# Human Upper Body Pose Estimation in Static Images

Mun Wai Lee and Isaac Cohen

Institute for Robotics and Intelligent Systems
Integrated Media Systems Center
University of Southern California
Los Angeles, CA 90089-0273, USA
{munlee, icohen}@usc.edu
WWW home page: http://www-scf.usc.edu/~munlee/index.html

**Abstract.** Estimating human pose in static images is challenging due to the high dimensional state space, presence of image clutter and ambiguities of image observations. We present an MCMC framework for estimating 3D human upper body pose. A generative model, comprising of the human articulated structure, shape and clothing models, is used to formulate likelihood measures for evaluating solution candidates. We adopt a data-driven proposal mechanism for searching the solution space efficiently. We introduce the use of proposal maps, which is an efficient way of implementing inference proposals derived from multiple types of image cues. Qualitative and quantitative results show that the technique is effective in estimating 3D body pose over a variety of images.

## 1 Estimating Pose in Static Image

This paper proposes a technique for estimating human upper body pose in static images. Specifically, we want to estimate the 3D body configuration defined by a set of parameters that represent the global orientation of the body and body joint angles. We are focusing on middle resolution images, where a person's upper body length is about 100 pixels or more. Images of people in meetings or other indoor environment are usually of this resolution. We are currently only concerned with estimating the upper body pose, which is relevant for indoor scene. In this situation the lower body is often occluded and the upper body conveys most of a person's gestures. We do not make any restrictive assumptions about the background and the human shape and clothing, except for not wearing any head wear nor gloves.

### 1.1 Issues

There are two main issues in pose estimation with static images, the high dimension state space and pose ambiguity.

**High Dimension State Space.** Human upper body pose has about 20 parameters and pose estimation involves searching in a high dimensional space with complex distribution. With static images, there is no preceding pose for initializing the search, unlike a video tracking problem. This calls for an efficient mechanism for exploring the solution space. In particular, the search is preferably data-driven, so that good solution candidates can be found easily.

**Pose Ambiguity.** From a single view, the inherent non-observability of some of the degrees of freedom in the body model leads to *forwards/backwards flipping ambiguities* [10] of the depth positions of body joints. Ambiguity is also caused by noisy and false observations. This problem can be partly alleviated by using multiple image cues to achieve robustness.

### 1.2 Related Work

Pose estimation on video has been addressed in many previous works, either using multiple cameras [3] or a single camera [2, 9]. Many of these works used the particle filter approach to estimate the body pose over time, by relying on a good initialization and temporal smoothness. Observation-based importance sampling scheme has also been integrated into this approach to improve robustness and efficiency [5].

For static images, some works have been reported for recognizing prototype body poses using shape context descriptors and exemplars [6]. Another related work involves the mapping of image features into body configurations [8]. These works however rely on either a clean background or that the human is segmented by a background subtraction and therefore not suitable for fully automatic pose estimation in static images.

Various reported efforts were dedicated to the detection and localization of body parts in images. In [4, 7], the authors modeled the appearance and the 2D geometric configuration of body parts. These methods focus on real-time detection of people and do not estimate the 3D body pose. Recovering 3D pose was studied in [1, 11], but the proposed methods assume that image positions of body joints are known and therefore tremendously simplify the problem.

## 2 Proposed Approach

We propose to address this problem, by building an image generative model and using the MCMC framework to search the solution space. The image generative model consists of (*i*) human model, which encompasses the articulated structure, shape and the type of clothing, (*ii*) scene-to-image projection, and (*iii*) generation of image features. The objective is to find the human pose that maximizes the posterior probability.

We use the MCMC technique to sample the complex solution space. The set of solution samples generated by the Markov chain weakly converges to a stationary distribution equivalent to the posterior distribution. Data-driven MCMC

framework [13] allows us to design good proposal functions, derived from image observations. These observations include face, head-shoulder contour, and skin color blobs. These observations, weighted according to their saliency, are used to generate *proposal maps*, that represent the proposal distributions of the image positions of body joints. These maps are first used to infer solutions on a set of 2D pose variables, and subsequently generate proposals on the 3D pose using inverse kinematics. The proposal maps considerably improve the estimation, by consolidating the evidences provided by different image cues.

## 3 Human Model

### 3.1 Pose Model

This model represents the articulated structure of the human body and the degree of freedom in human kinematics. The upper body consists of 7 joints, 10 body parts and 21 degree of freedom (6 for global orientation and 15 for joint angles). We assume an orthographic projection and use a scale parameter to represent the person height.

### 3.2 Probabilistic Shape Model

The shape of each body part is approximated by a truncated 3D cone. Each cone has three free parameters: the length of the cone and the widths of the top and base of the cone. The aspect ratio of the cross section is assumed to be constant. Some of the cones share common widths at the connecting joints. In total, there are 16 shape parameters. As some of the parameters have small variances and some are highly correlated, the shape space is reduced to 6 dimensions using PCA, and this accounts for 95% of the shape variation in the training data set.

### 3.3 Clothing Model

This model describes the person's clothing to allow the hypothesis on where the skin is visible, so that observed skin color features can be interpreted correctly. As we are only concerned with the upper body, we use a simple model with only one parameter that describes the length of the sleeve. For efficiency, we quantized this parameter into five discrete levels, as shown in Figure 1a.

## 4 Prior Model

We denote the state variable as $m$, which consists of four subsets: ($i$) global orientation parameters: $g$, ($ii$) local joint angles: $j$, ($iii$) human shape parameters: $s$, and ($iv$) clothing parameter: $c$.

$$m = \{g, i, s, c\} \ . \tag{1}$$

Assuming that the subsets of parameters are independent, the prior distribution of the state variable is given by:

$$p(m) \approx p(g)p(j)p(s)p(c). \tag{2}$$

**Global Orientation Parameters.** The global orientation parameters consist of image position ($x_g$), rotation parameters ($r_g$) and a scale parameter ($h_g$). We assume these parameters to be independent so that the following property holds:

$$p(g) \approx p(x_g)p(r_g)p(h_g). \tag{3}$$

The prior distributions are modeled as normal distributions and learned from training data.

**Joint Angles Parameters.** The subset $j$, consists of 15 parameters describing the joint angles at 7 different body joint locations.

$$j = \{j_i, i = neck, left\_wrist, left\_elbow, \ldots, right\_shoulder\}. \tag{4}$$

In general, the joint angles are not independent. However, it is impracticable to learn the joint distribution of the 15 dimensional $j$ vector, with a limited training data set. As an approximation, our prior model consists of joint distribution of pair-wise neighboring body joint locations. For each body location, $i$, we specify a neighboring body location as its parent, where:

$$\begin{array}{ll} parent(left\_wrist) = left\_elbow & parent(right\_wrist) = right\_elbow \\ parent(left\_elbow) = left\_shoulder & parent(right\_elbow) = right\_shoulder \\ parent(left\_shoulder) = torso & parent(right\_shoulder) = torso \\ parent(neck) = torso & parent(torso) = \emptyset \end{array}$$

The prior distribution is then approximated as:

$$p(j) \approx \lambda_{pose} \prod_i p(j_i) + (1 - \lambda_{pose}) \prod_i p(j_i, j_{parent(i)}) \tag{5}$$

where $\lambda_{pose}$ is a constant valued between 0 and 1. The prior distributions $p(j_i)$ and $p(j_i, j_{parent(i)})$ are modeled as Gaussians. The constant $\lambda_{pose}$ is estimated from training data using cross-validation, based on the maximum likelihood principle.

**Shape Parameters.** PCA is used to reduce the dimensionality of the shape space by transforming the variable $s$ to a 6 dimensions variable $s'$ and the prior distribution is approximated by a Gaussian:

$$p(s) \approx p(s') \approx N(s', \mu_{s'}, \Sigma_{s'}) \tag{6}$$

where $\mu_{s'}$ and $\Sigma_{s'}$ are the mean and covariance matrix of the prior distribution of $s'$.

**Clothing Parameters.** The clothing model consists of a discrete variable $c$, representing the sleeve length. The prior distribution is based on the empirical frequency in the training data.

**Marginalized distribution of image positions of body joints.** We denote $\{u_i\}$ as the set of image positions of body joints. Given a set of parameters $\{g, j, s\}$, we are able to compute the image position of each body joints $\{u_i\}$:

$$u_i = f_i(g, j, s) \tag{7}$$

where $f_i(.)$ is a deterministic forward kinematic function. Therefore, there exists a prior distribution for each image position:

$$p(u_i) = \int\int\int f_i(g, j, s)p(g)p(j)p(s)dgdjds \tag{8}$$

where $p(u_i)$ represents the marginalized prior distribution of the image position of the $i$-th body joint. In fact, any variable that is derived from image positions of the body joints has a prior distribution, such as the lengths of the arms in the image or the joint positions of the hand and elbow. As will be described later, these prior distributions are useful in computing weights for the image observations. The prior distribution of these measures can be computed from Equation (8) or it can be learned directly from the training data as was performed in our implementation.

## 5 Image Observations

Image observations are used to compute data driven proposal distribution in the MCMC framework. The extraction of observations consists of 3 stages: (*i*) face detection, (*ii*) head-shoulders contour matching, and (*iii*) skin blobs detection.

### 5.1 Face Detection

For face detection, we use the Adaboost technique proposed by [12]. We denote the face detection output as a set of face candidates,

$$I_{Face} = \{I_{Face\_Position}, I_{Face\_Size}\}, \tag{9}$$

where $I_{Face\_Position}$ is the detected face location and $I_{Face\_Size}$ is the estimated face size. The observation can be used to provide a proposal distribution for the image head position, $u_{Head}$, modeled as a Gaussian distribution:

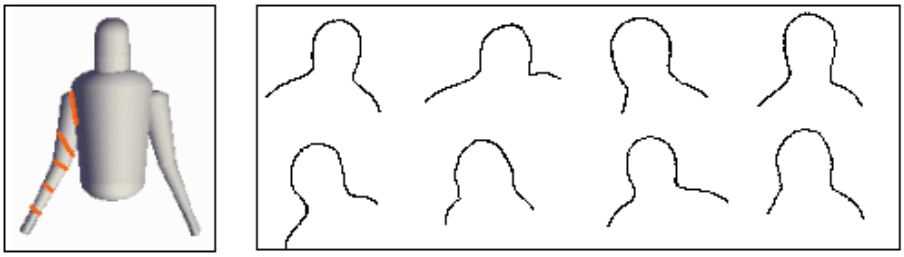
$$q(u_{Head}|I_{Face}) \sim N(u_{Head} - I_{Face\_Position}, \cdot, \cdot). \tag{10}$$

The parameters of the Gaussian are estimated from training data. The above expression can be extended to handle multiple detected faces.

### 5.2 Head-Shoulder Contour Matching

**Contour Model for Head-Shoulder.** We are interested in detecting 2D contour of the head and shoulders. Each contour is represented by a set of connected

points. This contour is pose and person dependent. For robustness, we use a mixture model approach to represent the distribution of the 2D contour space. Using a set of 100 training data, a K-mean clustering algorithm is used to learn the means of 8 components, as shown in Figure 1b. The joint distributions of these contour and the image position of head, neck and shoulders are also learned from the training data.



(a) (b)

**Fig. 1.** Models: (a) Quantized sleeve length of clothing model, (b) components of head-shoulder model.

**Contour Matching.** In each test image, we extract edges using the Canny detector, and a gradient descent approach is used to align each exemplar contour to these edges. We define a search window around a detected face, and initiate searches at different positions within this window. This typically results in about 200 contour candidates. The confidence of each candidate is weighted based on (*i*) confidence weight of the detected face, (*ii*) joint probability of contour position and detected face position, and (*iii*) edge alignment error. The number of candidates is reduced to about 50, by removing those with low confidence.

The resulting output is a set of matched contours $\{ I_{Head\_Shoulder,i}\}$. Each contour provides observations on the image positions of the *head*, *neck*, *left shoulder* and *right shoulder*, with a confidence weight $w_{HS,i}$:

$$I_{Head\_Shoulder,i} = \{w_{HS,i}, I_{Head\_Pos,i}, I_{Neck\_Pos,i}, I_{L\_Shoulder\_Pos,i}, I_{R\_Shoulder\_Pos,i}\}. \tag{11}$$

Each observation is used to provide proposal candidates for the image positions of the head ($u_{Head}$), left shoulder ($u_{L\_Shoulder}$), right shoulder ($u_{R\_Shoulder}$), and neck ($u_{Neck}$). The proposal distributions are modeled as Gaussian distributions given by:

$$\begin{aligned}
&q(u_{Head}|I_{Head_Shoulder,i}) \sim w_{HS,i}N(u_{Head} - I_{Head_Pos,i}, \cdot, \cdot)\\
&q(u_{Neck}|I_{Head_Shoulder,i}) \sim w_{HS,i}N(u_{Neck} - I_{Neck_Pos,i}, \cdot, \cdot)\\
&q(u_{L_Shoulder}|I_{Head_Shoulder,i}) \sim w_{HS,i}N(u_{L_Shoulder} - I_{L_Shoulder_Pos,i}, \cdot, \cdot)\\
&q(u_{R_Shoulder}|I_{Head_Shoulder,i}) \sim w_{HS,i}N(u_{R_Shoulder} - I_{R_Shoulder_Pos,i}, \cdot, \cdot)
\end{aligned} \tag{12}$$

The approach to combine all these observations is described in Section 5.4.

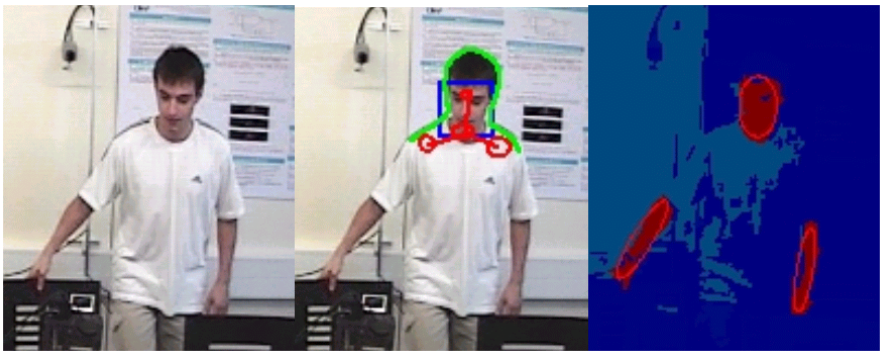


**Fig. 2.** Image observations, from left: (i) original image, (ii) detected face and head-shoulders contour, (iii) skin color ellipses extraction.

### 5.3 Elliptical Skin Blob Detection

Skin color features provide important cues on arms positions. Skin blobs are detected in four sub-stages: (*i*) color based image segmentation is applied to divide the image into smaller regions, (*ii*) the probability of skin for each segmented region is computed using a histogram-based skin color Bayesian classifier, (*iii*) ellipses are fitted to the boundaries of these regions to form skin ellipse candidates, and (*iv*) adjacent regions with high skin probabilities are merged to form larger regions (see Figure 2).

The extracted skin ellipses are used for inferring the positions of limbs. The interpretation of a skin ellipse is however dependent on the clothing type. For example, if the person is wearing short sleeves, then the skin ellipse represent the lower arm, indicating the hand and elbow positions. However, for long sleeve, the skin ellipse should cover only the hand and used for inferring the hand position only. Therefore the extracted skin ellipses provide different sets of interpretation depending on the hypothesis on the clothing type in the current Markov chain state.

For clarity in the following description, we assume that the clothing type is short sleeve. For each skin ellipse, we extract the two extreme points of the ellipse along the major axis. These points are considered as plausible candidates for the hand-elbow pair, or elbow-hand pair of either the left or right arm. Each candidate is weighted by (*i*) skin color probability of the ellipse, (*ii*) likelihood of the arm length, (*iii*) joint probability of the elbow, hand positions with one of the shoulder candidates (For each ellipse, we find the best shoulder candidate that provides the highest joint probability.)

### 5.4 Proposal Maps

In this section we present the new concept of proposal maps. Proposal maps are generated from image observation to represent the proposal distributions of the image positions of body joints. For this discussion, we focus on the generation of a proposal map for the left hand. Using the skin ellipse cues presented earlier, we generate a set of hypotheses on the left hand position, $\{I_{L\_Hand,i}, i = 1, \ldots, N_h\}$, where $N_h$ is the number of hypotheses. Each hypothesis has an associated weight $w_{L\_Hand,i}$ and a covariance matrix $\Sigma_{L\_Hand,i}$ representing the measurement uncertainty. From each hypothesis, the proposal distribution for the left hand image position is given by:

$$q(u_{L\_Hand}|I_{L\_Hand,i}) \propto w_{L\_Hand,i} N(u_{L\_Hand}, I_{L\_Hand,i}, \Sigma_{L\_Hand,i}). \quad (13)$$

Contributions of all the hypotheses are combined as follows:

$$q(u_{L\_Hand}|\{I_{L\_Hand,i}\}) \propto \max_i q(u_{L\_Hand}|I_{L\_Hand,i}). \quad (14)$$

As the hypotheses are, in general, not independent, we use the *max* function instead of the *summation* in Equation (14); otherwise peaks in the proposal distribution would be overly exaggerated. This proposal distribution is unchanged throughout the MCMC process. To improve efficiency, we approximate the distribution as a discrete space with samples corresponding to every pixel position. This same approach is used to combine multiple observations for other body joints. Figure 3 shows the pseudo-color representation of the proposal maps for various body joints. Notice that the proposal maps have multiple modes, especially for the arms, due to ambiguous observations and image clutter.

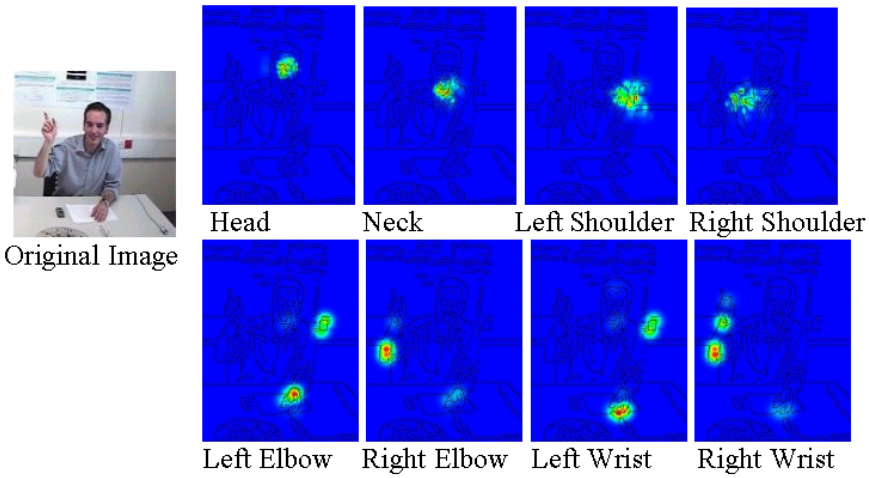



**Fig. 3.** Proposal maps for various body joints. The proposal probability of each pixel is illustrated in pseudo-color (or grey level in monochrome version).

## 6 Image Likelihood Measure

The image likelihood $P(I|m)$ consists of two components: ($i$) a region likelihood, and ($ii$) a color likelihood. We have opted for an adaptation of the image likelihood measure introduced in [14].

**Region Likelihood.** Color segmentation is performed to divide an input image into a number of regions. Given a state variable $m$, we can compute the corresponding *human blob* in the image. Ideally, the human blob should match to the union of a certain subset of the segmented regions.

Denoting $\{R_i, i = 1, \ldots, N_{region}\}$ as the set of segmented regions, $N_{region}$ is the number of segmented regions and $H_m$ the human blob predicted from the state variable $m$. For the correct pose, each region $R_i$ should either belong to the human blob $H_m$ or to the *background* blob $\bar{H}_m$. In each segmented region $R_i$, we count the number of pixels that belong to $H_m$ and $\bar{H}_m$.

$$\begin{array}{l} N_{i,human} = \text{count pixels } (u,v) \text{ where } (u,v) \in R_i \text{ and } (u,v) \in H_m, \\ N_{i,background} = \text{count pixels } (u,v) \text{ where } (u,v) \in R_i \text{ and } (u,v) \in \bar{H}_m. \end{array} \tag{15}$$

We define a binary label, $l_i$ for each region and classify the region, so that

$$l_i = \begin{cases} 1 \text{ if } N_{i,human} \geq N_{i,background} \\ 0 \qquad\qquad \text{otherwise} \end{cases} \tag{16}$$

We then count the number of *incoherent* pixels, $N_{incoherent}$, given as:

$$N_{incoherent} = \sum_{i=1}^{N_{region}} (N_{i,background})^{l_i} (N_{i,human})^{1-l_i}. \tag{17}$$

The region-based likelihood measurement is then defined by:

$$L_{region} = \exp(-\lambda_{region} N_{incoherent}) \tag{18}$$

where $\lambda_{region}$ is a constant determined empirically using a Poisson model.

**Color Likelihood.** The likelihood measure expresses the difference between the color distributions of the human blob $H_m$ and the background blob $\bar{H}_m$. Given the predicted blobs $H_m$ and $\bar{H}_m$ , we compute the corresponding color distributions, denoted by $d$ and $b$. The color distributions are expressed by normalized histograms with $N_{histogram}$ bins. The color likelihood is then defined by:

$$L_{color} = \exp(-\lambda_{color} B_{d,b}^2) \tag{19}$$

where $\lambda_{color}$ is a constant and $B_{d,b}$ is the Bhattachayya coefficient measuring the similarity of two color distributions and defined by:

$$B_{d,b} = \sum_{i=1}^{N_{histogram}} \sqrt{d_i b_i}. \tag{20}$$

The combined likelihood measure is given by :

$$P(I|m) = L_{region} \times L_{color}. \tag{21}$$

## 7 MCMC and Proposal Distribution

We adapted the data-driven MCMC framework [13]. which allows the use of image observations for designing proposal distribution to find region of high density efficiently. At the $t$-th iteration in the Markov chain process, a candidate $m'$ is sampled from $q(m_t|m_{t-1})$ and accepted with probability,

$$p = \min\left\{1, \frac{p(m'|I)q(m_{t-1}|m')}{p(m_{t-1}|I)q(m'|m_{t-1})}\right\}. \tag{22}$$

The proposal process is executed by three types of Markov chain dynamics described in the following.

**Diffusion Dynamic.** This process serves as a local optimizer and the proposal distribution is given by:

$$q(m'|m_{t-1}) \propto N(m', m_{t-1}, \Sigma_{diffusion}) \tag{23}$$

where the variance $\Sigma_{diffusion}$ is set to reflect the local variance of the posterior distribution, estimated from training data.

**Proposal Jump Dynamic.** This jump dynamic allows exploratory search across different regions of the solution space using proposal maps derived from observation. In each jump, only a subset of the proposal maps is used. For this discussion, we focus on observations of the left hand. To perform a jump, we sample a candidate of the hand position from the proposal map:

$$\hat{u}_{L\_hand} \sim q(u_{L\_hand}|\{I_{L\_hand,i}\}). \tag{24}$$

The sampled hand image position is then used to compute, via inverse kinematics (IK), a new state variable $m'$ that satisfies the following condition:

$$f_i(m') = \begin{cases} f_i(m_{t-1}) \text{ where } j \neq L\_hand \\ \hat{u}_{L\_hand} \text{ where } j = L\_hand \end{cases} \tag{25}$$

where $f_i(m_{t-1})$ is the deterministic function that generates image position of a body joint, given the state variable. In other words, IK is performed by keeping other joint positions constant and modify the pose parameters to adjust the image position of the left hand. When there are multiple solutions due to depth ambiguity, we choose the solution that has the minimum change in depth. If $m'$ cannot be computed (e.g. violate the geometric constraints), then the proposed candidate is rejected.

**Flip Dynamic.** This dynamic involves flipping a body part (i.e. head, hand, lower arm or entire arm) along depth direction, around its pivotal joint [10]. Flip dynamic is balanced so that forward and backward flips have the same proposal probability. The solution candidate $m'$ is computed by inverse kinematics.

## 8 Experimental Results

We used images of indoor meeting scenes as well as outdoors images for testing. Ground truth is generated by manually locating the positions of various body joints on the images and estimating the relative depths of these joints. This data set is available at http://www-scf.usc.edu/~munlee/PoseEstimation.html.

### 8.1 Pose Estimation.

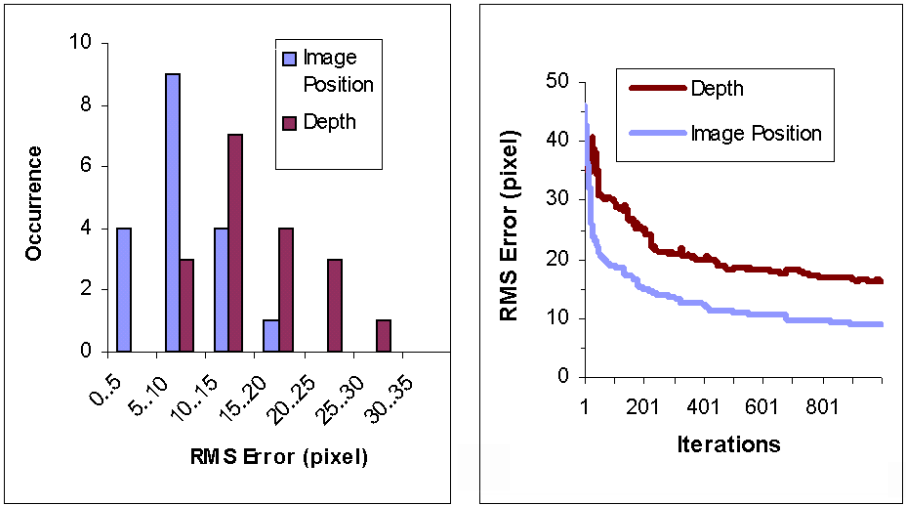
Figure 4 shows the obtained results on various images. These images were not among the training data. The estimated human model and its pose (solutions with the highest posterior probability) are projected onto the original image and a 3D rendering from a sideward view is also shown.

The estimated joint positions were compared with the ground truth data, and a RMS error was computed. Since the depth had higher uncertainties, we computed two separate measurements, one for the 2D positions, and the other for the depth. The histograms of these errors (18 images processed) are shown in Figure 5a. This set of images and the pose estimation results are available at the webpage: http://www-scf.usc.edu/~munlee/images/upperPoseResult.htm.

**Fig. 4.** Pose Estimation. First Row: Original images, second row: estimated poses, third row: estimated poses (side view).

### 8.2 Convergence Analysis.

Figure 5b shows the RMS errors (averaged over test images) with respect to the MCMC iterations. As the figure shows, the error for the 2D image position decreases rapidly from the start of the MCMC process and this is largely due to the observation-driven proposal dynamics. For the depth estimate, the kinematics flip dynamic was helpful in finding hypotheses with good depth estimates. It however required a longer time for exploration. The convergence time varies considerably among different images, depending on the quality of the image observations. For example, if there were many false observations, the convergence required a longer time. On average, 1000 iterations took about 5 minutes.




(a) (b)

**Fig. 5.** Results: (a) Histogram of RMS Error (b) Convergence Analysis.

## 9 Conclusion

We have presented an MCMC framework for estimating 3D human upper body pose in static images. This hypothesis-and-test framework uses a generative model with domain knowledge such as the human articulated structure and allows us to formulate appropriate prior distributions and likelihood functions, for evaluating samples in the solution space.

In addition, the concern with high dimensionality and efficiency postulates that the searching process should be more driven by image observations. The data-driven MCMC framework offers the flexibility in designing proposal mechanism for sampling the solution space. Our technique incorporates multiple cues to provide robustness.

We introduce the use of proposal map, which is an efficient way of consolidating information provided by observations and implementing proposal distributions. Qualitative and quantitative results are presented to show that the technique is effective over a wide variety of images. In future work, we will extend our work to full body pose estimation and video-based tracking.

**Acknowledgment.** This research was partially funded by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, under Cooperative Agreement No. EEC-9529152. We would like to thank the PETS workshop committee for providing images on meeting scene.

## References

1. Barron, C., Kakadiaris, I. A.: Estimating anthropometry and pose from a single image. CVPR 2000, vol.1, pp. 669-676.
2. Bregler, C., Malik, J.: Tracking people with twists and exponential maps. CVPR 1998, pp. 8-15.
3. Deutscher, J., Davison, A., Reid, I.: Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. CVPR 2001, vol. 2, pp. 669-676.
4. Ioffe, S., Forsyth, D.A.: Probabilistic methods for finding people. IJCV 43(1), pp.45-68, June 2001.
5. Lee, M. W., Cohen, I.: Human Body Tracking with Auxiliary Measurements. AMFG 2003, pp.112-119.
6. Mori, G., Malik, J.: Estimating Human Body Configurations using Shape Context Matching. ECCV 2002, pp 666-680.
7. Ronfard, R., Schmid, C., Triggs, B.: Learning to parse pictures of people. ECCV 2002, vol. 4, pp. 700-714.
8. Rosales, R., Sclaroff, S.: Inferring body pose without tracking body parts. CVPR 2000, vol.2, pp. 721-727.
9. Sminchisescu, C., Triggs, B.: Covariance Scaled Sampling for Monocular 3D Body Tracking. CVPR 2001, vol.1, pp. 447-454.
10. Sminchisescu, C., Triggs, B.: Kinematic Jump Processes for Monocular Human Tracking. CVPR 2003, vol.1, pp. 69-76.
11. Taylor, C.J.: Reconstruction of articulated objects from point correspondences in a single uncalibrated image. CVIU 80(3): 349-363, December 2000.
12. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. CVPR 2001, vol.1 , pp.511-518.
13. Zhu, S., Zhang, R., Tu, Z.: Integrating bottom-up/top-down for object recognition by data driven Markov chain Monte Carlo. CVPR 2000, vol.1, pp.738-745.
14. Zhao, T., Nevatia, R.: Bayesian Human Segmentation in Crowded Situations. CVPR 2003, vol.2 pp.459-466.