Journal of Fiber Bioengineering and Informatics 8:2 (2015) 249–258
doi:10.3993/jfbim00116

Pose Estimation Using Local Adjustment with Mixtures-of-parts Models ⋆

Peng Cai [a], Dehui Kong [a,*], Shaofan Wang [a], Baocai Yin [a]
Xiaogang Ruan [b], Yi Huo [a]

[a] *Beijing Key Laboratory of Multimedia and Intelligent Software Technology, College of Metropolitan Transportation, Beijing University of Technology, Beijing 100124, China*

[b] *School of Electronic Information and Control Engineering, Beijing University of Technology Beijing 100124, China*

Abstract

Articulated pose estimation with mixtures-of-parts decomposes human body into several local component templates with springs connecting each other. Such a method fails in precisely estimating human pose especially due to the defects of tree models when human has the complicated pose of body. To address this problem, we propose pose estimation using local adjustment with mixtures-of-parts models. We can achieve the most suitable pose of body by the blending and selecting strategy based on the full score and the corresponding attributes of limbs and body. The experiments show that the estimation effect of human pose of our method is better than the previous method based on articulated pose estimation with mixtures-of-parts.

Keywords: Articulated Model; Mixtures-of-Parts; Pose Estimation; Local Adjustment; Blending and Selecting Strategy

1 Introduction

Pose estimation refers to as localization of components of human bodies within images. Part representation [1] and pictorial structure [2, 3] are two important tools for pose estimation. The first approach models human body as a collection of parts while the second approach represents human body as a collection of component templates with springs connecting each other. Compared with part representation based approach, pictorial structure based approach improves the preciseness of pose estimation by adding the connection between adjacent components of bodies. However, difficulties of full-body pose estimation arise from many aspects such as a large number

⋆Project supported by NSFC (No. 61227004, 61370120), PHR (IHLB), Beijing Municipal Science and Technology Project (No. Z141100006014032) and Beijing Municipal Education Commission (No. km201410005013).

*Corresponding author.
Email address: kdh@bjut.edu.cn (Dehui Kong).

1940–8676 / Copyright © 2015 Binary Information Press & Textile Bioengineering and Informatics Society
June 2015

of degrees of freedoms, varied clothing and body shape, change of viewpoints, and articulated occlusion.

We propose pose estimation using a local adjustment of the localization of body components. The adjustment is achieved by blending a full-body score with important attributes of components, including the non-overlapping area of components, and the HoG convolutions of components.

The main contribution of our work is that, the most suitable pose of whole body/local components (upper arms, thighs, forearms, shanks) are captured by local pose adjustment, based on the blending and selecting strategy of full score and global attributes of body/local attributes of limbs.

The remainder of this paper is organized as follows: Section 2 discusses previous work on pose estimation of human; the local pose adjustment of articulated model with mixtures-of-parts is proposed in Section 3; Section 4 gives experimental results, which show that the pose estimation of our method can perform better than the previous method [3] of articulated pose estimation with mixtures-of-parts in most cases; finally, the conclusion and future work are given in Section 5.

2 Previous Work

Many scholars propose effective algorithms to estimate pose of body in video domain or static images and we list some important methods of estimating human pose as follows.

O'Rourke and Badler [4], Hogg [5], Rohr [6] propose the classic model-based methods to address pose estimation in the video domain. The pose estimation of body of static image is becoming more and more important, tree models are efficient in pose estimation [2], but have some difficulties in pose estimation of limbs (e.g., the cases of double-counting). Loopy models require approximate inference strategies such as importance sampling [2, 7], loopy belief propagation [8], or iterative approximations [9]. Recently, branch and bound algorithms [10, 11] with tree-based lower bounds are presented to globally solve such problems. Lan and Huttenlocher [12] consider tracking the double-counting cases by using stronger pose priors. Tran and Forsyth [13] present that the above solutions maybe more susceptible to overfitting to statistics of a particular dataset.

Yang and Ramanan [3] propose a method of human pose estimation in static images based on a new part models, which is a general, flexible mixture model for capturing contextual co-occurrence relations between parts, and a spring models for encoding spatial relations. Their method is almost invalid for the complicated pose of body (e.g., double counting of limbs) due to the defects of tree structure of articulated part models with mixtures-of-parts.

This paper presents a local pose adjustment method of articulated models with mixtures-of-parts based on tree structure. We achieve the better estimation effect of human pose than the method of [3] in most cases, by the blending and selecting strategy of full score and corresponding attributes of body and limbs.

3 Local Pose Adjustment Based on Articulated Model with Mixtures-of-parts

As [3], we let I be an image, let $l_i = (x, y)$ be the pixel location of part i and let t_i be the mixture type of part i, where $i \in \{1, \ldots, K\}$, $l_i \in \{1, \ldots, L\}$, and $t_i \in \{1, \ldots, T\}$ ($K = 26$ and $T = 6$ in

our paper). We also let $z_i = (l_i, t_i)$ be the pixel location and the mixture type of part i.

3.1 An Overview of Local Adjustment Based on Articulated Model with Mixtures-of-parts

We give the basic idea of local adjustment based on articulated model with mixtures-of-parts as below:

(1) We use articulated model with mixtures-of-parts to compute full score of body, and achieve the most suitable pose of body for local pose adjustment by the blending and selecting strategy of full score and global attributes of body.

(2) At the articulated positions of upper arms or thighs, we use articulated model with mixtures-of-parts to compute full score of them, and achieve the most suitable pose of them by the blending and selecting strategy of full score and local attributes of them.

(3) At the articulated positions of forearms or shanks, the procedure is the same as step 2 above.

3.2 Definition of Full Score of Body and Limbs

As shown in Fig. 1, we use articulated model with mixtures-of-parts to compute full score of body and limbs, where a circle represents a part of body and a line between two circles represents the connected relation between parts of body. We use 26 parts to represent the 26 nodes of tree structure of body, where upper arm or thigh or forearm or shank has the corresponding nodes respectively (see Fig. 1).

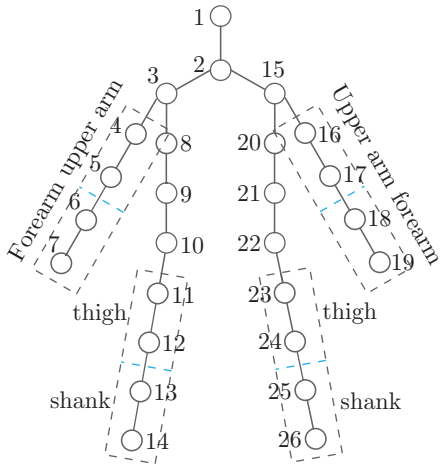


Fig. 1: The articulate model with mixtures-of-parts

We get the M poses of body by setting the threshold T of full score of body.

$$score_{i,j}(z_{i,j}) = \phi_{i,j}(I, z_{i,j}) + \sum_{k \in kids(i,j)} m_{i,k}(z_{i,j}) \tag{1}$$

$$m_{i,j}(z_{i,k}) = \max_{z_{i,j}} [score_{i,j}(z_{i,j}) + \psi_{j,k}(z_{i,j}, z_{i,k})]. \tag{2}$$

At the position of part j of pose i of body, we compute the corresponding full score by Eq. (1) ($1 \le i \le M$, $1 \le j \le 26$). $z_{i,j}$ is the location and the mixture type of part j of pose i of body, $\phi_{i,j}(I, z_{i,j})$ is corresponding to the convolution between the HoG vector [14] of part j of pose i of body and the HoG template by image I, and $\psi_{j,k}(z_{i,j}, z_{i,k})$ is corresponding to the relative location between parts j and k of pose i of body. Then $score_{i,1}(z_{i,1})$ is the full score of root part of body, which is the full score of pose of body.

We denote

$$score'_{i,j}(z_{i,j}) = m_{i,j+1}(z_{i,j}), \tag{3}$$

to be the full score of articulated position j of pose i of body by Eq. (2) ($j \in \{3, 10, 22, 15, 5, 12, 24, 17\}$), which is the full score of pose i of articulated position j of corresponding limb (upper arm: $j \in \{3, 15\}$; thigh: $j \in \{10, 22\}$; forearm: $j \in \{5, 17\}$; shank: $j \in \{12, 24\}$).

3.3 Definition of Corresponding Attributes of Body and Limbs

To select the most suitable pose of body for local pose adjustment and the most suitable poses of limbs, we define the corresponding attributes of body and limbs (i.e., upper arms, thighs, forearms and shanks) as below.

We denote

$$A^a_{i,j} = \widehat{A^a_{i,j}} + k \cdot A \tag{4}$$

to be the corresponding non-overlapping area (i.e., number of pixels) function between the last part of limb (arm or leg) and the corresponding location of body, where i represents the index of pose of upper arm or thigh ($1 \le i \le N$), j represents the index of the articulated position of upper arm or thigh ($j \in \{3, 10, 22, 15\}$), $\widehat{A^a_{i,j}}$ is the non-overlapping area between the part $(j+4)$ and the location of body besides parts $(j+1)$ to $(j+4)$, 1 and 2. k is the ratio corresponding to arm or leg, and A represents the area of part. When $\widehat{A^a_{i,j}}$ is larger, the overlapping area between the last part of limb and the corresponding location of body is smaller, which can effectively reduce the false pose estimations of upper arm or thigh due to the defects of tree models, and $k \cdot A$ can make a balanced role between true and false local pose adjustments of limb.

We denote

$$A^b_{i,j} = \widehat{A^b_{i,j}} + \overline{A^b_{i,j}} \tag{5}$$

to be the corresponding non-overlapping area function of between the four parts of limb (arm or leg) and the corresponding location of body, where i represents the index of pose of forearm or shank ($1 \le i \le N$), and j represents the index of the articulated position of forearm or shank ($j \in \{5, 12, 24, 17\}$). $\widehat{A^b_{i,j}}$ is the non-overlapping area between the four parts of limb and the location of body besides them and parts 1 and 2, $\overline{A^b_{i,j}}$ is the area of limb. When $A^b_{i,j}$ is larger, the overlapping area between forearm or shank and the corresponding location of body is smaller, which can effectively reduce the false pose estimations of forearm or shank due to the defects of tree models.

The normal vector between two parts of body is computed by their centers in the image. We denote

$$P_{i,j}(\alpha) = \min\{P^{j-1,j+2}_{i,j} \cdot P^{j,j+1}_{i,j}, \alpha\} \tag{6}$$

to be the dot product of normal vectors between four parts of limb (arm or leg), where i represents the index of pose of forearm or shank, j represents the index of the articulated position of forearm or shank ($j \in \{5, 12, 24, 17\}$), and $P_{i,j}^{j-1,j+2}$ is the normal vector between parts $(j-1)$ and $(j+2)$ of limb, and $P_{i,j}^{j,j+1}$ is the normal vector between parts j and $(j+1)$ of limb, α is the threshold. $P_{i,j}^{j-1,j+2} \cdot P_{i,j}^{j,j+1}$ of true pose of limb is close to 1, which can be the constraint of blending function of local attributes of limb.

We let I be the image of pose estimation of body, and compute the convolution between the HoG vectors of parts of image I and the corresponding HoG templates of articulated model with mixtures-of-parts. We rotate the image I by 90 degree counterclockwise, get the corresponding image $\widehat{I}$, and compute the convolution between the HoG vectors of parts of image $\widehat{I}$ and the corresponding HoG templates.

H_i^k is denoted to be the HoG convolution of part k of pose i of body by the image I ($k \in \{1, 26\}$). $H_{i,j}^k$ and $\widehat{H_{i,j}^k}$ are denoted to be the HoG convolutions of part k of articulated position j of pose i of forearm or shank by the images I and $\widehat{I}$ respectively, where $1 \le i \le N$, $j \in \{5, 12, 24, 17\}$, and $k \in \{j+1, j+2\}$ (the same below). H_i^k, $H_{i,j}^k$ and $\widehat{H_{i,j}^k}$ are corresponding to the matching degree between the HoG vectors of part and the corresponding HoG template by the image I and $\widehat{I}$ respectively, which are close to true values of the corresponding parts of local pose of limb with increase of them.

We let

$$H_i = \sum_{k=1}^{26} H_i^k \tag{7}$$

be the sum of the HoG convolutions of 26 parts of pose i of body by image I. We let

$$f_{i,j} = h_{i,j}^{j+1} * h_{i,j}^{j+2} * \widehat{h_{i,j}^{j+1}} * \widehat{h_{i,j}^{j+2}} \tag{8}$$

be the function of the quantitative HoG convolutions of parts of forearm or shin, which is corresponding to the matching degree between the HoG vectors of parts of forearm or shank and its HoG templates in the image I and $\widehat{I}$, where

$$h_{i,j}^{j+m} = \begin{cases} 1.15 & \text{if } H_{i,j}^{j+m} > 0 \\ 1 & \text{otherwise} \end{cases} \quad m = 1, 2$$

$$\widehat{h_{i,j}^{j+m}} = \begin{cases} 1.15 & \text{if } \widehat{H_{i,j}^{j+m}} > 0 \\ 1 & \text{otherwise} \end{cases} \quad m = 1, 2$$

$h_{i,j}^{j+m}$ and $\widehat{h_{i,j}^{j+m}}$ are the quantitative values of $H_{i,j}^{j+m}$ and $\widehat{H_{i,j}^{j+m}}$ respectively.

We let

$$F_{i,j} = H_{i,j}^{j+2} + \widehat{H_{i,j}^{j+2}} + \min\left\{H_{i,j}^{j+2}, \widehat{H_{i,j}^{j+2}}\right\}$$

be the function of the HoG convolutions of part $(j+2)$ of articulated position j of pose i of forearm or shin, which is corresponding to the matching degree between the HoG vectors of part $(j+2)$ and its HoG template in the image I and $\widehat{I}$.

We denote

$$\widehat{F_{i,j}(\beta)} = \min\{F_{i,j}/\overline{F_j^n}, \beta\} \tag{9}$$

to be the quantitative function of $F_{i,j}$ of pose i of articulated position j of forearm or shin, where

$$\overline{F_j^n} = \sum_{i=1}^{n} F_{i,j} * s(i,j) \Big/ \sum_{i=1}^{n} s(i,j) \tag{10}$$

is the weighted average of $F_{i,j}$ of the n poses of forearm or shin corresponding to articulate position j in descending order of $s(i,j)$ ($n = 0.15 \cdot N$), $s(i,j) = score'_{i,j}(z_{i,j})$ is computed by Eq. (3), and β is the threshold.

3.4 Local Pose Adjustment of Body

We propose a local pose adjustment method of human pose (see Fig. 2). First, we set the threshold T of full score to select the corresponding poses of body, and from them we select the most suitable pose of body for local pose adjustment, by the blending and selecting strategy of full score and global attributes of body; Second, we achieve the most suitable poses of upper arm and thigh, by the blending and selecting strategy of full score and local attributes. Third, we achieve the most suitable poses of forearm and shank, by the blending and selecting strategy of full score and local attributes.

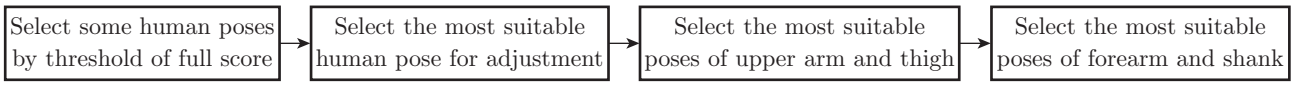


Fig. 2: The frame of local adjustment of human pose based on articulated model with mixtures-of-parts

We let

$$S_i = score_{i,1}(z_{i,1}) - T \tag{11}$$

be the full score of the selected pose i of body by setting the threshold T which makes $score_{i,1}(z_{i,1}) \geq T$ (i.e. $S_i \geq 0$), where $score_{i,1}(z_{i,1})$ is the full score of position of root part of pose i by Eq. (1).

Based on the selected poses i of body, we denote

$$S'_i = S_i \cdot H_i \cdot A_i^c \tag{12}$$

to be the blending function of body, where S_i is the full score of pose i of body by Eq. (11), H_i is the sum of the HoG convolutions of 26 parts of pose i of body by Eq. (7), and A_i^c is the area of bounding box of all parts of pose i of body.

We get the N poses of body in descending order of S_i, and from them we select the most suitable pose i corresponding to the maximum value of S'_i for the local pose adjustment of next step by Eq. (12).

Based on the most suitable pose of body for local pose adjustment of body, we compute the most suitable pose of upper arm or thigh. We denote

$$S'_{i,j} = S_{i,j} \cdot A_{i,j}^a \tag{13}$$

to be the blending function of upper arm or thigh corresponding to articulated position j ($j \in \{3, 10, 22, 15\}$), where $S_{i,j} = score'_{i,j}(z_{i,j})$ is the full score of arm or leg corresponding to part j by Eq. (3), and $A_{i,j}^a$ is the non-overlapping area function of arm or leg by Eq. (4).

We get the N poses of upper arm or thigh in descending order of $S_{i,j}$, and from them we select the most suitable pose of upper arm or thigh corresponding to the maximum value of $S'_{i,j}$ by Eq. (13).

We denote

$$\overline{S'_{i,j}} = \widehat{S_{i,j}} \cdot f_{i,j} \cdot \widehat{F_{i,j}(\beta)} \cdot P_{i,j}(0.5) \tag{14}$$

to be the blending function of forearm or shank corresponding to articulated position j ($j \in \{5, 12, 24, 17\}$), where

$$\widehat{S_{i,j}} = \min\{S_{i,j}/\overline{S_j}, 0.8\}$$

is the quantitative function of $S'_{i,j}$, and

$$\overline{S_j} = \sum_{i=1}^{n} S_{i,j}/n$$

represents the weighted average of $S_{i,j}$ of the n poses of forearm or shank corresponding to articulate position j in descending order of $S_{i,j}$ ($n = 0.15 \cdot N$), where $S_{i,j} = score'_{i,j}(z_{i,j})$ is the full score of forearm or shank corresponding to articulated position j by Eq. (3), $f_{i,j}$ is the function of the quantitative HoG convolutions of parts of forearm or shank by Eq. (8), $\widehat{F_{i,j}(\beta)}$ is the quantitative function of $F_{i,j}$ of forearm or shank by Eq. (9), and $P_{i,j}(0.5)$ is the dot product of normal vectors between parts of arm or leg corresponding to articulated position j by Eq. (6).

Based on the most suitable pose of upper arm or thigh above, we get the N poses of forearm or shank in descending order of $S_{i,j}$, and from them we select the n poses corresponding to the n maximum value of $\overline{S'_{i,j}}$ by Eq. (14) ($n = 0.15 \cdot N$).

We denote

$$\overline{\overline{S'_{i,j}}} = S_{i,j} \cdot A^b_{i,j} \cdot f_{i,j} \cdot \widehat{F_{i,j}(\beta)} \cdot P_{i,j}(0.7) \tag{15}$$

to be the another blending function of forearm or shank corresponding to articulated position j ($j \in \{5, 12, 24, 17\}$), $A^b_{i,j}$ is the non-overlapping area function of arm or leg defined by Eq. (5), $S_{i,j}$, $f_{i,j}$, $\widehat{F_{i,j}(\beta)}$ and $P_{i,j}(0.7)$ are defined the same as Eq. (14).

We achieve the most suitable pose of forearm or shank by selecting the maximum $\overline{\overline{S'_{i,j}}}$ value of the above n poses of forearm or shank by Eq. (15).

4 Experimental Results

All experiments are performed on a single Intel(R) 3.20 GHz CPU, 4 G RAM. We compare the articulated pose estimation method based on mixtures-of-parts [3] with our method.

We use articulated model with mixtures-of-parts to estimate human pose, and adjust local pose of body by the blending and selecting strategy based on full score and corresponding attributes of body and limbs. The number of parts of articulated model is set to 26, all parts are set to 5×5 HoG cells in size and 6 mixture types per part. For the method of [3], the parameter is threshold of full score of body: T. For our method, the parameters include:

T : threshold of full score of body pose

N : number of poses of upper arm or thigh or forearm or shank

β : threshold of quantitative function $F_{i,j}$ of pose i of forearm or shank corresponding to articulated position j

k : ratio of area of part corresponding to arm or leg

For the method of [3], T is set to -1; for our method, T is set to -1, N is set to 100, β is set to 0.4 and 0.8 for forearm and shank respectively, and k is set to 3 and 1 for arm and leg respectively.

In Fig. 3, there are 12 groups of pictures of comparative results. The left picture is the estimation effect of pose of the method of [3], and the right picture is the estimation effect of pose of our method. We see that the estimation effect of body pose of our method is better than the method of [3] in most cases. The reason is that, there exists exponential number of combination of poses of limbs, we can get the better pose of limbs by the blending and selecting strategy based on full score and local attributes of limbs. In the last group of pictures of Fig. 3, the estimation effect of pose of the method of [3] is better than our method because of some non-perfect aspects of ours.

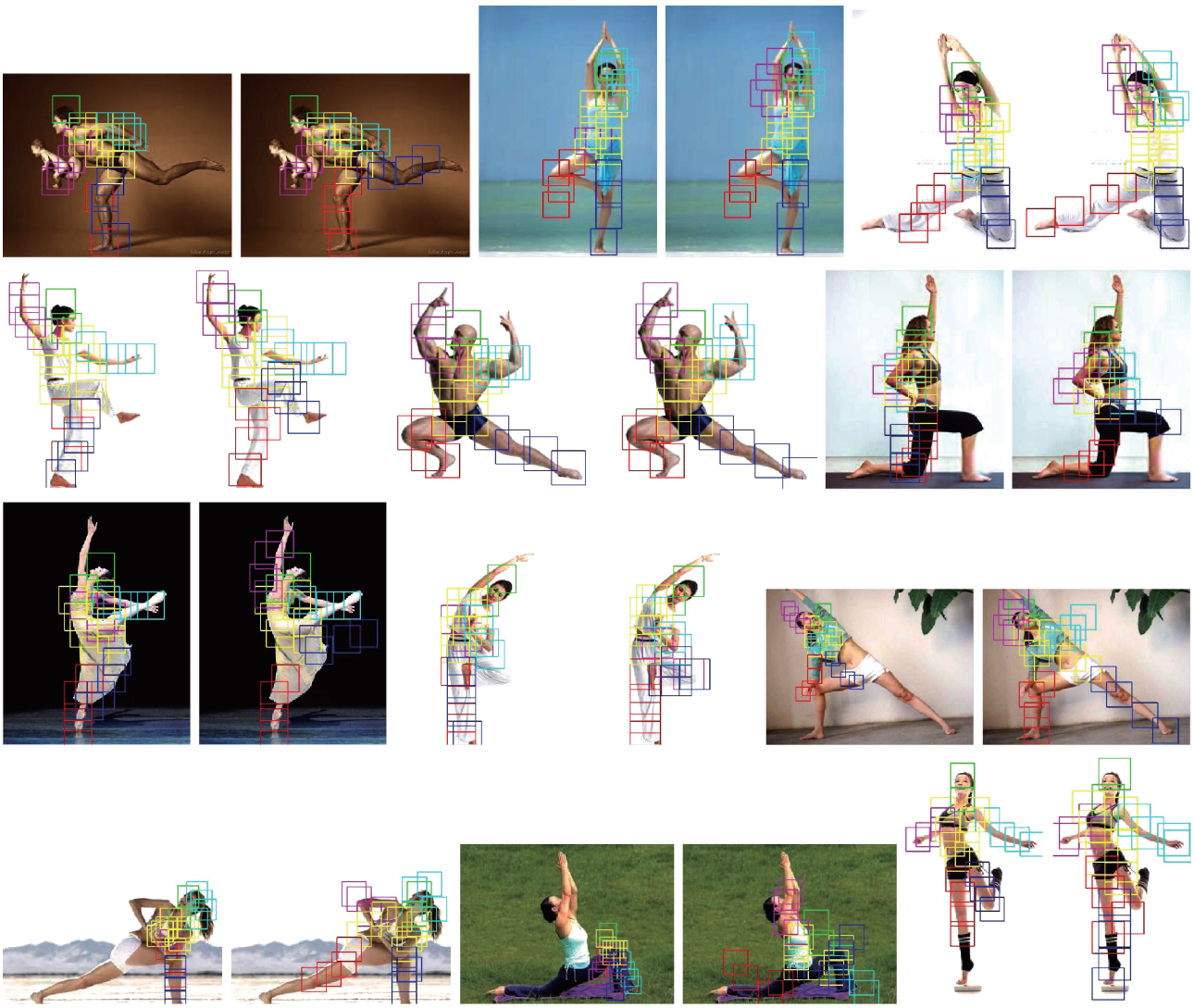


Fig. 3: The effect comparison of pose estimation of 12 pictures (left: the method of [3]; right: our method)

In Fig. 4, the pictures of the first column are the poses of arms or legs before the adjustment of them, the pictures of the second column are the poses of upper arms or thighs after the adjustment of them by Eq. (13), the pictures of the third column are the poses of forearms or shanks after the adjustment of them by Eq. (14) and Eq. (15). In the adjustment procedure of local poses of Fig. 4, we see that the adjusted poses of arm or leg from upper arm or thigh to forearm or shank is close to the true pose of limb step by step.

From Table 1, we see that the time of pose estimation of the method of [3] is 3–6 s which is 8–11 s later than the method of [3]. The time of our method can be largely reduced with decrease of threshold N, and we will reduce the time of our method under the premise of estimation effect of poses.

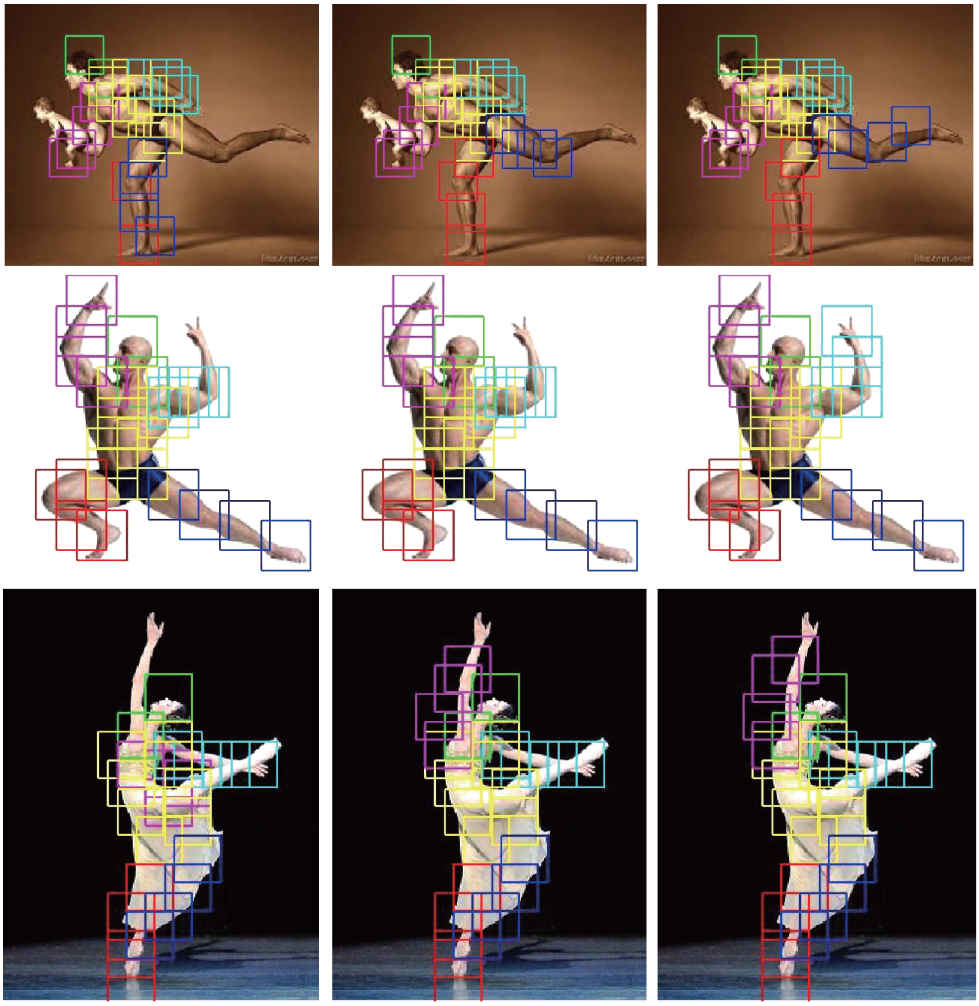


Fig. 4: The corresponding effects of pose estimation of local adjustment of our method for Pictures 1 (right leg), 5 (right arm) and 7 (left arm) in Fig. 3

Table 1: The time of pose estimation of body for the method of [14] and our method for 12 pictures in Fig. 3 (time unit: s)

The method of [3]						Our method					
3.79	3.77	3.71	3.70	3.78	3.51	14.62	14.11	13.81	12.92	14.79	14.32
5.85	3.71	4.97	3.75	3.30	3.70	17.71	14.60	16.23	14.60	13.81	13.54

5 Conclusions

We propose a local adjustment method based on articulated model with mixtures-of-parts. Our method adjusts the poses of limbs of body by the blending and selecting strategy based on full score and local attributes of limbs, which may make us fully use the exponential number of combination of poses of limbs with all kinds of full score and local attributes. The experiments show that our method can produce the better effect of pose estimation of limbs than the method of [3]. We shall propose the better blending function of full score and corresponding attributes, and achieve the more robust estimation effect of human pose.

References

[1] D. Marr, H. Nishihara, Representation and recognition of the spatial organization of three-dimensional shapes, in: Proceedings of the Royal Society of London, Series B, Biological Sciences, 200 (1140), 1978, 269-294

[2] P. Felzenszwalb, D. Huttenlocher, Pictorial structures for object recognition [J], IJCV, 61(1), 2005, 55-79

[3] Y. Yang, D. Ramanan, Articulated human detection with flexible mixtures of parts [J], TPAMI, 35(12), 2013, 2878-2890

[4] J. O'Rourke, N. Badler, Model-based image analysis of human motion using constraint propagation [J], TPAMI, 2(6), 1980, 522-536

[5] D. Hogg, Model-based vision: A program to see a walking person [J], Image and Vision Computing, 1(1), 1983, 5-20

[6] K. Rohr, Towards model-based recognition of human movements in image sequences [J], CVGIP - Image Understanding, 59(1), 1994, 94-115

[7] M. Lee, I. Cohen, Proposal maps driven mcmc for estimating human body pose in static images, in: CVPR, 2004, 334-341

[8] L. Sigal, M. Black, Measure locally, reason globally: Occlusion-sensitive articulated pose estimation, in: CVPR, 2006, 2041-2048

[9] Y. Wang, G. Mori, Multiple tree models for occlusion and spatial constraints in human pose estimation, in: ECCV, 2008, 710-724

[10] T. Tian, S. Sclaroff, Fast globally optimal 2d human detection with loopy graph models, in: CVPR, 2010, 81-88

[11] V. Singh, R. Nevatia, C. Huang, Efficient inference with multiple heterogenous part detectors for human pose estimation estimation, in: ECCV, 2010, 305-318

[12] X. Lan, D. Huttenlocher, Beyond trees: Common-factor models for 2d human pose recovery, in: ICCV, 2005, 470-477

[13] D. Tran, D. Forsyth, Improved human parsing with a full relational model, in: ECCV, 2010, 227-240

[14] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: CVPR, 2005, 886-893