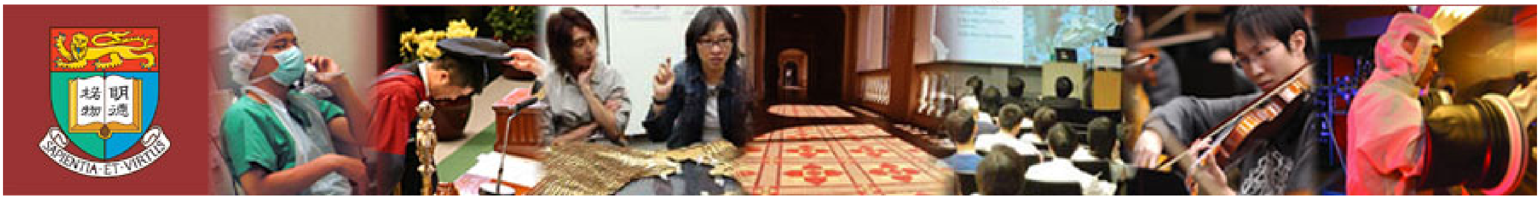


Arm Pose Modeling for Visual Surveillance

Chong Guo Li and N. H. C. Yung
Department of Electrical and Electronic Engineering, The University of Hong Kong
Pokfulam Road, Hong Kong, H.K.S.A.R., China

Abstract—*Arm pose of individual person often indicates his/her behavior and interaction in a crowd. This paper focuses on how to model arm pose with respect to walking styles in visual surveillance. We propose to use a Bayesian network to represent the relationship between the upper arm, forearm and hand. Two sets of probabilistic templates are used to estimate the likelihoods of these three arm parts against their boundary, foreground and skin color features. The prior distributions, a critical part of the Bayesian formulation, are estimated using a kernel density estimator based on the CMU motion dataset, whereby MAP is applied to estimate the part's orientation and size before fitting them to the arms. The proposed method is found to perform effectively on the CAVIAR sub-dataset and a HKU dataset, and compare well against another popular method. Potentially, it can be extended to track arm movements or to analyze the objects carried by individuals.*

Keywords: Arm pose modeling, Bayesian network, Maximum a posterior probability, Kernel density estimation

1. Introduction

Arm pose is the manifestation of a behavior. It could be a gentle swinging action by the side when walking, or a waving action to mean goodbye. According to Mehrabian [1], 93% of our communication is non-verbally expressed by our body. Moving our arms and hands not only interacts with the environment but also conveys meaningful information [2]. Therefore, arm pose detection is the first step towards understanding human behavior individually or as a whole in a crowd. The next is to derive positions, orientations and sizes of the upper arms, forearms and hands for action classification later on.

The goal of this research is to model arm pose in one step that includes detection and the estimation of positions, orientations and sizes of the various arm parts. However, arm pose modeling is challenging for a number of reasons. Firstly, arms can be bare or covered by sleeves. A bare arm is typically dominated by skin color, whereas sleeved arm may have different appearance due to the length of the sleeves. Secondly, sleeved arm is dominated by the color and texture of the sleeve, which may be similar to the person's body color, but there could be exceptions as well. Thirdly, arm normally exhibits complex degree of freedom. Fourthly, arm may be partially occluded, i.e., behind the body or inside a pocket. We propose in principle to adopt a graphic model to describe the relationship between the parameters of the arm model. Then the likelihoods of arm parts against the boundary, foreground and skin color features of the arm are combined with the learnt prior probability of a specific action (e.g. walking), whereby the arm pose is modeled by the maximum a posterior probability (MAP).

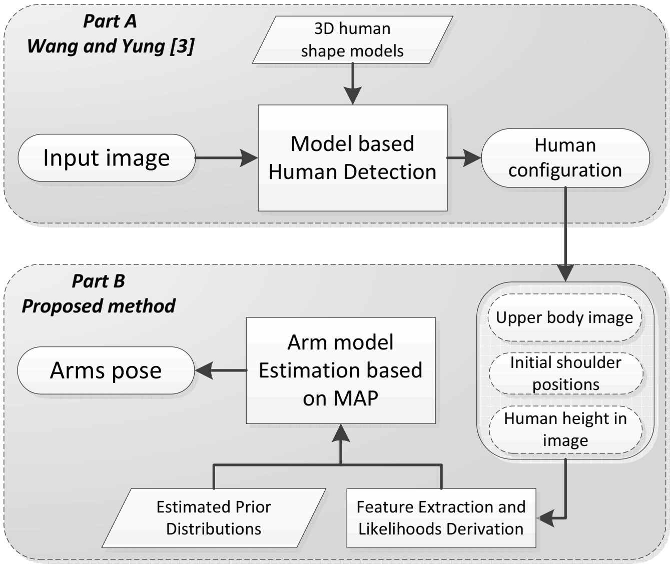
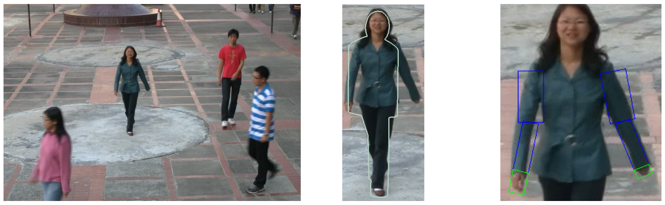

Part A
Wang and Yung [3]
3D human shape models
Input image
Model based Human Detection
Human configuration
Part B
Proposed method
Arms pose
Arm model Estimation based on MAP
Upper body image
Initial shoulder positions
Human height in image
Estimated Prior Distributions
Feature Extraction and Likelihoods Derivation

Fig. 1: Arm pose modeling of the proposed method (Part B) based on the human model-based detection[3] (Part A).



(a) Original image (b) 3D model (c) Arm model

Fig. 2: An example of arm pose modeling.

This arm pose modeling approach extends the model-based human detection work by Wang and Yung [3] as depicted in the Part A of Fig.1. In [3], human pose shown in Fig.2(b) is estimated by matching pre-defined 3D human models, also in the MAP sense with image evidence nominated by a head detector and a foot detector. From the human model, the initial left and right shoulder joints and human body height can be easily determined. Based on these parameters, the arm pose depicted in Fig.2(c) is modeled by our proposed method as shown in the Part B of Fig.1.

In this paper, the main contributions are:

1) An effective Bayesian network is adopted to depict the causal dependency relationship between arm parameters;

2) Realistic prior probabilities are estimated by a kernel estimator from real 3D motion data and two sets of probabilistic templates are used to derive likelihoods of arm parts;

3) The total Percentage of Correctly estimated body Parts (PCP) of arm model reaches 0.82 and 0.93 with 0.3 PCP threshold on the CAVIAR sub-dataset and the HKU campus dataset respectively, which is better than the method described in [4].

The organization of this paper is as follows: Section 2 outlines the related works on arm pose or upper body pose estimation. Section 3 presents our proposed approach. Section 4 describes the experiments and results, and Section 5 concludes this paper.

2. Related works

There are a large number of ongoing research works that focus on arm pose modeling and/or estimation from monocular image or video. When arm pose, head and torso are considered together, it is known as upper body pose modeling. In general, research can be divided to two main categories: Analysis/Synthesis (AS) based approach and Pictorial Structure (PS) based approach.

AS-based methods model pose by optimizing the similarity between model projection and the observed image [5]. In the work of Hsu [6], 3D human arm is modeled as a kinematic linkage connected by movable joints. The projected motion vector of images are used to approximate the 3D kinematic arm motion. A new 2D arm image is then recovered according to the estimated 3D arm pose and the corresponding texture. This method heavily depends on the correctness of motion estimation. Moeslund et al. [7] proposed a compact model representation of human arm with just two parameters, which is based on the screw-axis representation. The solution space is pruned according to the arm's kinematic constraints, from which an exhaustive silhouette matching is performed between the image and 3D arm models. The matching time is reduced due to the reduction in the solution space. However, for some poses, this silhouette-matching framework has depth error of up to ±20 cm since some silhouettes are not unique. AS-based methods are also used in Human Computer Interaction (HCI) applications for reliable real-time 3D arm pose estimation and user motion reconstruction [8]. Background estimation and skin color detection are incorporated in estimating the Region of Interest (ROI) for the upper arm and forearm. Within the ROIs, the 2D back-projected model edges from 3D sticks-figure model are compared with the silhouette of the foreground to find an edge score. The 3D pose with maximum edge score is considered the estimated pose. To achieve 60 ms/frame, constraint of arm pose and subsampling technique are adopted. However, it suffers from the same drawback as [7], and it has only been tested on one video with limited clothing style over a simple background.

In recent years, PS-based methods for upper body pose modeling have become more popular. PS was first introduced by Fishchler and Elschlager [9] in 1973. It represents an object by a collection of parts in a deformable model and pairs of parts are represented by spring-like connections. Bayesian formulation has also been incorporated in the PS recently by Felzenszwalb and Huttenlocher [10] for object recognition. Moreover, Ramanan [11] proposed an iterative parsing process based on the PS with trained conditional random field (CRF) based deformable models. Its performs better compared with [12], which only used trained deformable models. Based on the work of Ramanan [11], the priors of head and torso are included in [4], [13] to fit model by human detection and tracking, foreground highlighting, appearance models estimating and pose parsing. Its PCP reaches 79% and 51.5% with 0.5 PCP threshold on the Buffy and ETHZ PASCAL datasets respectively. The method is able to cope with different scale, illumination condition, as well as sleeve texture and color. It assumes that people are upright and seen mainly from a frontal or back viewpoint. It is selected to be the baseline method for the proposed method. PS is also used in long term arm and hand tracking for continuous sign language [14].

In summary, AS-based methods rely on silhouette matching. Unfortunately, arm silhouettes are sometimes not unique. This makes it impossible to derive correct 3D arm pose sometimes. On the other hand, PS-based methods do not rely on a model to avoid matching confusion, although it can only deal with 2D arm pose estimation. Under the Bayesian framework, it contains the prior structure relationship and deformable models of arm parts. However, the main drawback of PS-based methods is the difficulty in determining the prior distributions [10]. In this proposed method, features such as boundary, foreground and skin color are used rather than just silhouette. Furthermore, prior distributions between arm model parameters are directly estimated from a 3D motion dataset, in order to make it realistic.

3. Proposed method

3.1 2D Arm model and its Bayesian network

The 2D human arm model used is based on the skeletal articulation structure in which the upper arm, forearm and hand are modeled as rectangles and are connected as shown in Fig.3. This arm model is similar to the one used in [14]. In the xy plane, a specific arm pose can be determined by four points: the shoulder joint p_s, elbow joint p_e, wrist joint p_w and hand endpoint p_h. θ_U, θ_F and θ_H are orientation angles

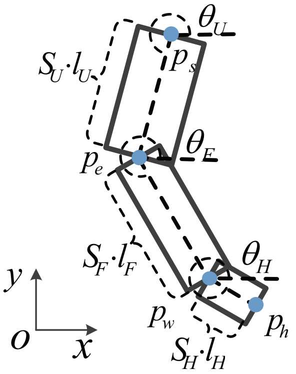


Fig. 3: 2D arm model in *xy* plane.

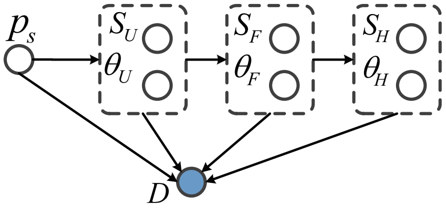


Fig. 4: Bayesian network for the 2D arm model.

w.r.t. *x* axis for the upper arm, forearm and hand respectively. Whereas $_U$, l_F and $_H$ are the lengths of upper arm, forearm and hand in the image. S_U, S_F and S_H are scaling factors of the upper arm, forearm and hand for a specific arm pose. As such, θ_U, θ_F and θ_H are continuous from 0 to 360 degrees. Similarly, S_U, S_F and S_H are continuous from 0 to 1. Human height is assumed known from Wang and Yung [3]. The statistic ratios of upper arm, forearm and hand to human height are given by [15], from which l_U, l_F and l_H are estimated.

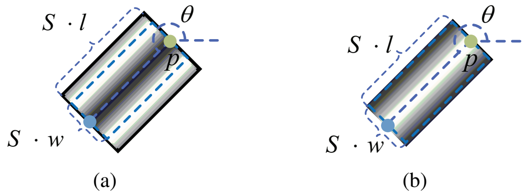
There is a causal dependency relationship between the arm parts parameters. The upper arm is determined by p_s, θ_U, S_U and l_U. The endpoint of the upper arm is the elbow joint p_e. In the same way, the forearm is defined by p_e, θ_F, S_F, l_F and p_w. The hand pose is modeled similarly. Hence, a specific arm pose can be represented by a vector θ $(p_s, \theta_U, S_U, l_U, \theta_F, S_F, l_F, \theta_H, S_H, l_H)$ in which all parameters are to be estimated except l_U, l_F and l_H.

Given the observation data $\mathbf{D}$, which contains the boundary, foreground and skin color features of an upper body image from [3], the posterior probability of the specific arm pose is $P(\theta|\mathbf{D})$. According to the Bayesian rule, this posterior probability is proportional to its joint probability:

$$P\left(\theta \mid \mathbf{D}\right) \propto \; P\left(\theta, \mathbf{D}\right). \tag{1}$$

The Bayesian network as shown in Fig.4 depicts this relationship. It can efficiently encode the joint probability as follows:

$$\begin{aligned} P(\theta, \mathbf{D}) =& P(p_s)P(\mathbf{D}|p_s)P(\theta_U, S_U|p_s)P(\mathbf{D}|\theta_U, S_U) \\ & P(\theta_F, S_F|\theta_U, S_U)P(\mathbf{D}|\theta_F, S_F) \\ & P(\theta_H, S_H|\theta_F, S_F)P(\mathbf{D}|\theta_H, S_H), \end{aligned} \tag{2}$$



(a)

(b)

Fig. 5: Two sets of probabilistic templates for every arm part likelihood calculation: (a) $b_{(\theta,S)}$ for the probability boundary image; (b) $g_{(\theta,S)}$ for the foreground image and skin color mask image. The brighter point has higher weight or probability than the other points.

where $P(\mathbf{D}|\cdot)$ are likelihoods, and the rest are prior probabilities. The likelihoods are the probabilities for the specific parameters given the observation $\mathbf{D}$. The prior probabilities describe the parameters dependency.

In order to derive the arm pose for a human object, the vector $\widehat{\theta}(\widehat{p}_s, \widehat{\theta}_U, \widehat{S}_U, \; \widehat{\theta}_F, \widehat{S}_F, \widehat{\theta}_H, \widehat{S}_H)$ with MAP is the best 2D arm pose model, or

$$\widehat{\theta} = \underset{\theta}{argmax}\, P(\theta|\mathbf{D}) = \underset{\theta}{argmax}\, P(\theta, \mathbf{D}). \tag{3}$$

3.2 Likelihood Derivation

Given the image of an upper body, the likelihoods $P(\mathbf{D}|p_s)$, $P(\mathbf{D}|\theta_U, S_U)$, $P(\mathbf{D}|\theta_F, S_F)$ and $P(\mathbf{D}|\theta_H, S_H)$ quantify the fitness between different parameters and the observed data of arm parts. To determine $P(\mathbf{D}|p_s)$, we start with the point $p_{s_0}(x_0, y_0)$, which is given as part of the model from [3]. Assuming that $p_{s_0}(x_0, y_0)$ has high probability to be the shoulder joint, a helical spring loci is used to select the sampling points as given in Eqn.(4), which can reduce the search of solution space substantially,

$$\begin{cases} x_\phi = \rho_x \phi \cos(\phi) + x_0 \\ y_\phi = \rho_y \phi \sin(\phi) + y_0 \end{cases}, \tag{4}$$

where (x_0, y_0) is the initial shoulder joint coordinate p_{s_0} and (x_ϕ, y_ϕ) is sampling point p_{s_ϕ} with angle ϕ. ρ_x and ρ_y control the range of sampling points. Only the sampling points are considered to be shoulder joints candidate. Therefore,

$$P(\mathbf{D}|p_s) = \begin{cases} 1 & if \; p_s = p_{s_\phi} \\ 0 & otherwise \end{cases}.$$

In order to recognize arm parts, two different probabilistic templates are proposed as shown in Fig.5 to capture the features of boundary, foreground and skin color. For instance, if we want to generate probabilistic templates for the upper arm given θ_U, S_U, p_s and l_U, the template parameters shown in Fig.5 are: $p = p_s$, $\theta = \theta_U$, $S = S_U$, $l = l_U$ and $w = r_U \cdot l_U$, where r_U is the average ratio between the length and width of the upper arm according to [15]. Similarly, the templates for the forearm and hand can be generated.

As depicted in Fig.5(a), $b_{(\theta,S)}$ contains two Gaussian distributions on both sides, whereas $g_{(\theta,S)}$ as shown in

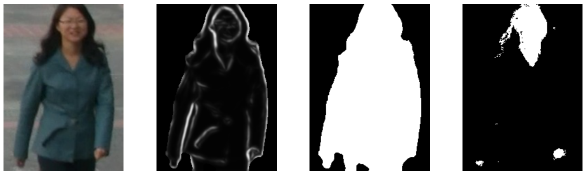

(a) (b) (c) (d)

Fig. 6: Features for likelihood calculation: (a) original image; (b) corresponding probability boundary f_{pb}; (c) corresponding foreground f_{fg}; (d) skin color mask f_{sc}.

Fig.5(b) contains only Gaussian distributions in the middle. The variances of the Gaussian distributions take into account of the variations of arm boundary for the different clothing style and body shape.

The likelihood calculation considers three features: boundary, foreground, and skin color mask shown in Fig.6. The probability boundary (*pb*) approach [16] is employed to generate the probability boundary image f_{pb} shown in Fig.6(b). The foreground f_{fg} shown in Fig.6(c) is extracted by the Wang and Yung method [17]. The skin color mask f_{sc} shown in Fig.6(d) is calculated by the Conaire et al. method [18]. These methods are chosen for their robustness and performance.

For upper arm and forearm, their templates $b_{(\theta,S)}$ and $g_{(\theta,S)}$ should have high responses on f_{pb} and f_{fg}. For the hand, its template $g_{(\theta,S)}$ has high response on f_{sc}. The likelihoods of upper arm, forearm and hand are given below:

$$\begin{aligned} P(\mathbf{D}|\theta_U, S_U) &= P(f_{pb}, f_{fg}|\theta_U, S_U) \\ &= P(f_{pb}|\theta_U, S_U)P(f_{fg}|\theta_U, S_U), \end{aligned} \tag{5}$$

$$\begin{aligned} P(\mathbf{D}|\theta_F, S_F) &= P(f_{pb}, f_{fg}|\theta_F, S_F) \\ &= P(f_{pb}|\theta_F, S_F)P(f_{fg}|\theta_F, S_F), \end{aligned} \tag{6}$$

$$P(\mathbf{D}|\theta_H, S_H) = P(f_{sc}|\theta_H, S_H), \tag{7}$$

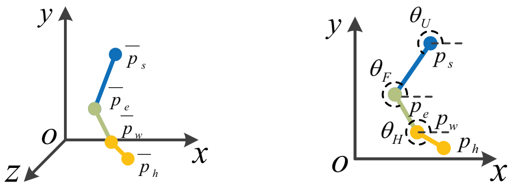
where

$$\begin{aligned} P(f_{pb}|\theta_., S_.) &= \sum_x \sum_y b_{(\theta_.,S_.)}(x,y) f_{pb}(x,y), \\ P(f_{fg}|\theta_., S_.) &= \sum_x \sum_y g_{(\theta_.,S_.)}(x,y) f_{fg}(x,y), \\ P(f_{sc}|\theta_H, S_H) &= \sum_x \sum_y g_{(\theta_H,S_H)}(x,y) f_{sc}(x,y), \end{aligned}$$

and $\theta_., S_.$ can be either θ_U, S_U or θ_F, S_F. All the templates are normalized such that $\sum_x \sum_y b_{(\theta_.,S_.)}(x,y) = 1$ and $\sum_x \sum_y g_{(\theta_.,S_.)}(x,y) = 1$.

3.3 Prior Distributions

Within the framework of Bayesian network, the challenge is how to estimate representative prior distributions [10]. The requirement for prior distributions is both informative and generality. For instance, the arm movement is unable to violate the constraints of physiological structure as described in [15]. As different person may have slightly different arm pose when performing the same action, a prior distribution should cover all possible variations. Gaussian distribution as the prior distribution of human body part relationship is widely used in pose estimation [14], [10]. In this paper, we estimate the prior distributions directly from a real 3D arm articulation data of a specific action instead. A non-parametric procedure is used to estimate them without assuming that the forms of the underlying densities are known.



(a) (b)

Fig. 7: Projection of arm joints from *xyz* space to *xy* plane. (a) $\overline{p}_s$, $\overline{p}_e$, $\overline{p}_w$ and $\overline{p}_h$ in *xyz* space; (b) corresponding joints p_s, p_e, p_w and p_h, as well as the orientation angles θ_U, θ_F and θ_H in *xy* plane.

For a set of joints with 3D coordinates, first a coordinate transformation is applied to make sure that the torso is parallel to the *xy* plane and faces along the *z* direction, as well as the line between the two shoulder joints is parallel to the *x* axis. The shoulder joint $\overline{p}_s$, the elbow joint $\overline{p}_e$, the wrist joint $\overline{p}_w$ and the hand endpoint $\overline{p}_h$ in the *xyz* space are depicted in Fig.7(a). The projected joints on the *xy* plane are depicted in Fig.7(b). Besides, every arm part has 2 Degree of Freedom (DOF), in which one is the degree in the *xy* plane such as θ_U shown in Fig.7(b), and another is φ_U between $\overline{p}_s\overline{p}_e$ and the *xy* plane. θ_U and φ_U are independent since they represent the two DOF of the upper arm. The projected upper arm length in the *xy* plane is $L_U = |p_s p_e| = cos\varphi_U |\overline{p}_s\overline{p}_e| = S_U \overline{L}_U$, where $\overline{L}_U = |\overline{p}_s\overline{p}_e|$ is the real upper length in the *xyz* space, and $S_U = cos\varphi_U$ is the scaling factor of upper arm length in the *xy* plane to its real length. Hence, orientation angle θ_U and scaling factor S_U are independent.

The prior probabilities $P(p_s)$, $P(\theta_U, S_U|p_s)$ $P(\theta_F, S_F|\theta_U, S_U)$ and $P(\theta_H, S_H|\theta_F, S_F)$ can be decomposed as:

$$P(\theta_U, S_U|p_s) = P(\theta_U|p_s)P(S_U|p_s), \tag{8}$$

$$P(\theta_F, S_F|\theta_U, S_U) = P(\theta_F|\theta_U)P(S_F|S_U), \tag{9}$$

$$P(\theta_H, S_H|\theta_F, S_F) = P(\theta_H|\theta_F)P(S_H|S_F). \tag{10}$$

Assuming that the prior distribution of p_s is uniform and the shoulder position does not directly influence the orientation and size of the upper arm, $P(p_s)$ is a constant and $P(\theta_U|p_s) = P(\theta_U)$, $P(S_U|p_s) = P(S_U)$.

As the parameter spaces θ and S of every arm parts are large, it is necessary to discretize them. Based on the

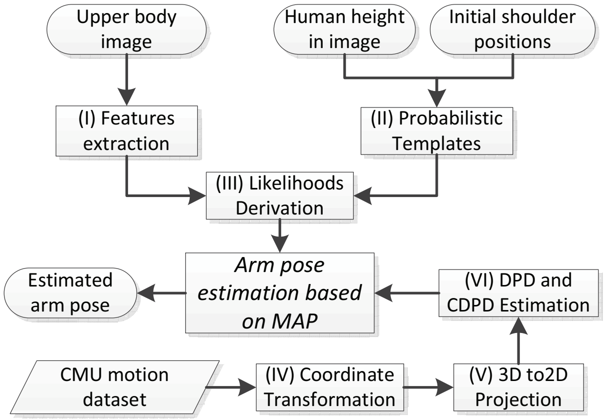

Upper body image
Human height in image
Initial shoulder positions
(I) Features extraction
(II) Probabilistic Templates
(III) Likelihoods Derivation
Estimated arm pose
Arm pose estimation based on MAP
(VI) DPD and CDPD Estimation
CMU motion dataset
(IV) Coordinate Transformation
(V) 3D to2D Projection

Fig. 8: Flowchart of arm pose modeling based on MAP.

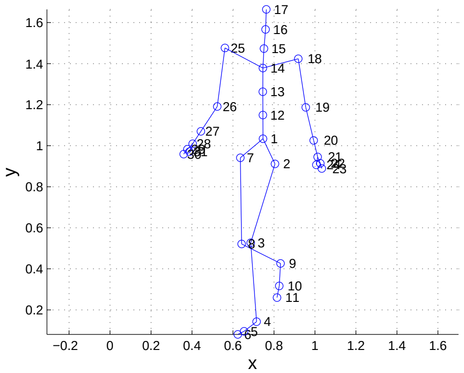


Fig. 9: Coordinates of 31 joints in 3D for a walking person.

projected data θ_U, S_U, θ_F, S_F, θ_H, S_H of a specific action, the probability density function (PDF) of the upper arm is estimated first, and then this PDF is discretized to give the discrete probability distributions (DPD). The conditional discrete probability distribution (CDPD) of the forearm and hand are estimated in the same manner. In this paper, a Gaussian kernel density estimator is adopted to conduct the PDF estimation.

3.4 Implementation

The steps of arm pose estimation modeling are depicted in Fig.8. The inputs are upper body image, initial shoulder position and estimated human height. Features extraction (Fig.8(I)), probabilistic templates generation (Fig.8(II)) and likelihood derivation (Fig.8(III)) are described in Section 3.2, and prior distribution estimation (Fig.8(IV-VI)) is described in Section 3.3.

The CMU motion dataset [19] is used to estimate the prior distributions. It contains various human action categories, in which the whole body joints movement 3D information are recorded in real time. Since this paper focuses on arm pose modeling, only the walking category and the related arm joints are needed for the prior distribution estimation. When walking, the hand orientation is assumed to be the same as the forearm. Only the parameters of upper arm and forearm are estimated in this case.

As depicted in Fig.9, there are 31 joints in 3D to depict a body pose. The shoulder, elbow and wrist for the left arm of a person are the joints numbered 18, 19 and 20, respectively. The shoulder, elbow and wrist of the right arm are the joints numbered 25, 26 and 27, respectively.

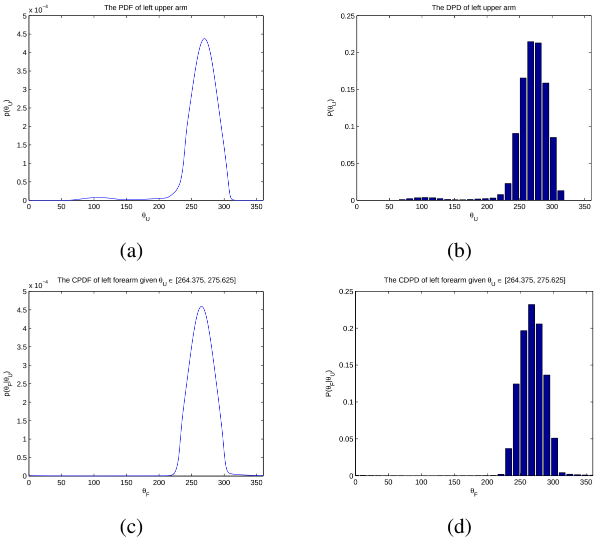
First, a coordinate transformation (Fig.8(IV)) is required for the 31 joints to make sure that the plane with joints 25, 18 and 1 is parallel to the *xy* plane and the line through joints 25 and 18 is parallel to the *x* axis. Second, the six joints of the arms are projected to the *xy* plane (Fig.8(V)) to derive θ_U, S_U and θ_F, S_F. Third, DPD of θ_U is estimated, then for every bin of θ_U, CDPD of θ_F is estimated (Fig.8(VI)). In the same manner, DPD of S_U and CDPD of S_F are estimated. Since the hand is assumed to be in the same direction as the forearm when walking, we have $P(\theta_H, S_H|\theta_F, S_F) = 1$ if $\theta_H = \theta_F$, otherwise $P(\theta_H, S_H|\theta_F, S_F) = 0$.

The total number of walking style frames used for prior probability is 275,665. For the orientation angles of an arm, there is a DPD of θ_U for the upper arm and 32 CPDPs of θ_F for the forearm in which θ_U, θ_F are discretized to 32 bins and the range of every bin is 11.25 degrees. For the length scaling factor of an arm, there is a DPD of S_U for the upper arm and 16 CPDPs of S_F for the forearm in which S_U, S_F are discretized to 16 bins and the range of every bin is 1/16. In order to keep the shape of the continuous PDFs and consider the size of solution space, the bin number of orientation angles and length scaling factors are selected as 32 and 16 respectively.Fig.10 shows examples of PDF, CPDF, DPD and CDPD estimation for orientation angles θ_U, θ_F.

4. Experiment and results

The proposed method has been evaluated by the videos of shopping center in Portugal (2nd set) of CAVIAR dataset and an outdoor video taken at the HKU campus. All of them were annotated as stick-man with the positions of shoulder, elbow and wrist in image for both arms. As in [4], PCP is used to measure the performance of the proposed method. In principle, if the segment endpoints of an estimated body part lie within a ratio of the length of the ground-truth segment from their annotated location, the estimation is considered correct. The ratio is called PCP threshold. First, the video was processed by the Wang and Yung method [3] to obtain the upper body image, initial shoulder joints and estimated human height. Then, this proposed approach and the method of Eichner et al. [13] are used to estimate and evaluate arm pose of the two datasets. We run the code from [20] of the method proposed by Eicher et al. for comparison.

There are a total of 420 upper body images from the CAVIAR dataset and 157 upper body images from HUK



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

Fig. 10: Example of probability distributions estimation for θ_U and θ_F: (a) PDF of θ_U for left upper arm; (b) DPD of θ_U; (c) CPDF of left forearm θ_F where its θ_U corresponds to the 25th bin with bin center 270 degrees; (d) CPDP of forearm θ_F.

PCP threshold	0.1	0.3	0.5
All parts	0.12	0.82	0.96
LUA	0.18	0.93	0.99
LFA	0.03	0.66	0.92
RUA	0.22	0.96	1.00
RFA	0.06	0.74	0.93

Table 1: PCP of the proposed method with various PCP threshold for the CAVIAR sub-dataset (LUA: left upper arm, LFA: left forearm, RUA: right upper arm, RFA: right forearm).

campus dataset in different clothing styles, color and textures for testing. Fig.13 depicts some results of CAVIAR sub-dataset by the proposed method whereas Fig.14 depicts the results of HKU campus dataset with upper arm, forearm and hand parts. Fig.11 and Fig.12 depict the average PCP curves of the proposed method and the Eichner's method with PCP threshold from 0.1 to 0.5 for the two test datasets. Table 1 and Table 2 also give the PCP values of the whole body and every body part with threshold of 0.1, 0.3 and 0.5 for both datasets of this proposed approach. According to these figures, this proposed approach gains better performances on both datasets. According to these tables, the average PCP of our proposed method can reach up to 0.82 and 0.93 even with a 0.3 PCP threshold for CAVIAR sub-dataset and HKU campus datasets respectively.

We found that this proposed approach has better performance on the higher resolution images such as this HKU campus dataset. The main reason is the sizes of probability

PCP threshold	0.1	0.3	0.5
All parts	0.38	0.93	0.97
LUA	0.52	0.95	0.96
LFA	0.42	0.89	0.95
RUA	0.29	0.97	0.99
RFA	0.31	0.94	0.98

Table 2: PCP of the proposed method with various PCP threshold for the HKU campus dataset.

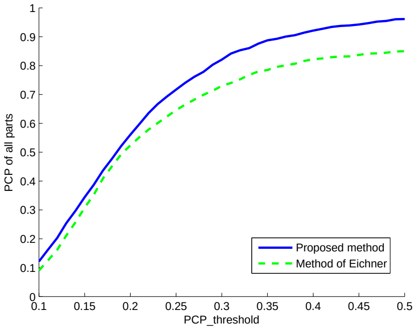


Fig. 11: Average PCP curve of the proposed method and Eichner's method [4] for CAVIAR sub-dataset.

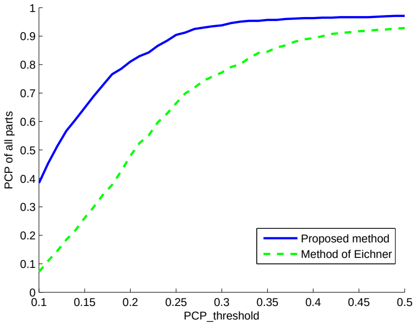


Fig. 12: Average PCP curve of the proposed method and Eichner's method [4] for the HKU campus dataset.

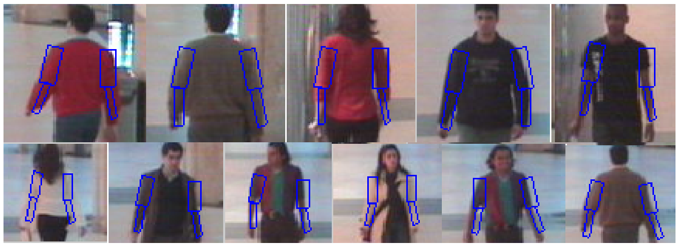


Fig. 13: Results of the proposed method for CAVIAR sub-dataset in which the wider rectangles in blue are upper arms, the other thin rectangles in blue are forearms.

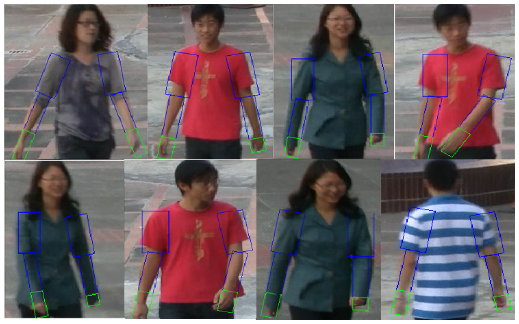


Fig. 14: Results of the proposed method for HKU campus dataset in which the wider rectangles in blue are upper arms, the other thin rectangles in blue are forearms and the rectangles in green are hands.

templates are small that make it is hard to capture the likelihoods for the low resolution images. The computational complexity is $O(h^2)$ where h is the estimated human height in pixel from the work of Wang and Yung [3]. This MAP based approach also can be implemented efficiently by GPU.

5. Conclusions

This paper represents the statistical framework based on Bayesian network for arm pose modeling. The probabilistic templates are generated from the *pb*, foreground and skin color features for the likelihoods of arm parts. The prior distributions are estimated based on the walking category in the CMU motion dataset by the kernel density estimation. The results achieved suggest that the proposed probabilistic templates and the prior distributions are effective to model arm pose of the walking pedestrian in outdoor scenes. The future works includes estimation of prior distributions for other actions, model selection, and viewpoint variations.

References

[1] A. Mehrabian, "Communication without words," *Psychological Today*, vol. 2, pp. 53–55, 1968.

[2] S. Mitra and T. Acharya, "Gesture recognition: A survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 37, no. 3, pp. 311–324, 2007.

[3] L. Wang and N. Yung, "Bayesian 3d model based human detection in crowded scenes using efficient optimization," in *Applications of Computer Vision (WACV), 2011 IEEE Workshop on*. IEEE, 2011, pp. 557–563.

[4] M. Eichner, V. Ferrari, and S. Zurich, "Better appearance models for pictorial structures," in *Proc. BMVC*, vol. 2, no. 4. Citeseer, 2009, p. 6.

[5] T. Moeslund, A. Hilton, and V. Kruger, "A survey of advances in vision-based human motion capture and analysis," *Computer vision and image understanding*, vol. 104, no. 2-3, pp. 90–126, 2006.

[6] R. Hsu, M. Kageyama, H. Fukui, Y. Nakaya, and H. Harashima, "Human arm modeling for analysis/synthesis image coding," in *Robot and Human Communication, 1993. Proceedings., 2nd IEEE International Workshop on*. IEEE, 1993, pp. 352–355.

[7] T. Moeslund and E. Granum, "Modelling and estimating the pose of a human arm," *Machine Vision and Applications*, vol. 14, no. 4, pp. 237–247, 2003.

[8] S. Salti, O. Schreer, and L. Di Stefano, "Real-time 3d arm pose estimation from monocular video for enhanced hci," in *Proceeding of the 1st ACM workshop on Vision networks for behavior analysis*. ACM, 2008, pp. 1–8.

[9] M. Fischler and R. Elschlager, "The representation and matching of pictorial structures," *Computers, IEEE Transactions on*, vol. 100, no. 1, pp. 67–92, 1973.

[10] P. Felzenszwalb and D. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[11] D. Ramanan, "Learning to parse images of articulated bodies," *Advances in Neural Information Processing Systems*, vol. 19, p. 1129, 2007.

[12] D. Ramanan and C. Sminchisescu, "Training deformable models for localization," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, vol. 1. IEEE, pp. 206–213.

[13] M. Eichner, M. Marin-Jimenez, A. Zisserman, and V. Ferrari, "Articulated human pose estimation and search in (almost) unconstrained still images," in *ETH Zurich, D-ITET, BIWI, Technical Report No.272*, 2010.

[14] P. Buehler, M. Everingham, D. Huttenlocher, and A. Zisserman, "Long term arm and hand tracking for continuous sign language tv broadcasts," in *Proc. BMVC*, vol. 1281. Citeseer, 2008.

[15] A. Tilley and H. D. Associates, *The measure of man and woman: human factors in design*. Wiley, 2002.

[16] D. Martin, C. Fowlkes, and J. Malik, "Learning to detect natural image boundaries using local brightness, color, and texture cues," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 5, pp. 530–549, 2004.

[17] L. Wang and N. Yung, "Extraction of moving objects from their background based on multiple adaptive thresholds and boundary evaluation," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 11, no. 1, pp. 40–51, 2010.

[18] C. Conaire, N. O'Connor, and A. Smeaton, "Detector adaptation by maximising agreement between independent data sources," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2007, pp. 1–6.

[19] "Carnegie Mellon University motion capture database," http://mocap.cs.cmu.edu/.

[20] "2D articulated human pose estimation software v1.21," http://www.vision.ee.ethz.ch/~calvin/, 2011.