HUMAN BODY POSTURE REFINEMENT BY NONPARAMETRIC BELIEF PROPAGATION

Ruixuan Wang and Wee Kheng Leow

School of Computing, National University of Singapore,
3 Science Drive 2, Singapore 117543, Singapore.
{wangruix, leowwk}@comp.nus.edu.sg

ABSTRACT

Accurate human body posture refinement from single or multiple images is essential in many applications, such as vision-based sport coaching and physical rehabilitation. Two main causes of difficulty to solve the refinement problem are high degree freedom of human body and self-occlusion. One of the most recent algorithms is nonparametric belief propagation (NBP) that solves the problem in a lower dimensional state space. However, it is difficult to handle self-occlusion. This paper presents an NBP-based algorithm that can refine body posture even in self-occlusion case, which has been shown by experimental results. The experimental results also show that our algorithm can accurately refine body posture even if the initial posture has large difference from the true posture.

1. INTRODUCTION

Human body posture refinement tries to accurately recover human posture from single or multiple images, given one or multiple initial postures. Accurately recovering posture is essential in such applications as vision-based sport coaching and physical rehabilitation. In this paper, we present an algorithm to refine articulated human body posture from images.

Posture refinement is an optimization problem. Three kinds of search strategies are often used to solve the problem: local descent method, multiple random start, and sampling method including NBP. Local descent method [1, 2] can be used to incrementally update an existing posture estimate, e.g., using the gradient to guide the search direction toward a local optimal posture, but it cannot guarantee globally optimal. Multiple random start [3, 4] performs local descent method from each of multiple initial postures to generate multiple local optimality, and then selects the local optimal posture with the best similarity measurement as the solution. This method works better, although it cannot guarantee to find the best posture. Another search strategy is sampling method [5, 6], which generates a large number of samples in the posture state space, and selects one with the most similarity between the sample and the image observation as the globally optimal. Although densely sampling the entire state space can guarantee a good solution, it is infeasible to densely sample from even a local region of the high (e.g., 30) dimensional posture state space.

Instead of directly recovering whole body posture, NBP [7, 8, 9] divides human body into several body parts and recovers the low dimensional (e.g., 6) pose of each body part by considering the relationships between every two adjacent body parts. It represents human body and the relationships between body parts by a graphical model. Every body part is encoded by one node in the graph, and every edge connecting two nodes indicates that there are relationships between the two nodes. However, the original NBP [9] cannot deal with self-occlusion. In the original NBP, each node (or body part) has its own observation function that is estimated by the similarity measurement between state of the body part pose and the corresponding image part. If the body part is partially or fully occluded by other body parts, the observation function cannot be correctly estimated by the similarity measurement. Although the observation function can be learned from training images that include the case of self-occlusion [10], learning is often a complex process and not easy to collect training images. What is more, such learned observation function is used not to refine but to initialize body postures, and it is limited to initialize a small set of postures.

In this paper, one modified NBP algorithm is presented that can deal with the self-occlusion of body parts, without any learning process. Based on a graphical model(Section 2), the modified NBP algorithm is designed (Section 3). Initial experimental results show that the algorithm can refine body posture even in the case of self-occlusion between body parts. The results also show that, even if the initial posture is largely different from the true posture, accurate posture can be recovered.

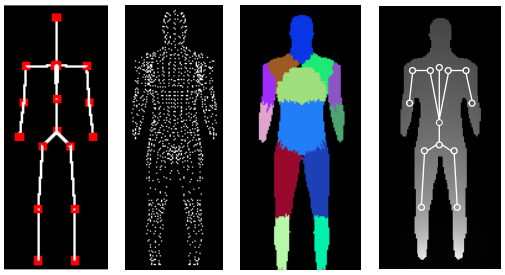
2. ARTICULATED HUMAN BODY MODEL

A human skeleton model (Figure 1(a)) is used to represent body joints and bones, and a triangular mesh model (Figure 1(b)) is used to represent the body shape. Each vertex in the mesh is attached to the related body bone, and each body part consists of one or multiple bones and the corresponding vertexes (Figure 1(c)). For each body part's size, assuming there is a fixed ratio between width and thickness, two parameters (length and width) can be used to represent the size of each body part.

Human body posture $\mathcal{X}$ is represented by a set of body parts' poses, $\mathcal{X} = \{\mathbf{x}_i | i \in \mathcal{V}\}$, where $\mathcal{V}$ is the set of body parts, and body part pose $\mathbf{x}_i = (\mathbf{p}_i, \boldsymbol{\theta}_i)$ represent the i^{th} body part's 3D position $\mathbf{p}_i$ and 3D orientation $\boldsymbol{\theta}_i$. Given the shape and size of human body, any body posture $\mathcal{X}$ can be rendered (by rotating body bones and the corresponding mesh vertexes) and projected to generate a synthetic image observation. During posture refinement, each synthetic observation will be used to compared with a real image observation $\mathcal{Z} = \{\mathbf{z}_i | i \in \mathcal{V}\}$, where $\mathbf{z}_i$ represents the image observation for the i^{th} body part. The relationship between $\mathbf{x}_i$ and $\mathbf{z}_i$ is represented by the observation function $\phi_i(\mathbf{x}_i, \mathbf{z}_i)$. In addition due to the articulation, every pair of adjacent body parts $\mathbf{x}_i$ and $\mathbf{x}_j$ must be connected. This constraint is enforced by the potential function $\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$.

A tree-structured graphical model (Figure 1(d)) is used to represent the articulated human body model. The tree consists of a

set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$. Each node $i \in \mathcal{V}$ is associated with $\mathbf{x}_i$ and $\mathbf{z}_i$ of i^{th} body part, and each edge $(i,j) \in \mathcal{E}$ is associated with the potential function $\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$.



(a) (b) (c) (d)

Fig. 1. Human body model. (a) Human skeleton model. (b) The vertexes in the body mesh model. (c) Each vertex and triangle in the mesh model is assigned to one specific body part. (d) A tree-structured graphical model for representing human body model.

3. POSTURE REFINEMENT ALGORITHM

NBP can be used to estimate each body part's pose. However, it assumes that observation of each part can be obtained independently [7, 9]. This limits it to cases where there is no self-occlusion. We modified NBP to handle occlusion by changing the joint probability of body posture $\mathcal{X}$ and corresponding image observation $\mathcal{Z}$ to Equation (1). Similar to NBP [7], we may calculate marginal distributions by Equations (2) and (3),

$$p(\mathcal{X}, \mathcal{Z}) = \alpha_1 \prod_{(i,j)\in\mathcal{E}} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) \prod_{i\in\mathcal{V}} \phi_i(\mathcal{X}, \mathbf{z}_i) \quad (1)$$

$$m_{ij}^n(\mathbf{x}_j) \approx \alpha_2 \int_{\mathbf{x}_i} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)\phi_i(\mathbf{x}_i, \tilde{\mathcal{X}}_{-i}^{n-1}, \mathbf{z}_i) \times \prod_{k\in\Gamma(i)\backslash j} m_{ki}^{n-1}(\mathbf{x}_i) d\mathbf{x}_i \quad (2)$$

$$\hat{p}^n(\mathbf{x}_j|\mathcal{Z}) \approx \alpha_3 \phi_j(\mathbf{x}_j, \tilde{\mathcal{X}}_{-j}^{n-1}, \mathbf{z}_j) \prod_{i\in\Gamma(j)} m_{ij}^n(\mathbf{x}_j) \quad (3)$$

where $m_{ij}^n(\mathbf{x}_j)$ is the message propagated from node i to j in iteration n. $\Gamma(i) = \{k|(i,k) \in \mathcal{E}\}$ is the neighbor of node i, and $\Gamma(i)\backslash j$ is the neighbor of i except j. $\tilde{\mathcal{X}}_{-i}^{n-1}$ is the set of body parts' estimations except the i^{th} body part which come from the previous $(n-1)^{th}$ iteration (or from initial estimates when n is 1).

When body part i is partially occluded by some others, its image observation $\mathbf{z}_i$ is generated by both this part and the others. Together with the other body parts' estimations $\tilde{\mathcal{X}}_{-i}^{n-1}$ coming from previous iteration, each estimate of $\mathbf{x}_i$ can generate corresponding observations to measure the observation functions, although the convergence of Equation (3) is remained to be proved.

3.1. Potential Functions

Potential function $\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$ can be used to enforce relationships between body part i and j, such as kinematic constraint and joint angle limits. In our work, $\psi_{ij}(\mathbf{x}_i, \mathbf{x}_j)$ is used to enforce just kinematic constraint between two adjacent body parts, i.e. the corresponding two ends of the two body parts should be at the same 3D position (i.e. the position of body joint connecting them) without noise. Therefore in general, assuming node i is the parent of j, potential function can be a single Gaussian with a zero mean vector, i.e.

$$\psi_{ij}^n(\mathbf{x}_i, \mathbf{x}_j) = \mathcal{N}(T(\mathbf{x}_i) - \mathbf{p}_j; 0, \Lambda_{ij}^n) \quad (4)$$

where $\psi_{ij}^n(\mathbf{x}_i, \mathbf{x}_j)$ represents the probability of $\mathbf{x}_j$ given $\mathbf{x}_i$. T is a rigid transformation that is obtained from the the pose $\mathbf{x}_i$ and the size information of the i^{th} body part. $\Lambda_{i,j}^n$ is the variance matrix of the gaussian function $\mathcal{N}$ in the n^{th} iteration of NBP. Note that $\Lambda_{i,j}^n$ may be different in different iterations. Here $\Lambda_{i,j}^n$ is gradually decreasing by $\Lambda_{i,j}^n = \lambda\Lambda_{i,j}^{n-1}$, where λ is a decreasing factor between 0 and 1. It is easy to get $\psi_{ij}^n(\mathbf{x}_i, \mathbf{x}_j) = \{\psi_{ij}^{n-1}(\mathbf{x}_i, \mathbf{x}_j)\}^{1/\lambda}$, which makes λ have similar effect to simulated annealing in annealed particle filtering [6].

Similarly, when i is a child of node j, there is

$$\psi_{ij}^n(\mathbf{x}_i, \mathbf{x}_j) = \mathcal{N}(T'(\mathbf{x}_j) - \mathbf{p}_i; 0, \Lambda_{ij}^n) \quad (5)$$

3.2. Observation Functions

Observation function $\phi_i(\mathbf{x}_i, \tilde{\mathcal{X}}_{-i}^{n-1}, \mathbf{z}_i)$ measures the likelihood of $\mathbf{z}_i$ given $\mathbf{x}_i$. In order to measure the likelihood, each estimate of body part pose $\mathbf{x}_i$ is required to be rendered and then projected together with $\tilde{\mathcal{X}}_{-i}^{n-1}$, and then the similarity between the projected image and the input image observation $\mathcal{Z}$ is computed to estimate the likelihood. Since $\phi_i(\mathbf{x}_i, \tilde{\mathcal{X}}_{-i}^{n-1}, \mathbf{z}_i)$ is estimated by the similarity of the two whole images, it can deal with self-occlusion where one body part is partially occluded by others. In our work, edge and silhouette are used as the feature for the similarity measurement. Chamfer distance is used to measure the edge similarity, and overlapping area of the projected image and the human body image region in the input image is used to measure the silhouette similarity. The relative weight between edge and silhouette similarity is experimentally determined.

3.3. Nonparametric BP

After designing potential functions and observation functions, NBP can be used to search for body parts' states by iteratively updating each message and each marginal distribution. In our NBP algorithm, like BP Monte Carlo (BPMC) [11], each message $m_{ij}^n(\mathbf{x}_j)$ is represented by a set of K weighted samples,

$$m_{ij}^n(\mathbf{x}_j) = \{(\mathbf{s}_j^{(n,k)}, \omega_{ij}^{(n,k)})|1 \le k \le K\} \quad (6)$$

where $\mathbf{s}_j^{(n,k)}$ is the k^{th} sample of the j^{th} body part state in the n^{th} iteration and $\omega_{ij}^{(n,k)}$ is the weight of the sample. Correspondingly, the marginal distribution is represented by

$$\hat{p}^n(\mathbf{x}_j|\mathcal{Z}) = \{(\mathbf{s}_j^{(n,k)}, \pi_j^{(n,k)})|1 \le k \le K\} \quad (7)$$

where $\mathbf{s}_j^{(n,k)}$ is the same as that in Equation 6 and $\pi_j^{(n,k)}$ is the corresponding weight.

In each iteration, each message $m_{ij}^n(\mathbf{x}_j)$ and each marginal distribution $\hat{p}^n(\mathbf{x}_j|\mathcal{Z})$ are updated based on Equations (2) and (3). Since the messages and marginal distributions are nonparametric, the update is based on the Monte Carlo method. The update process is described in the following:

1. Modify potential functions by $\Lambda_{i,j}^{n+1} = \lambda \Lambda_{i,j}^{n}$.
2. Use importance sampling to generate new samples $\mathbf{s}_j^{(n+1,k)}$ from related marginal distributions of previous iteration. The related marginal distributions include the neighbors' and its own marginal distributions of previous iteration. The new samples are to be weighted respectively in the following two steps to represent corresponding messages and marginal distributions.
3. Update messages. For each new sample $\mathbf{s}_j^{(n+1,k)}$ and each neighboring node $i \in \Gamma(j)$, calculate the weight $\omega_{ij}^{(n+1,k)}$:

$$\omega_{i,j}^{(n+1,k)} = \sum_{k'=1}^{K} [\psi_{ij}(\mathbf{s}_i^{(n,k')}, \mathbf{s}_j^{(n+1,k)}) \frac{\pi_i^{(n,k')}}{\omega_{ij}^{(n,k')}}] \quad (8)$$

Equation (8) is the nonparametric version of Equation (9), which represents message in terms of marginal distribution [9].

$$m_{ij}^{n}(\mathbf{x}_j) = \alpha_2 \int_{\mathbf{x}_i} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) \frac{\hat{p}^{n-1}(\mathbf{x}_i|\mathcal{Z})}{m_{ji}^{n-1}(\mathbf{x}_i)} d\mathbf{x}_i \quad (9)$$

The updated messages will be used to update marginal distributions.

4. Based on the updated messages, each marginal distribution is updated. For each sample $\mathbf{s}_j^{(n+1,k)}$, calculate the weight $\pi_j^{(n+1,k)}$, where

$$\pi_j^{(n+1,k)} = \phi_j(\mathbf{s}_j^{(n+1,k)}, \tilde{\mathcal{X}}_{-j}^{n}, \mathbf{z}_j) \prod_{l \in \Gamma(j)} \omega_{lj}^{(n+1,k)} \quad (10)$$

Then $\pi_j^{(n+1,k)}$ is re-weighted because we use importance sampling to generate sample $\mathbf{s}_j^{(n+1,k)}$. The updated marginal distributions will be used to update messages in the next iteration.

Human body posture can be estimated from the set of marginal distributions. The mean of $\hat{p}^n(\mathbf{x}_j|\mathcal{Z})$ or the sample with the maximum weight in $\hat{p}^n(\mathbf{x}_j|\mathcal{Z})$ can be used to represent the estimation of j^{th} body part state.

There are several differences between our algorithm and others' NBPs. Sudderth et al. [7, 9] used Gaussian mixtures to represent messages and marginal distributions, and a complex Gibbs sampler is required to generate samples in each iteration. In our algorithm, like BPMC [11], it just uses a set of weighted samples to represent messages and marginal distributions, and uses importance sampling to generate samples. Compared to BPMC in which importance function comes from the same node's marginal distribution of previous iteration, and in which they re-weight messages by the importance function, our importance function comes from multiple marginal distributions, i.e. both the neighboring marginal distributions and the same node's marginal distribution of previous iteration. And we re-weight the marginal distributions, not messages, by the importance function. We believe that such re-weighting is more reasonable because the samples from importance sampling are essentially used for marginal distributions. In addition, BPMC is used for rigid object whereas our algorithm deals with articulated body. More important, compared to these existing algorithms [9, 11], our algorithm can deal with self-occlusion by using estimated postures of previous iteration. Furthermore, our algorithm embeds the simulated annealing idea into the algorithm by modifying potential functions in each iteration.

4. QUANTITATIVE EVALUATION

Quantitative evaluation of our NBP algorithm was performed with two tests. The first test evaluated the algorithm's ability to refine body posture in the self-occlusion case. The second test evaluated the accuracy of recovered body posture when the difference between initial posture and true posture is large.

In order to quantitatively evaluate our algorithm, we captured motion using Gypsy motion capture system and extracted 3D postures from the motion. Each posture was mapped to 54-DOF human skeleton model with mesh model for skin, and rendered using OpenGL to get input image. To obtain initial posture, we added some uniform random noise to joint angles of the true posture.

In the tests, our algorithm was used to refine posture from a single image, therefore the depth information cannot be accurately estimated. As a result, 2D joint position error $E_{2D} = \frac{1}{nh} \sum_{i=1}^{n} \|\hat{\mathbf{p}}_{2i} - \mathbf{p}_{2i}\|$ is computed to assess the performance of our algorithm. $\hat{\mathbf{p}}_{2i}$ and $\mathbf{p}_{2i}$ are the estimated and true 2D image position of the i^{th} body joint. h is the articulated body height and it is about 195 pixels in the tests. Note that each estimated joint position was obtained from the estimated pose of one corresponding body part.

4.1. Posture refinement in self-occlusion case

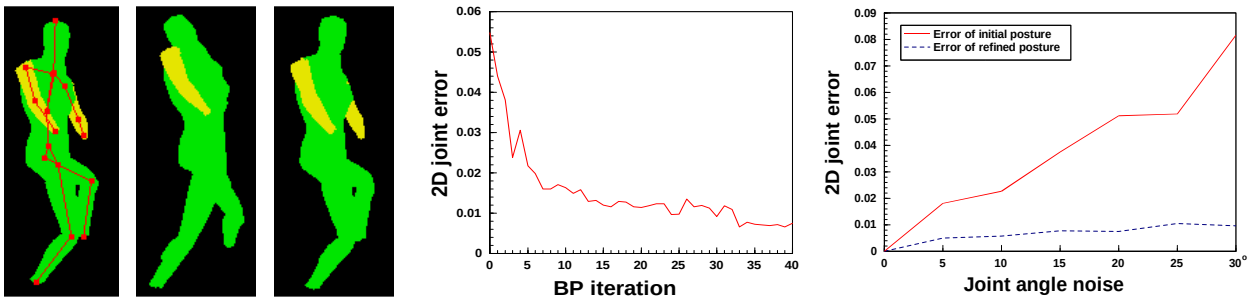
In this test, the performance of our NBP algorithm in refining self-occluded posture was evaluated. 150 weighted samples were used to represent each message and each marginal distribution respectively. Generally, the decreasing factor λ should be close to 1 (e.g., 0.95) according to the simulated annealing theory [6]. But this makes the algorithm very slowly to obtain the final solution. When λ is smaller (e.g., 0.5), it converges fast but may converge to a local optimal. For balance, here we used a two-level iteration framework, λ is set 0.6 and 5 iterations for the NBP are repeated 8 times (i.e., totally 40 iterations). Test results showed that such two-level iteration alternative resulted in similar good solutions to the original annealing method with λ close to 1, while obtaining faster convergence. For each iteration, around 20 seconds is spent, most of which is used to compute observation functions.

Figures 2(a) (b) and (c) illustrate one input image (with posture truth represented by the skeleton) and the corresponding images of an initial posture and the best refined posture. We can see that the estimated (projected 2D) posture is very close to the true posture. For this pair of input image and the initial posture, Figure 2(d) represents the 2D joint position error E_{2D} with respect to the iteration number. It shows that, after 15 iterations, the error has decreased to a relatively small value (i.e., 0.1% of body height h, or 2 pixels when h is 195 pixels). Note that the edge information plays an important role in posture refinement especially when arms fall inside the torso image region, in which case silhouette information itself often cannot get good refinement result.

Sudderth et al. [9] reported a similar result on refining body posture from initial posture estimates. However, their NBP algorithm required a complex learning process and was tested on a simple walking posture using a multi-camera system. In comparison, our algorithm does not require any learning process and can deal with more complex postures.

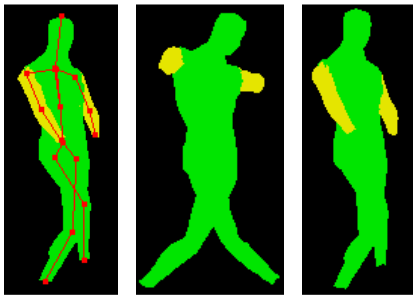
4.2. Posture refinement from different initial postures

In the second test, we tested the accuracy of the recovered body posture when the difference between initial posture and true pos-



(a) (b) (c) (d) (e)

Fig. 2. Test results. (a) Input image with the truth. (b) Image of initial posture. (c) Image of refined posture. (d) Error E_{2D} with respect to BP iteration. (e) Error E_{2D} with respect to joint angle noise. The estimated posture error in (e) is computed after 40 iterations of NBP.



(a) (b) (c)

Fig. 3. One result in the second test. (a) Input image with the truth. (b) Image of initial posture. (c) Image of refined posture.

ture increases. The uniform random noise of each joint angle was increased from 0^o to $[-30^o,\ 30^o]$. For each range of random noise, the mean of E_{2D} is averaged over a sequence of 20 images.

Figure 2(e) illustrates the mean errors of initial posture and estimated posture. It shows that when the joint angle noise is increased, the 2D error of estimated posture remains small even when the joint angle random noise is large (e.g., 30^o). Figure 3 illustrates one test result, from which we can see that the projection of the refined posture is very similar to the input image even if the initial posture is very different from the true posture.

5. CONCLUSION

This paper presented an algorithm to refine human body posture from single or multiple images. Compared to existing NBP algorithms, our algorithm can deal with self-occlusion and can accurately recover the projected 2D postures from single image, which has been shown by the test results. In the future work, tests on posture refinement from multi-view images and on real images will be performed. We also plan to apply the algorithm to articulated human body tracking and vision-based sport coaching applications.

6. REFERENCES

[1] C. Bregler and J. Malik, "Tracking people with twists and exponential maps," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, pp. 8-15, 1998.

[2] D.E. Difranco, T.J. Cham and J.M. Rehg, "Reconstruction of 3D figure motion from 2D correspondences," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 307-314, 2001.

[3] T.J. Cham and J.M. Rehg, "A multiple hypothesis approach to figure tracking," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 239-245, 1999.

[4] C. Sminchisescu and B. Triggs, "Covariance scaled sampling for monocular 3D body tracking," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 447-454, 2001.

[5] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," *Proc. European Conf. Computer Vision*, vol. 1, pp. 343-356, 1996.

[6] J. Deutscher, A. Blake and I. Reid, "Articulated body motion capture by annealed particle filtering," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 126-133, 2000.

[7] E.B. Sudderth, A.T. Ihler, W.T. Freeman and A.S. Willsky, "Nonparametric belief propagation," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 605-612, 2003.

[8] M. Isard, "PAMPAS: Real-valued graphical models for computer vision," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 613-620, 2003.

[9] E.B. Sudderth, M.I. Mandel, W.T. Freeman and A.S. Willsky, "Visual hand tracking using nonparametric belief propagation," *Proc. IEEE CVPR Workshop on Generative model based vision*, 2004.

[10] L. Sigal, S. Bhatia, S. Roth, M.J. Black and M. Isard, "Tracking loose-limbed people," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 421-428, 2004.

[11] G. Hua and Y. Wu, "Multi-scale visual tracking by sequential belief propagation," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 826-833, 2004.