                   | 1443                                  | 1717    | 30            | 145          | 250         | 54         |
| 7          | 1500                                 | 1565                                   | 1550                                  | 1767    | 0             | 115          | 160         | 65         |
| 8          | 874                                  | 989                                    | 989                                   | 1283    | 113           | 0            | 233         | 227        |

5. Conclusion
In this paper, we use genetic algorithms to find the best multiple person body part assembly. The results show that our model can be consistent with the connection result of the OpenPose model using only joints coordinate information. Achieved the purpose of reducing the dependence of the human pose estimation model on features.

References
[1] Zhang F, Zhu X, Ye M. 2019. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3517-3526
[2] Kreiss S, Bertoni L, Alahi A. Pitpaf 2019. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 11977-11986
[3] Wei S, Ramakrishna V, Kanade T, et al. 2016 *IEEE Conference on Computer Vision and Pattern Recognition*. 4724-4732
[4] Papandreou G, Zhu T, Kanazawa N, et al. 2017 *IEEE Conference on Computer Vision and Pattern Recognition*. 3711-3719
[5] He K, Gkioxari G, Dollar P, et al. 2017 *Int. Conference on Computer Vision*. 2980-2988
[6] Fang H, Xie S, Tai Y, et al. 2017 *Int. Conference on Computer Vision*. 2353-2362
[7] Xiao B, Wu H, Wei Y, et al. 2018 *European Conf. on computer vision (ECCV)*. 466-481
[8] Cao Z, Simon T, Wei S, et al. 2017 *Computer Vision and Pattern Recognition (CVPR)*. 1302-1310

Acknowledgement
This work is supported by the National Natural Science Foundation of China (Nos. 61966014 and 61962023), and the Jishou University Graduate Research and Innovation Project (JDY19045).