World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:7, 2008

Human Body Configuration using Bayesian Model

Rui. Zhang, and Yiming. Pi

Open Science Index, Computer and Information Engineering Vol:2, No:7, 2008 publications.waset.org/7193/pdf

***Abstract*—**In this paper we present a novel approach for human Body configuration based on the Silhouette. We propose to address this problem under the Bayesian framework. We use an effective Model based MCMC (Markov Chain Monte Carlo) method to solve the configuration problem, in which the best configuration could be defined as MAP (maximize a posteriori probability) in Bayesian model. This model based MCMC utilizes the human body model to drive the MCMC sampling from the solution space. It converses the original high dimension space into a restricted sub-space constructed by the human model and uses a hybrid sampling algorithm. We choose an explicit human model and carefully select the likelihood functions to represent the best configuration solution. The experiments show that this method could get an accurate configuration and timesaving for different human from multi-views.

***Keywords*—**Bayesian framework, MCMC, model based, human body configuration.

I. Introduction

HUMAN Body configuration based on the Silhouette is the key point for automatic recognition of human activities and image understanding. It has a wide use in the surveillance. For certain applications some methods have been presented. For example, Pictorial structures have been brought up by Felzenszwalb [1], although it performs well in certain situation, it still has drawbacks such as its time-consuming and the need for a generic model, which restricts its wider use in different human and multi views. 2D or 3D structure model has also been presented and the body pose is recovered by extracting image features and mapping them to structural components of the model [2] [3] [4] but it needs a articulate model for a special person and need much train data for this specific person which restrict its wide application in real system for unknown person and quick movement. There are other kinds of methods that directly model how the person moves through the spatiotemporal distribution generated by the motion [5] [6]. These methods only consider the contour projection of the body so we can not get some detail configuration information. Rosales and Sclaroff inferred the body poses without tracking body parts [7], they tried to map low-level features to high level representation (configuration) using the machine learning strategy but there results are not so good.

Manuscript received September 12, 2005.

Rui zhang is with the department of electronics engineering of University of Electronic Science and Technology of China (UESTC) Chengdu, 610054 (e-mail:ruizhang331@yahoo.com.cn).

Yiming pi is with the department of electronics engineering of University of Electronic Science and Technology of China (UESTC) Chengdu, 610054 (e-mail:ympi@uestc.edu.cn).

In this article we propose to address this problem basing on the bayesian framework and the best configuration could be defined as MAP (maximize a posteriori probability) in Bayesian model. It considers both prior knowledge and the likelihood Model. So we transfer our problem into a optimization method.. But the computation of the global optimum is NP-hard and, thus, intractable. The important question concerning the approximation of the problem arises: How good is a computed minimizer relative to the unknown global optimum? Can a certain quality of solutions in terms of its sub-optimality be guaranteed in each application? None of the approaches seems to be immune against getting trapped in some poor local minimum and, hence, doses not meet these criteria.

In this case, we consider fitting the Bayesian model requires estimating the distribution of the parameters (considered random) conditional on the observed data. MCMC (Markov Chain Monte Carlo) [8] is a widely used method to estimate the MAP solution in a high dimension space by drawing samples from the posterior distribution using Markov chain. We can estimate virtually all summaries of interests directly from the simulation. In our works, we use a special human model based sampling. This hybrid sampling algorithm combines the Gibbs sampling and Metropolis Hasting sampling to get a higher performance than a single sampler because it utilizes the human model which could converse the original high dimension solution space into a restricted sub-space constructed by the structure. The experiments show that the model based MCMC could get a higher efficiency both in the performance and time-saving.

This paper is organized as follows: Section 2 gives an introduction about the MCMC method under the Bayesian model and the two major sampling algorithms; Section 3 describes the model based strategy; section 4 presents the human model and likelihood model. Markov dynamics and the results are presented in section 5; section 6 gives the conclusion for this paper.

II. Markov Chain Monte Carlo Framework under the Batesian Model

In this section we describe the MCMC framework. Based on the upper discussions we know that the problem is to find the solution which maximizes a posteriori probability. We formulate this model based human configuration problem as computing the maximum posteriori probability under the Bayesian model:

$$\theta = \arg\max P(\theta \mid I) \tag{1}$$

Open Science Index, Computer and Information Engineering Vol:2, No:7, 2008 publications.waset.org/7193/pdf

When θ is assemble of human model and its parameters, I contains the image observation. Following the bayesian rule the posteriori probability is decomposed into a likelihood term and a prior term:

$$P(\theta \mid I) \sim P(I \mid \theta)P(\theta) \tag{2}$$

Fitting the Bayesian model requires estimating the distribution of the parameters (considered random) conditional on the observed data. Let us assume that instead of wanting to approximate $p(x)$, we want to find its global maximum. For example, if $p(x)$ is the likelihood or posterior distribution, we often want to compute the ML and maximum a posteriori (MAP) estimates. As mentioned earlier, for the optimization problem, one of the most important design decision concerns compromise between the adequacy of the optimization criterion and the difficulty in computing the solution. But the computation of the global optimum is NP-hard. Many current optimization methods, such as steep decent, will be trapped in a local peak especially the solution space has multiple peaks. Another optimization approach is based on stochastic sampling. We could run a Markov chain of invariant distribution $p(x)$ and estimate the global model.

MCMC is a suitable methodology for finding the solution by drawing samples from the posteriori distribution using a Markov chain. There are two major reasons for MCMC's success [8]. Firstly: MCMC provides a systematic way, firmly rooted in Bayesian estimation theory, for including prior constraints about the shape and average size of homogenous regions in an image; A wide variety of behaviors may be obtained, simply by varying a few parameters in the definition of local potentials in the MCMC model. The second reason is that, even when exact optimal estimators cannot be precisely computed, it is possible to design reasonable approximate algorithms that work well in many cases.

A. Metropolis-Hasting Algorithm

There are two major algorithms to simulate sampling from the posteriori distribution. The first one is the Metropolis Hasting algorithm [9]. At each iteration t, we sample a candidate state θ from a proposal distribution

$$q(\theta_2 \mid \theta_1) = P_r(\theta_1 \to \theta_2) \tag{3}$$

The new state θ' is accepted with the following probability:

$$P = \min(1, \frac{Q(\theta')q(\theta_t \mid \theta')}{Q(\theta t)q(\theta' \mid \theta t)}) \tag{4}$$

Where $Q(\theta) = p(\theta \mid I)$. The Markov chain constructed in this way has its stationary distribution in $p(\theta \mid I)$ and independent from the initial state [11]. This algorithm has its advantage that it could search for a wide range but it has some drawbacks that when the dimension of the solution space is high its efficiency will drop [12].

B. Gibbs Algorithm

Gibbs Sampling is a special case of the Metropolis Hasting sampling [13] that one simulate n random variables sequentially from the n univariate conditionals rather than generating a single n-dimensional vector in a single pass using the full joint distribution, the value of the k variable is drawn from the distribution $p(\theta^{(k)} \mid \Theta^{(-k)})$, where $\Theta^{(-k)}$ denote a vector containing all of the variable but k. During the ith iteration, we draw from the distribution to obtain $\theta_i^{(k)}$:

$$\theta_i^{(k)} \sim p(\theta^{(k)} \mid \theta^{(1)} = \theta_i^{(1)}, \cdots, \theta^{(k-1)} = \theta_i^{(k-1)}, \theta^{(k+1)} = \theta_{i-1}^{(k+1)}, \cdots, \theta^{(n)} = \theta_{i-1}^{(n)}) \tag{5}$$

The random value is always accepted in the Gibbs sampling. The key to the Gibbs sampling is that one only consider univariate conditional distribution. The Gibbs sampling has the advantage that can use some conditional information and give the random search some help but it converges too slow especially when the dimension is high.

III. Model Based Strategy

In this section we discuss our model based strategy. This algorithm uses the human structure features to simulate the sampling. It converses the original high dimension solution space into a restricted sub-space constructed by the structure features and drawing a hybrid sampling combining the metropolis hasting algorithm and the Gibbs sampling.

We know that people have a great variety of appearances, gestures, activities, scales and illuminations. So we can not use a single model to match the human body. The feature such as color, edge and texture is not suitable for depicting human. Some efficient features used in the face detection such as Gabor based or haar based features have been used in the human detection, but these applications require stabile gestures and views which is not all satisfactory in practice. But all human body has some common structures in a certain configuration. These structure features are robust due to that people have the stabile skeleton and body parts configurations. We conclude two major structure features as follows:

1) Geometric Relations:

Geometric relations are robust features because all human body components are constructed in certain configuration. For example, the shank is related with thigh by the knee and the head lies in the up of the torso. So the parameters of a single body component are not independent but restricted by other components. A natural way to express them is in terms of a set G = (V, R), V corresponds to the body components and R is a set $((V_i, V_j) \mid \forall V_i \in R)$ which represents all geometric relations.

2) Geometric Proportions

Geometric Proportions are other important features because the components of human body obey some proportional relationships. For example, the height of the torso and head equals to the height of legs approximately. for a special human component V_i, we construct the Proportion feature as

Open Science Index, Computer and Information Engineering Vol:2, No:7, 2008 publications.waset.org/7193/pdf

$F = (V_i^i, V_i^j)$ .where the V_i^i represents its inner proportions such as the ratio of Length and width, and the V_i^j represents the correlation proportions between the i and j component.

Based on these features we could use a more efficient search strategy in the solution space called model based MCMC. We use a hybrid sampling algorithm to obtain the value θ from the distribution.

$$\theta_i \square P(\theta_i \mid Y_i) \tag{6}$$

And $Y_i = (Y_i^1, Y_i^2, \cdots, Y_i^n)$ are the model restrictions at ith iteration which make the random metropolis-hasting sampling into a conditional sampling. The transfer from the state θ_{i-1} to θ_i becomes:

$$\theta_i = \theta_{i-1} + (\Delta\theta \mid Y_i) \tag{7}$$

And we choose the proposal distribution that is normally centered on the current value $q(\theta_i \mid \theta_{i-1}) = q(\theta_{i-1} \mid \theta_i)$.

The new state is accepted by the probability that:

$$P = \min(1, \frac{p(\theta_i \mid I, Y_i)}{p(\theta_{i-1} \mid I, Y_i)}) \tag{8}$$

This algorithm makes the random search becomes the conditional search. It makes the Markov Chain dynamic jumps in a restricted sub-space constructed by the structure features rather than a totally unknown high dimension solution space; even we assume we do not know the proposal distribution. So it could greatly reduce the search range and reach the optimal results using less iteration.

IV. Generative Model

A. Human Model and Hybrid Strategy

We choose a human model that considers both the explicit representation and searching dimensions. As the human has various appearances, we do not use the color model or shape model, only consider the structure information.

Our model includes two stages: the coarse stage and the fine stage. In coarse stage we choose a low dimension simple model which includes three parts: head, torso, and legs .We present each of them with a rectangle and the parameters of each component include positions, height and fatness, represented as M ={x, y, h, f}. So the DOF (dimension of freedom) of the human model is 12. The coarse stage model has a moderate dimensions for the MCMC to simulate but it can not work well when the human is in fast walk or some other gesture when the two legs are separated. In this case one rectangle can not depict the legs explicitly. Instead we use a fine model for legs which is constructed by four rectangles representing two thighs and two shanks respectively, The parameters of the leg model include positions, rotation ,height and fatness, represented as M = {x, y, θ ,h, f}. So the DOF of the leg model is 20. We use the coarse model to simulate our Markov Chain in the first half of the iterations and the fine model to simulate for the rest iterations. In coarse stage we simulate with large steps and get the head and torso parameters, then in the fine stage we focus our search on the leg model, only change the head and torso parameters slightly with a low probability.

There are two reasons for us to choose the two stage hybrid strategy:

Firstly, a very powerful property of MCMC is that it is possible to combine several samplers into mixtures and cycles of the individual samplers [11]. If the transition kernels $K1$ and $K2$ have invariant distribution $p(\bullet)$ each, then the *cycle hybrid kernel* $K1K2$ and the *mixture hybrid kernel* $\nu K1 + (1 - \nu)K2$, for $0 \leqslant \nu \leqslant 1$, are also transition kernels with invariant distribution $p(\bullet)$.

Secondly, we note a tradeoff between the model dimension and the MCMC efficiency [12]. The MCMC efficiency drops reciprocally with number of dimensions but the human body is so highly articulate that it requires high dimension to model. The two stage model could solve it efficiently and the results show that it performs well. We do not construct the arm model because the arms are invisible in some case especially in profile views.

B. Likelihood Model

There is no general way to choose likelihood model. For specific application the principle is to choose the cue that is most discriminative compared to the rest of the scene. The first factor is its optimality; the desired solution should give the highest value. Another factor is that the likelihood should be realistically peaked, which means we do not have exaggerate peaks.

We choose the Silhouette based model in our application. Basically this likelihood reflects the number of wrongly classified pixels according to the current state. Body silhouette extraction is achieved by simple background subtraction and thresholding [6]. The wrongly classified pixels include the false positives refer to that do not belong to human but within the hypothesis and true negatives refer to that belong to human but are not within any hypothesis object [10]. That is:

$$p(I \mid \theta) = e^{-\lambda(N_{fp} + N_{tn})} \tag{9}$$

λ is a coefficient which controls the number of correct pixels a human hypothesis has to contribute. It is mainly decided by the size of human which the application is interested in.

We assume the prior distribution is flat in our application and the posteriori probability is equal to the likelihood model. $p(\theta \mid I) = p(I \mid \theta)$.

V. Human configuration using Model based MCMC

A. Markov Dynamics

We use the following Markov dynamics to simulate our experiments:

- **Change**: (only used in the coarse stage): we use three basic visual angles: {*frontal (back); partly profile, totally profile*}. And we can randomly select a visual

Open Science Index, Computer and Information Engineering Vol:2, No:7, 2008 publications.waset.org/7193/pdf

angle, the human body parts parameters are changed respectively using the structure features.

- **Diffusion**: We randomly select a human component and change the x and y positions in some random diffusions to get the component position changed, so the new human position is: $(x+\Delta x, y+\Delta y)$
- **Scale**: We randomly select a human component and scale the height and width of a component in the ratio of 0.9 or 1.1 compared with the former one and get a new hypothesis.
- **Rotation**: (only used in the fine stage): We randomly select a leg part and rotate it in a randomly chosen angle, notes that if we choose the thigh to rotate the respective shank must be changed according to the joint coordinates.

In each iteration, one of these is chosen randomly according to predefined probabilities and the sum of the probabilities equal to 1. For coarse stage it is:

$p_{change} + p_{diffusion} + p_{scale} = 1$ and for fine stage it is

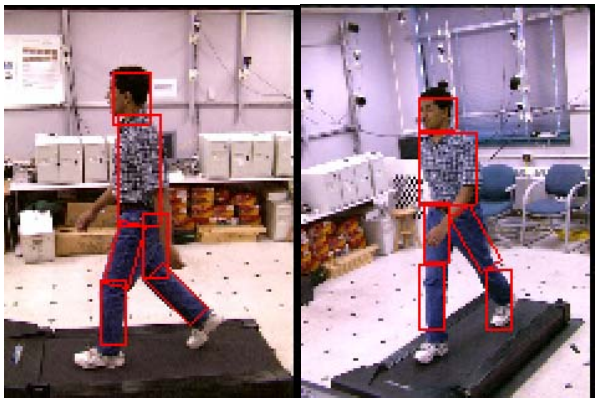
$p_{rotation} + p_{diffusion} + p_{scale} = 1$ .The above dynamics guarantee the Markov chain is irreducible (any state is reachable within finite iterations) and aperiodic (Markov chain is not periodic) since all of them are stochastic.

B. Experiments

We use the CMU Motion of body (Mobo) database to demonstrate our methods. The Mobo database contains 25 individuals walking on a treadmill. The subjects perform four different walk patterns: slow walk, fast walk, incline walk and walking with a ball. All subjects are captured from six views. The sequences are each 11 seconds long, recorded at full frame rate (30 frames/second), capturing more than 8000 pictures with a resolution of 640x480.

We do our experiments as follows: First we extract the silhouette and then we initialize the human model in a standard upright pose in the center of the image with the initial parameters. We assume the prior distribution is flat and the choice of proposal distribution is arbitrary.

The test was set after about 1000 iterations in about 100ms on a Pentium IV 2.4G Hz PC with un-optimized C++ code. Fig. 1 shows the human configuration results in two different individuals from multi-views. More results are shown in the Fig. 2.



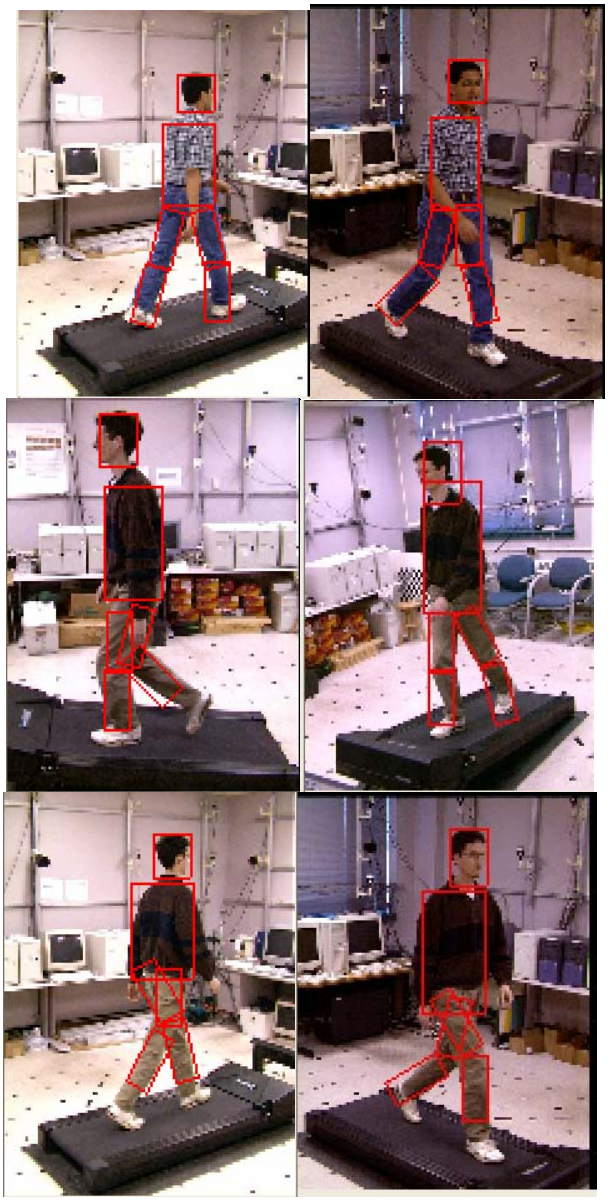
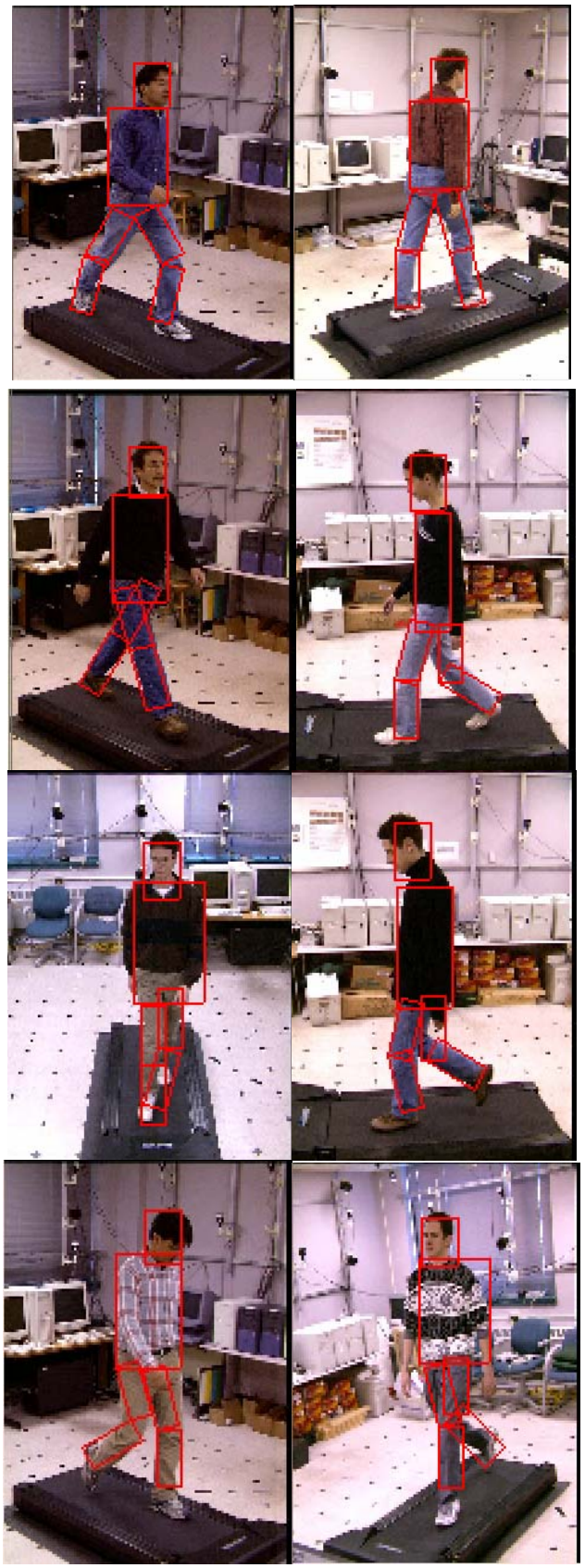


Fig. 1 Human configuration results in two different individuals from multi-views

Open Science Index, Computer and Information Engineering Vol:2, No:7, 2008 publications.waset.org/7193/pdf



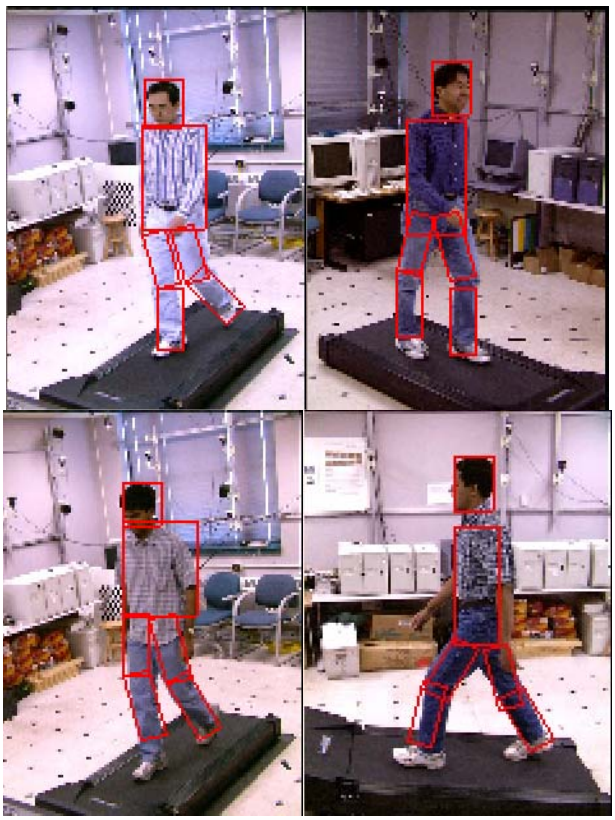


Fig. 2 Other Human body configuration results

We also show our experiment results in 1000 frames randomly selected from the database that contain all 25 individuals and 6 views in Fig. 3. The results show that the posteriori probabilities are very high (most of them are between 0.8 and 0.9).Considering that the rectangle representation for body parts can not match the silhouette precisely because we do not include the arm model and the shape of each part is not exactly the rectangle. Our methods could get a suitable result for human configuration.

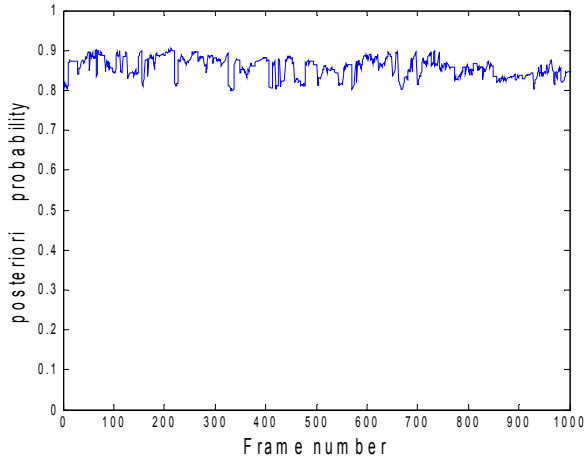


Fig. 3 posteriori probabilities in 1000 frames randomly selected from the database that contain all 25 individuals and 6 views. The results show that the posteriori probabilities are very high (most of them are between 0.8 and 0.9).Considering that the rectangle representation can not match the silhouette precisely because we do not include arm model and the shape of each part is not exactly the rectangle.

Open Science Index, Computer and Information Engineering Vol:2, No:7, 2008 publications.waset.org/7193/pdf

We show a 2-D projected distribution of the posteriori probabilities in the following Fig. 4. The X, Y labels represent the torso positions and the Z labels represent the posteriori probabilities using our likelihood model. We can see from this figure that the distribution of the values complies GMM(Gaussian mixture model), which comprises a number of component functions and has multiple peaks. In this case, most optimization method could be trapped in a local peak. So they are all not suitable for our problem. In contrast, the MCMC could estimate the distribution of such multiple peaks.

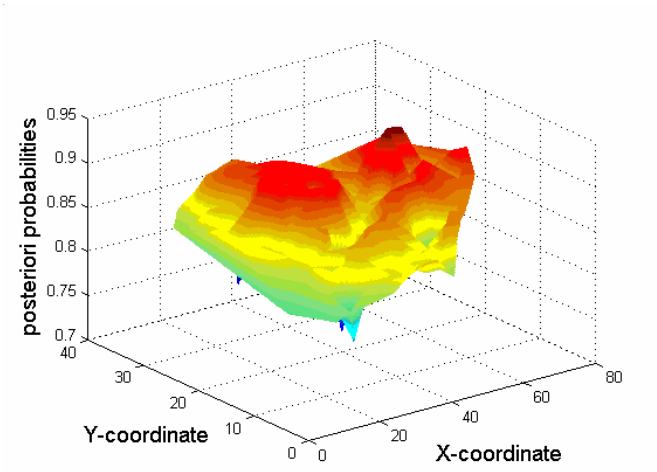


Fig. 4 2-D projected distribution of the posteriori probabilities. The X, Y labels represent the torso positions and the Z labels represent the posteriori probabilities. Note that the distribution of the values complies GMM (Gaussian mixture model), which comprises a number of component functions and has multiple peaks. In this case, most optimization method could be trapped in a local peak. So they are all not suitable for our problem. In contrast, the MCMC could estimate the distribution of such multiple peaks.

If we can get more about the proposal probability and use some other optimization techniques such as dynamic programming we can get a real-time configuration.

VI. Conclusion and Future Work

We have presented an approach using a model based MCMC for human segmentation in static images. The model based algorithms use the human structure features to form a hybrid sampling algorithm to simulate the MAP distributions. It converses the original high dimension solution space into a restricted sub-space constructed by the structure features. We choose a proper human model and evaluation function to construct the likelihood model. We use the Mobo database to demonstrate that this algorithm is effective on a wide variety of images and views.

A. Future Work

Our current system has some limitations and future works could be improved in various aspects since the MCMC is an open framework.

1) The speed of the Markov chain strongly depends on the proposal probability and we can use some domain knowledge to compute the proposal probability. Furthermore, we assume that the prior distribution is totally unavailable. In special application we can get some prior distributions which could be helpful for our search.
2) Currently we do not use any other low level features such as edge or color, although they are not robust for a wide range. In some particular applications the human may have common appearances or shapes. We can use these features in those cases which may be helpful both in the efficiency and time-saving.
3) We can implement our model based strategy using some other knowledge such as kinematics model and distance transform to increase the sampling algorithm. We can use some dynamic programming techniques to increase the speed.

References

[1] P.F.Felzenswalb and D.P.Huttenlocher, "Efficient matching of pictorial structure" in *Proceeding of the computer vision and pattern Recognition*, 2000, pp. 66-73.
[2] K.Rohr, "Towards model based recognition of human movements in image sequence" in *CVGIP*, 1994, vol.59
[3] Y.Song, X.Feng and P.perona, "Towards detection of human motion". in *Proceeding of the computer vision and pattern Recognition*, 2000
[4] D.M.Gavrila and L.Davis, "Towards 3-D model-based Tracking and recognition of Human Movement: a Multi-view Approach" *In IEEE international conference on automatic Face and Gesture Recognition.*1995
[5] S.Niyogi and E.Adelson, "Analyzing and recognizing walking figures in XYT", in *Proceeding of the computer vision and pattern Recognition.* 1994, pp. 469-474.
[6] Robert.T.Collins, Ralph Gross and Jianbo shi, "Silhouette-based Human Identification from Body Shape and Gait" *In IEEE international conference on automatic Face and Gesture Recognition 2002*
[7] R.Rosales and S.Sclaroff, "Inferring Body Pose without Tracking Body Parts". in *Proceeding of the computer vision and pattern Recognition*, 2000.
[8] C.Andrieu, N.De.Freitas,A.Doucet and M.jordan, "An Introduction to MCMC for Machine Learning". *Machine Learning*, 50, 5-43, 2003
[9] W.K.Hasting, "Monte Carlo Sampling Methods Using Markov Chain and Their Application". Biometrika, 1970.
[10] T. Zhao and R. Nevatia. "Bayesian Human Segmentation in Crowded Situations", *in Proceeding of the computer vision and pattern Recognition,* pp.459-466. 2003
[11] W.Gilk, S.Richardson and D.Spiegelhalter, *Markov Chain Monte Carlo in Practice* Chapman and Hall, 1996
[12] B.Walsh, *Markov Chain Monte Carlo and Gibbs sampling*. Lecture Notes for EEB 596z, 2002.
[13] M. Hason "Tutorial on Markov Chain Monte Carlo" Technical report, Los Alamos National laboratory, 2000.

Rui. Zhang received the B.S degree from the department of information engineer at Xi'An Jiao Tong university. P.R.China, in 2003. Currently, he is a master student in the University of Electronic Science and Technology of China. P.R.China His research interests include the computer vision, image understanding, and the hardware realization of image process system.

Yiming. Pi received his B.S degree from the department of electronic engineer in Harbin Engineering University Institute, 1987, and his M.S and PhD degree from the University of Electronic Science and Technology of China, in 1990 and 1993 respectively. He has been the professor in UESTC since 2002. He has published more than 30 papers in international journal and conferences, half of them are indexed by SCI and EI. His research interests include Radar image processing, remote sensing and GPS.