Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/081L39

Iterative Human Pose Estimation based on A New Part Appearance Model

*Wang Hao, Meng Fanhui and Fang Baofu**

School of Computer and Information, Hefei Universty of Technology, Hefei, China

Received: 15 May. 2013, Revised: 9 Sep. 2013, Accepted: 10 Sep. 2013
Published online: 1 Apr. 2014

Abstract: Human pose estimation has become a hot topic in the field of computer vision , it can be used in human activity analysis, video, and any other field ,it main purpose is that detect the position, scale and direction of parts of people .Because of the result in the iterative human pose estimation based on tree-structure model is susceptible to the background, In this paper(consider static images),part appearance model is improved, an appearance model based on colour and texture for iterative human pose estimation is proposed, experimental results show this method give a better performance and accuracy while reduce the search space by people detector and grab cut.

Keywords: Computer vision ,human pose estimation,part appearance model, color texture, reduce search space.

1 Introduction

Human pose estimation can be used in the human body activities analysis, human-computer interaction and visual surveillance, etc. It is a hot topic in the field of computer vision recently. Human pose estimation is often approached in a video setting, within the context of tracking. Recent focus in the area has expanded to single image pose estimation, because such algorithms are likely useful for initialization in video tracking. Nevertheless, human pose estimation in static images has not been a molding solution at this stage. This is because: first, the human body is composed of many parts, motion is very complex, and there lacks 3d information in monocular image and video, which makes it hard to describe 2d pose change of human body simply using a unified model. Second, in different situations, background and lighting in image and video will change greatly, and the cloth of human body itself will also be changing, leading to great changes happen in the appearance of human body, which is difficult to describe unified model [1].

We briefly discuss related work on human detection in section 2, give an overview of Iterative human pose estimation based on tree-structured model in section 3, we describe our method in section 4 and give a detailed experimental results in section 5. The main conclusions are summarized in section 6.

2 Related Work

The existing work for human pose estimation can be generally divided into model-free and model-based method. Model-free pose estimation method also can be divided into learning-based method and samples-based method. Learning-based method uses training sample to study regression model from the image feature space to the human body posture space. Thus enrich the amount of the training samples into compact function expression. Extract features from new test images and put into regression model, so it can estimate human body posture now [2]. For example, Ankur Agarwal [3] use associated shape of human silhouettes as descriptors, the Relevance Vector Machine as regression, with sparse bayesian nonlinear regression learn a compact mapping model and map the feature to the parameters space, at last ,output the input features' corresponding human pose parameters; Romer Rosales[4] divide input space into many simple areas, here every small region has the corresponding mapping function, and used a feedback mechanism to reconstruct the matching pose, because the areas of training data is small and mapping functions have better fitting effect, so this method can greatly improve the accuracy. Although learning-based methods perform faster, does not need special initialization, it has a small storage costs, and need not save sample database, the

* Corresponding author e-mail: baofufang@163.com

scale of the training sample have a big effect to learning-based results. Samples-based method must first establish template library, and store reams of features and training samples which poses are already known. When the test image is given, samples-based method first extract corresponding features and compare them with the sample in template library with some measure method, which is equivalent to search for sample that is similar with observation data, then use k-nearest-neighbors or any other method to estimate human body pose. The human pose is very complex, image features which are projected from different human pose may be very similar with each other, it means that the relationship between image features and human pose is 1:N. For example, Nicholas R [5]search out several candidate sample from the template database, then select the best candidate as the finally match with time domain similarity constraints; Alexei A.Efros[6] calculate the human movement features with light flow, and according to the motion sequence match ,search out the nearest neighbor human pose sequence from motion sample library, this method can estimate human motion pose successfully from a distance low resolution video. Samples-based method must have cover all possible human pose sample, but because the human pose is too complex, limited sample is difficult to contain the whole body posture space, samples-based only can be applied to the specific pose estimation.

Model-based method divides human body into several connected components, using pictorial model to represent the whole human body. And use pictorial inference method to estimate human pose, i.e. use a priori model of human body in the process of human pose estimation, and the parameters of the model update with the change of current state. Model-based method is mainly divided into three parts, respectively is pictorial model, optimization algorithm and part appearance model. Pictorial model is used to express the constraint relationship between the parts. One of most commonly used is tree model [7,8,9, 10,11], which is defined according to the connection relations between parts and is the most intuitive. Sam Johnson [11] puts forward the attitude spatial clustering to different category in order to overcome the shortage which can't capture multiple model-based human body characteristics (such as the head of human may be in the front, side or rear of picture). It gives a mixed structure of pictorial model which a model corresponds to a category. In a more complex LSP database [17] it gives better performance. Xiao Feng [12] introduces constraint of non-tree probabilistic graphical model on the basis of pictorial model which improves a certain precision. Vivek Kumar Singh [13] puts forward method that make more precise pose estimation according to the tree structure of the object's context, which is on the base of the prior knowledge of other objects (such as football, etc.).

Inference algorithm estimates the final pose according to pictorial model and observation model. Belief Propagation is a common pictorial model optimization algorithm. However, in the problem of human pose estimation, it's not practical to use BP algorithm directly because of state vector of components has high dimension. Deva Ramanan [8] uses Sum-product algorithm which inherits the message passing mechanism [7]. By introducing the factor graph, the global probability density function is divided into the product of several local probability density functions. It expands the scope of algorithm's usage to undirected graph (such as CRF (Conditional Random Fields)). But it still has a limit that only in the factor graph without rings can guarantee the algorithm's convergence.

Part appearance model represent body parts, used to measure each part of the image likelihood, determine the location of the components. Marginal position characteristic of the literature [1] regards gradient image as a full field with "electricity" particle, and there are forces between particles. Pedro F.Felzenszwalb [7] defined the image likelihood of parts with the results of background subtraction, and is suitable for the video sequence which background is simple, in which the components based on the edge of the template. Deva Ramanan,Vittorio Ferrari [8,9] adopts color information templates and edge template, and it can be applied to video and a separate picture. Literature [8] puts forward a new simple parts observation model. This hybrid model is a structure of graphical model without direction (only close to the horizontal and vertical body). The reason why can do this is because the mixture model can extract background characteristics statistics on particular direction.

In recent years some successful methods are based on the pictorial model [1,7,8,9,10,11], and in the paper we use the same tree pictorial model. In allusion to the problem that human pose estimation algorithm which is based on tree pictorial model is easily confused with background, and usually people's clothing compose of homogeneous texture regions, there is great difference between clothes texture and background texture [14]. So in this paper, we present a new appearance model based on the color and texture.

In our method, First, we use human body detector to detect[16,18] the general location in static image, and expand the human body detection window so that can be enlarged to include all parts of the body. Secondly, Grub cut segmentation algorithm is used in the expanding detecting window to reduce the search space, extract the edge feature in foreground and match with variable edge model to establish the general location of body's all parts. Through the analysis of these pictures for position information and pictures color texture characteristics and edge character set up the regional model and background model of all parts of body. Use sum-product algorithm to infer body posture, and then repeat the analysis process to obtain the final result.

3 Iterative human pose estimation based on tree-structured model

Here, human's body parts are tied together in tree-structured conditional random field [17], typically , parts l_i are rectangular image patches(it is assumed that the rectangular is fixed-size) and their position is parametrized by location (x_i, y_i),orientation θ_i.The edge-based deformable model can be written as a log-linear model:

$$P(L|I) \propto EXP((\sum_{i,j\in E} \varphi(l_i - l_j) + \sum_i \phi(l_i)) \quad (1)$$

The pairwise potential$\varphi(l_i - l_j)$ corresponds to a prior on the relative position of parts,E is a tree,here we parameterize $\varphi(l_i - l_j)$ with discrete binning, as in formula (2), bin() is for the vectorized count of spatial and angular histogram bins,α_i is a model parameter that favors certain spatial and angular bins for part i with respect to its parent. $\phi(l_i)$ corresponds to the local image evidence for a part, in formula (3)$f_i(I(l_i))$ is for feature vector extracted from the oriented image patch at location l_i and β_i is the weight information of edge-based deformable model.

$$\varphi(l_i - l_j) = \alpha_i^T bin(l_i, l_j) \quad (2)$$

$$\phi(l_i) = \beta_i^T f_i(I(l_i)) \quad (3)$$

The machinery used here for inference is sum-product algorithm. Since E is a tree, sum-product algorithm is that first compute upstream messages from part i to its parent j,with the formula (4) and (5).Then starting from the root, we compute messages downstream from part j to part i with formula (6) and find the positions that maximize $P(L|I)$ as the results.

$$m_i(l_j) \propto \sum_{l_i} \varphi(l_i - l_j) a_i(l_i) \quad (4)$$

$$a_i(l_i) \propto \phi(l_i) \prod_{k\in kids_i} m_k(l_i) \quad (5)$$

$$P(l_i|I) \propto a_i(l_i) \sum_{l_j} \varphi(l_i - l_j) P(l_j|I) \quad (6)$$

Iterative human pose estimation based on tree-structure model, firstly match the extracted edge information (fig.1(b)) with edge-based deformable model to get the general location of each parts of body, and establish each part appearance model and background model(color model), And then create a region-based model to find other body parts (as shown in fig.1(c)). The new model is used to get human body position and establish new region deformable model (as shown in fig.1(d)), the inference algorithm give the result of pose estimation, this process is repeated (fig.1(e)). This repetitive process is very sensitive on initial process, which is based on the edge of the model initialization and in the first iteration establishing region deformable model is critical.

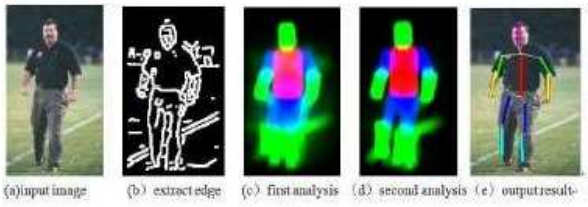

Fig. 1: The Process of Iterative Human Pose Estimation.

Vittorio Ferrari[9]first run a generic human detector to find the approximate location and scale(x, y, s) of the person ,second take a segmentation algorithm in person area to extract the human foreground area, and then run iterative human pose estimation based on tree structured model.

The result of the experiment indicated that human pose estimation algorithm which is based on tree pictorial model is easily confused with background, and usually people's clothing compose of homogeneous texture regions. There is great difference between clothes texture and background texture. So in this paper, we put forward a new appearance model based on the color and texture (AMBCT).

4 Iterative human pose estimation based on color and texture part appearance model(AMBCT)

In this section, because of the shortage of iterative human pose estimation based on tree structured model and Vittorio Ferrari's algorithm, we propose an improved part appearance model based on color and texture to estimate the human pose, i.e. we build a region model and region-based deformable model with color feature and texture feature, we also find that the segmentation algorithm is not always essential in the iterative human pose estimation, on the contrary sometimes without the grab cut segmentation algorithm based on color and texture appearance model(NG-AMBCT) give a better result.

4.1 Human Detector.

First we use a histogram of gradient(HOG) human detector [16] to find the general position of the human, enlarge the detect result so that form a region can always cover the human body(fig.2(a),fig.2(b)).Briefly, the HOG

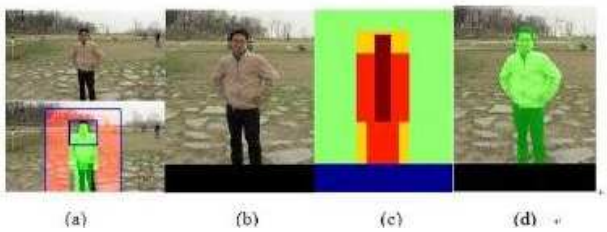

Fig. 2: Human detect and Grab cut segmentation.

human detect method tiles the detector window with a dense grid of cells, with each cell containing a local histogram over orientation bins. At each pixel, the image gradient vector is calculated and converted to an angle, voting into the corresponding orientation bin with a vote weighted by the gradient magnitude. Votes are accumulated over the pixels of each cell. The cells are grouped into blocks and a robust normalization process is run on each block to provide strong illumination invariance. The normalized histograms of all of the blocks are concatenated to give the window-level visual descriptor vector for learning and detecting.

4.2 Grab cut.

Second we extract the foreground in the enlarged region in section 4.1 with grab cut segmentation algorithm. Grab cut is a kind of interactive color image segmentation algorithm with a high accuracy of segmentation (such as figure 2 (d)) . When the human and the human enlarged region is successfully built, according to the prior knowledge that the head generally in the middle upper-half of the detection window, and the torso is directly underneath it, so we can divide the human enlarged region(such as figure2(c)) into four region for grab cut's initialization. In figure 2(c)red area contains mostly foreground, but some background as well; green area covers mostly background but it might also include part of the arms; deep purple red area is known to belong to the human; yellow region is neutral, i.e. probabilities of background and foreground are equal.

4.3 Build Region Model.

After highlighting the foreground in section 4.2, we extract edge feature with canny edge detection algorithm in the region of foreground (as figure 3 (b)), and match edge feature with the edge-based deformable model in section 3 of this paper, preliminary build general region of each body parts, then we should build region model in these general region. As in paper [8], we define a parse to be a soft labeling of region type (torso, left lower arm, etc.) and use the initial parse to build a region model, learn foreground and background color and texture models. We also exploit symmetry in appearance so that we can learn a single color and texture model for left/right limb pairs. Then label each pixel with the color and texture model, one parse process is end(as figure 3(c)). Later, we use these masks as features for a deformable model that re-computes $P(L|I)$, begin next analysis process and the procedure is repeated. For example, we can use the edge deformable model and edge information in section 4.2 to define a soft labeling for the image into head/non-head pixels, we can give the position and orientation of head by repeatedly sampling $P(l_i|I)$.Let the rendered appearance for part i be an image patch s_i,in the limit of infinite samples, one will obtain an image(eq.7), i.e. a parse for part i,it is computed by convolving $P(l_i|I)$ with rotated versions of patch s_i.

$$p_i(x,y) = \sum_{x_i,y_i,\theta_i} P(x_i,y_i,\theta_i|I)s_i^{\theta_i} \tag{7}$$

Given the parse image p_i,we learn a color and texture model for part i and color and texture model of background of p_i.Then we can use the part-specific color and texture models to label each pixels as foreground or background with a likelihood ratio test(eq.(8) and eq.(9)),In formula (8) and formula (9),k is the threshold that label a pixel is foreground or background. As the paper [9], the color appearance model used here are color histograms over the RGB cube discretized into 16*16*16 bins, each bin as a color value. To capture texture, we extract features from co-occurrence matrices. Co-occurrence matrices represent second order texture information- i.e., the joint probability distribution of gray-level pairs of neighboring pixel in a parse image region. In order to enhance extraction time of texture and meanwhile we can also describe the texture feature well, so here we use 4 descriptors which is different with the 12 descriptors in the paper [14], they are angular second-moment, contrast, correlation, inverse difference moment. Co-occurrence features are useful in part appearance model since they provide information regarding homogeneity and directionality of patches. In general, a person wears clothing composed of homogeneous textured regions and there is a significant difference between the regularity of clothing texture and background textures. For each color band, we create four co- occurrence matrices, one for each of the $(0°, 45°, 90°, 135°)$ directions, The displacement considered is 3 pixel.

$$P(fg_i(k)) = \sum_{x,y} p_i(x,y)\delta(im(x,y) = k) \tag{8}$$

$$P(bg_i(k)) = \sum_{x,y} (1 - p_i(x,y))\delta(im(x,y) = k) \tag{9}$$

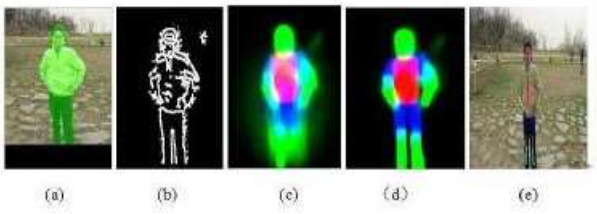

Fig. 3: AMBCT Iterative human pose estimation.

Iterative human pose estimation based on color and texture appearance model(AMBCT)

Input Test Image
Output Human Pose

(1) Reduce search space.
① Use body detector to detector the position of human body in the picture;
② On the base of ①, expand detecting window to contain the entire human body. According to the prior knowledge （Human's head is generally at the top of detection window, while the torso is always located under the head.）, so it can be initially divided into foreground and background, with which we can effectively carry out Grub cut initialization. Then Grub cut segmentation processes;
(2) Initial body part position.
① Build model based on edge. According to the characteristics of human body structural model , edge deformable model can be expressed as formula (1) log-linear regression model and we can use conditional random fields study to get edge variable model;
② Extract the edge characteristics of foreground area in (1), and match them with edge variable model to get the estimation of body part's initial position （Rough order of magnitude estimation);
(3) Build region model.
After building body's initial position, we need analyses the initial position of each component , build components' color and texture model (including background). Then use new model to mark pixels in pictures, and according to the component's posterior probability give the position after analysis, namely establish a regional model;
(4) New region-based deformable model.
① After step (3), we have built component's regional model. By re-learning the color and texture models in these areas, we re-estimate the body posture in the picture;
② repeat process ① in step (4), create new models and estimate. When it achieves convergence, output the results.

Fig. 4: The Detail Process of AMBCT Iterative human pose estimation

4.4 Build region-based deformable model.

After section 4.3, we have built an initial region color and texture model for each part and its background. We use these models to construct binary label images for part i:$P(fg_i(im) > P(bg_i(im)))$. The oriented patch features extracted from these label images can be written as f_i^r.We learn model parameters for a region-based deformable model by CRF parameter estimation, as in [8,15]. Given the parse from the region-based model, we can re-learn a color and texture model for each part and the background (and re-parse given the new models, and repeat). Experience show that both the parses and the color and texture models empirically converge after 1-2 iterations (fig. 3(d)).

5 Experiments

Given an image, through the parsing procedure that return a distribution over poses $P(L|I)$, select a high probability as a result and all other poses have a low value. We have tested our iterative human pose estimation on two datasets: PARSE and LSP, which have 305 images and 2000 images respectively.

Table 1: Results of human pose estimation

Accuracy of Estimation	PARSE(305 images)	LSP(2000 images)
AMBCT	22.3%	16.7%
Deva Ramanan[8]	17.2%	13.1%
Vittorio Ferrari[9]	18.6%	13.6%
NG-AMBCT	21.7%	15.9%

Table 1 show the comparison of four kind of algorithms of human pose estimation, they are AMBCT(our algorithm) , Iterative human pose estimation based on tree-structured model(Deva Ramanan [8]), Vittorio Ferrari [9],and no grab cut segmentation in AMBCT algorithm(NG-AMBCT). We can find that our algorithm have a high accuracy than other three algorithms. We also find that the same algorithm in PARSE dataset the accuracy of estimation is obviously high than LSP dataset, it mainly because that LSP dataset is much complex than PARSE dataset. Even so it reflects our algorithm have an advantage compare with other three algorithm.

In addition, we can find that NG-AMBCT and AMBCT have a close accuracy of estimation through table 1, sometimes NG-AMBCT give a better performance than AMBCT, sometimes on the contrary. For example, figure 5(a)(c) are the results of NG-AMBCT, which give the accurate results, but figure 5(b)(d) output the wrong result with AMBCT; figure 6 and figure 3 show that AMBCT result is better than NG-AMBCT. By our experiment, we assume that AMBCT give a better performance when the background and foreground information are relatively close to each other, and NG-AMBCT is better when the background and foreground have a obvious difference. How to weigh the use of segmentation algorithm, it is determined to our further research and experiment.

The images in Figure 7 are the result of our experiment in PARSE and LSP datasets, it shows that our method successfully mark the human pose.

Figure 8 is the comparison of AMBCT and literature [9], figure (a)(c) with the method in literature [9] give wrong results, when build part appearance model based on color and texture in AMBCT in figure 8(b)(d) gives accurate results. It proves that our algorithm to some degrees can solve the problem of methods in literatures [8] and [9] are easily interfered by the background.

The images in the Figure 5.5 are in the LSP dataset and cut from the real world. We find that the estimation usually fail in the environment with a complex background or posture or occlusion phenomenon (figure 9 (a)(c)(e)), in these complex cases human pose estimation is still hard now.

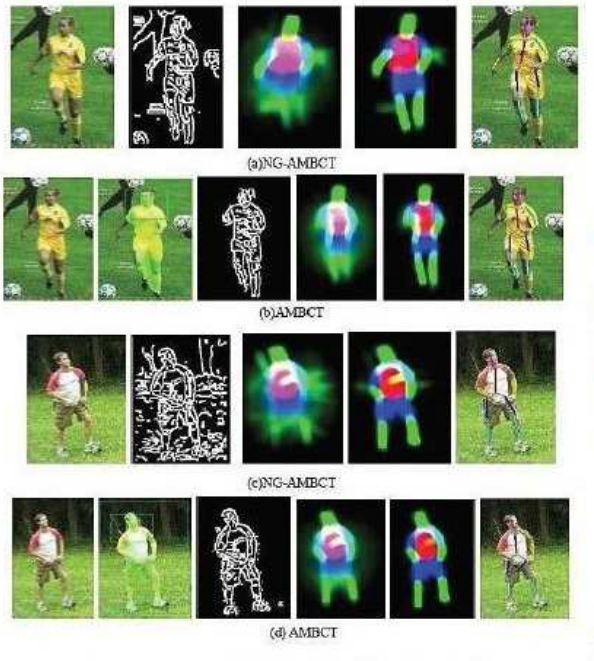

Fig. 5: Comparison of NG-AMBCT and AMBCT.

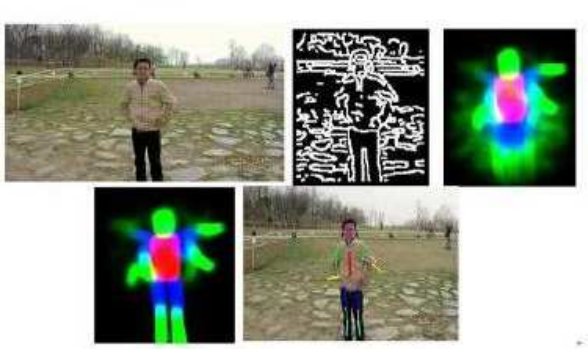

Fig. 6: Result of NG-AMBCT.

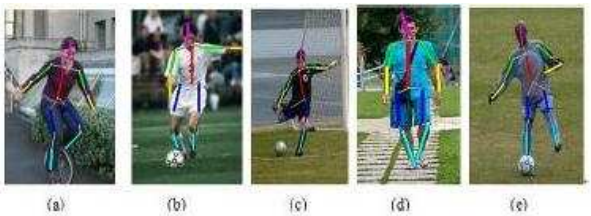

Fig. 7: Some estimation results in PARSE and LSP datasets.

6 Conclusion

For iterative human pose estimation base on tree-structured model is easily interfered by the background, because of there is a significant difference between the regularity of clothing texture and background textures and a person wears clothing composed of homogeneous, we propose a iterative human pose estimation algorithm with a new part appearance model based on color and texture(AMBCT), with human detector and Grab cut segmentation to reduce search space, build color and texture region model by edge-based deformable model, iterative infer the final human pose, experiment show that our method improve the accuracy of estimation of human pose.

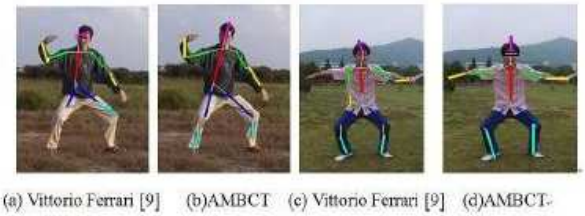

Fig. 8: Comparison of method in literature [9] and AMBCT.

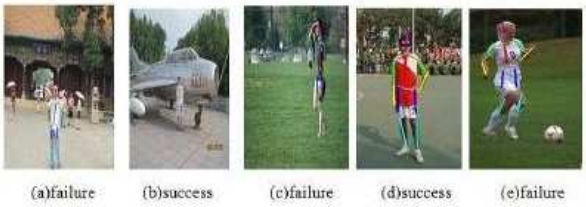

Fig. 9: other results.

Acknowledgement

The authors acknowledge the financial support Natural Science Foundation of China, project No. 61070131 The author is grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] Su Yan-chao , AI Hai-zhou , Lao Shi-hong . Part Detector Based Human Pose Estimation in Images and Videos, Journal of Electronics & Information Technology, **33**, 1413-1419 (2011).

[2] Sun Yun-da. Research on Touch-free Human Body Motion Capture Under Multiple View points, Signal and Information Processing of Beijing Jiaotong University, (2006).

[3] Ankur Agarwal,Bill Triggs.3D human pose from silhouettes by relevance vector regression, Computer Vision and Pattern Recognition(CVPR), **2**, 882-888 (2004).

[4] Romer Rosales, assilis Athitsos, Leonid Sigal, et al. 3D hand pose reconstruction using specialized mappings, ICCV, **1**, 378-385 (2011).
[5] Nicholas R. Howe. Silhouette Lookup for Automatic Pose Tracking, Computer Vision and Pattern Recognition Workshop, 15-22 (2004).
[6] Alexei A.Efros, lexander C. Berg, Greg Mori,et al. Recognizing action at a distance, Computer Vision, **2**, 726-733 (2003).
[7] Pedro F. Felzenszwalb, Daniel P. Huttenlocher. Pictorial Structures for Object Recognition, International Journal of Computer Vision, **61**, 55-79 (2005).
[8] Deva Ramanan. Learning to parse images of articulated bodies, Neural Information Processing Systems (NIPS), 1129-1136 (2006).
[9] Vittorio Ferrari, Manuel Marin-Jimenez, Andrew Zisserman.Progressive search space reduction for human pose estimation, Computer Vision and Pattern Recognition(CVPR), Anchorage :IEEE Computer Society, 1-8 (2008).
[10] Yi Yang,Deva Ramanan. Articulated pose estimation with flexible mixtures-of-parts, Computer Vision and Pattern Recognition (CVPR) .RI: IEEE Computer Society, 1385-1392 (2011).
[11] Sam Johnson, Mark Everingham. Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation, British Machine Vision Conference (BMVC), 1-8 (2010).
[12] Xiao Feng, Zhou Jie. Human pose estimation in static images based on region segmentation and Monte Carlo sampling, CAAI Transactions on Intelligent Systems, **6**, 38-43 (2011).
[13] Vivek Kumar Singh, Furqan Muhammad Khan,Ram Nevatia. Multiple Pose Context Trees for estimating Human Pose in Object Context, Computer Vision and Pattern Recognition Workshops (CVPRW), 17-24 (2010).
[14] William Robson Schwartz,Aniruddha Kembhavi,David Harwood,et al.Human Detection Using Partial Least Squares Analysis, International Conference on Computer Vision (ICCV), 24-31 (2009).
[15] Deva Ramanan,Cristian Sminchisescu.Training Deformable Models for Localization, Computer Vision and Pattern Recognition(CVPR), **1**, 206-213 (2006).
[16] N. Dalai, B. Triggs, I. Rhone-Alps, et al. Histograms of oriented gradients for human detection, Computer Vision and Pattern Recognition(CVPR), **1**, 886-893 (2005).
[17] Notes: http://www.comp.leeds.ac.uk/mat4saj/lsp.html.
[18] Meng Fan-hui, Wang Hao, Fang Bao-fu, et al. Research and Implementation of Human Detection based on Extended Histograms of Oriented Gradients, Journal of Guangxi Normal University, **29**, 168-172 (2011).
[19] Navneet Dalal, Bill Triggs, Cordelia Schmid.Human Detection Using Oriented Histograms of Flow and Appearance, European Conference on Computer Vision (ECCV). Graz, Austria: Springer, **3952**, 428-441 (2006).
[20] Ankur Agarwal, Bill Triggs. 3D human pose from silhouettes by relevance vector regression, Computer Vision and Pattern Recognition (CVPR), **2**, 882-888 (2004).
[21] Marcin Eichner,Vittorio Ferrari. Better appearance models for pictorial structures, British Machine Vision Conference (BMVC), (2009).
[22] Kyoung-Sic Cho, In-Ho Choi, and Yong-Guk Kim. Robust Facial Expression Recognition Using a Smartphone Working against Illumination Variation, Applied Mathematics & Information Sciences, **6**, 53-59 (2012).
[23] Carsten Rother,Vladimir Kolmogorov,Andrew Blake."Grab Cut":interactive foreground extraction using iterated graph cuts,ACM Transactions on Graphics (TOG), **23**, 309-314 (2004).

Wang Hao received the MS degree in Computer science from Hefei University of Technology in 1989, and the PhD degree in Computer science from Hefei University of Technology in 1997. He is currently a professor in Hefei University of Technology. His main research interests artificial intelligence.

Meng Fanhui received the MS degree in School of Computer and Information from Hefei University of Technology in 2012,His research interests are in the areas of artificial intelligence and computer vision.

Fang Baofu is the Corresponding author of this article. He received the MS degree in Computer software and theory from Hefei University of Technology in 2003, and will have the PhD degree in Computer application from Harbin Institute of Technology in 2012.He is associate professor and master's supervisor of Hefei University of technology. In 2011, He have won the Science and Technology Award for Youth in the national Artificial Intelligence Youth Forum held by Chinese Association for Artificial Intelligence. And he has been the academic advisor of RoboCup simulation 3D Team named HfutEngine which got the third Place of RoboCup2010 and the second place in RobocupChina2009. He has published more than 50 research articles in Robotics and artificial intelligence.