

# Extraction of Parametric Human Model for Posture Recognition Using Genetic Algorithm

Changbo HU   Qingfeng YU   Yi LI   and   Songde MA  
National Lab of Pattern Recognition, Institute of Automation,  
Chinese Academy of Sciences  
P.O.Box 2728, Beijing, 100080, China  
Cbhu, qfyu, liyi, masd@nlpr.ia.ac.cn

## Abstract

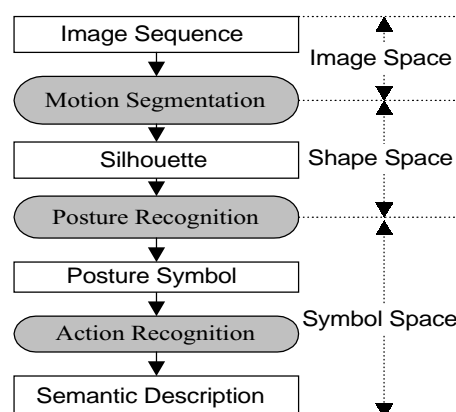
*We present in this paper an approach to extract human parametric 2-D model for the purpose of estimating human posture and recognizing human activity. This task is done in two steps. In the first step, human silhouette is extracted from complex background under a fixed camera through a statistical method. By this method, we can reconstruct the background dynamically and obtain the moving silhouette. In the second step, genetic algorithm is used to match the silhouette of human body to a model in parametric shape space. In order to reduce the searching dimension, a layer method is proposed to take the advantage of human model. Additionally we apply structure-oriented Kalman filter to estimate the motion of body parts. Therefore initial population and value in GA can be well constrained. Experiments on real video sequences show that our method can extract human model robustly and accurately.*

## 1. Introduction

Recognition of human activity is an attractive goal for computer vision [1,15]. Applications of human activity recognition include video indexing, efficient video coding, visual surveillance, virtual reality, human-machine interaction and athletic performance analysis, etc. Posture analysis is the base of activity analysis. Obtaining human posture robustly and accurately is in great demand.

Generally, posture recognition is a two-step task. The first step is to estimate the relative position parameters of important parts of human body, and the second step is the semantic classification of all possible position arguments. Because of the simple way human body is connected, a posture classification tree (PCT) corresponding to the

connected structure of human body can work very well. Through this classification, postures are mapped to a symbolic space, which constructs the basic symbol set of action analysis using frames such as HMM [2,3,13]. Although a single posture symbol may not have actual meaning, a posture symbol flow will construct an action with exact meaning, which is very similar to the recognition of natural language. Figure 1 shows the framework we apply. In this paper we will focus on the first step, i.e. image space to shape space.



**Figure 1. Action Recognition: A Possible Framework**

The key point in the first step is how to model human body from image. Generally there are two typical approaches to the motion analysis of human body parts, depending on whether a priori models are used [4,12]. In each type of approach, the representation of human body evolves from stick figures to 2-D contour models to 3-D volumes as the complexity of the model increases.[5,6,15]. We notice the necessity of 3-D model of human body to do precise recognition. But 3-D reconstruction is a difficult ill-posed problem because human movement is non-rigid

and deformable. And in 3-D methods, edges often play the central role in tracking and motion estimation. Some works are done by setting markers in human joints in order to obtain stable correspondence. But this method is intrusive, and furthermore can not be used in video analysis---a large application area. Although the principle of human's ability to recognize postures is still unclear, we can discern most posture only from silhouette, even when there is significant loss of several important parts such as head or leg.

So in this paper we propose a method to match 2-D parametric model with human silhouette. It is our aim developing a posture recognition system to analyze athletic video. Genetic algorithm (GA) [7,8,9] is widely used in solving global optimization problems for its ability of avoiding converging to local minimum points. We apply GA to search suitable models in parametric space.

The paper is organized as following: Section 2 presents the technique to extract silhouette of human body. Section 3 gives the way modeling human body. Section 4 details the genetic algorithm to match silhouette. In section 5 we test our method on real data. And finally conclusions are given in section 6.

## 2. Silhouette extraction

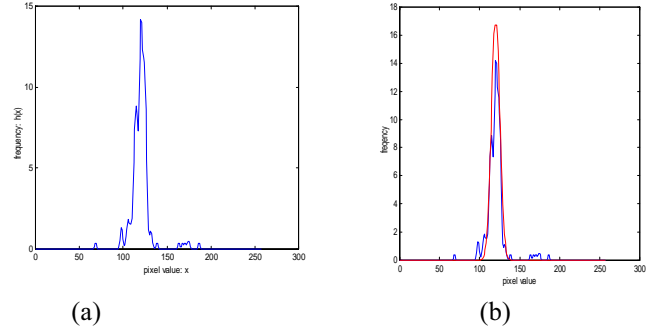
We formulate the silhouette extraction as classifying pixels in 3-D XYT space. By sampling the volume at the position  $(u_0, v_0)$ , we can get a line in the 3D space representing the recent history of the pixel at  $(u_0, v_0)$ . Then the problem of detecting moving objects may be tackled in two steps: We first process each pixel independently to restore the background, then we obtain the moving human body by background subtraction and applying adaptive threshold to subtraction result.

### 2.1. Background image restoration

For a single pixel in a video sequence captured with a static camera, we can find that the pixel often belongs to the background. Sometimes it is occluded by moving human body. The recent history  $\{X_1, \dots, X_t\}$  of each pixel can be represented by a histogram. (see Figure 2(a))

Assume that the values of pixels belonging to the background satisfy a Gaussian distribution. We then have:

$$p_b(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (2.1)$$



**Figure 2. (a):Histogram representation of a pixel's history. (b):Resultant Gaussian overlaid on the original histogram.**

What we need to know is the values of  $\mu$  and  $\sigma$ . We represent the distribution of values belonging to foreground as  $p_f(x)$ . Under the assumption that most of the time the pixel is part of the background, we regard  $p_f(x)$  as noise. Then we can estimate  $\mu$  and  $\sigma$  from the observed data [10]. Figure 2(b) shows the resultant Gaussian overlaid on the original histogram. After processing all pixels, we can restore the background image. The value of each pixel is equal to the mean value at the corresponding position. The background constructed in this way can be updated continually and moderate light variance and scene changes will not influence the silhouette extraction.

### 2.2. Silhouette acquisition

After a background image is acquired, a common strategy is to subtract with the background image.

$$\Delta_n = |I_n - B| \quad (2.2)$$

If we apply a changeless threshold to classify pixels, the result may often be wrong. For pixels similar to the background, they are often mis-classified as part of the background. Here we use  $\sigma(u, v)$  as the threshold in each pixel  $(u, v)$

$$L_n(u, v) = \begin{cases} 1, & \Delta_n(u, v) \geq 2 * \sigma(u, v) \\ 0, & \Delta_n(u, v) \leq 2 * \sigma(u, v) \end{cases} \quad (2.3)$$

Additionally, we first label all pixels as “static”. If a pixel is classified as foreground consecutively for a few frames, we label it as “moving”. If a “moving” pixel is

classified as background for several frames, its label is changed back to “static”. Thus, we can eliminate the false motion caused by scene noises, and some clutter pixels which belong to human but are similar to background can also be labelled as “moving” correctly. The silhouette extracted in this way is of higher quality.

### 3. Modeling Human Body

Fig. 5(a) shows a typical video flow of human motion. Using our silhouette extraction technique, silhouette segmented from this flow is like Fig. 5(b). Each main part of the body is represented by a rectangle  $R$ , which can be defined by a 5-elements array (Figure 3(c))

$$R = (x, y, w, h, \alpha)$$

The total DOF of  $R$  is 5.

A human body  $H$  can be represented by 10 rectangles:  $H = (R_i, i \in [0, 1, \dots, 9])$  (Fig. 3(a)(b)), as well as constraints on their parameters respectively.

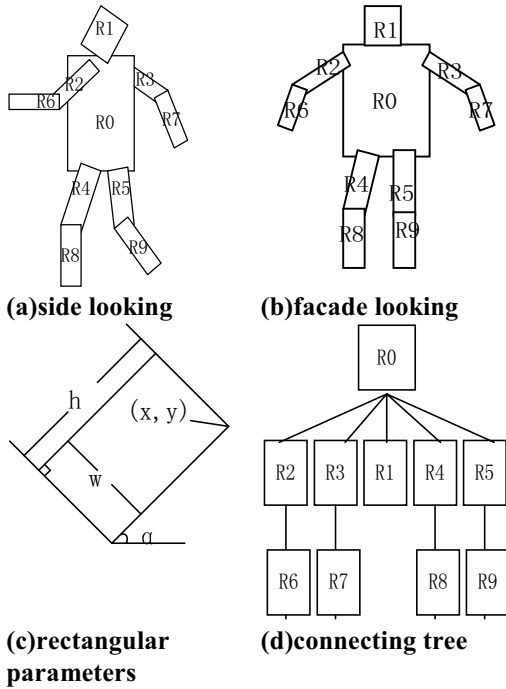


Figure 3. Human body model

We propose a Connecting Tree (CT) to describe the structural knowledge about human body. The CT is a simple 3-level tree with 10 nodes describing the connecting relation between the rectangles and the constraints on their parameters. Under the direction of CT, the  $(x, y)$  parameters of all rectangles except  $R_0$ , can be calculated, so the total DOF of our model is 32. We will

take the advantage of the CT, to simplify searching in parametric space, as discussed in section 4. In our experiment, we further simplify the model by assuming the scale relation of upper and lower limb. In matching algorithm, the upper limb and lower limb are considered together as an articulated object. One benefit of this assumption is that total DOF is 24. Another benefit is that the searching algorithm will not be confused when the upper and lower limb are collinear. We do not model hands and feet because in most silhouettes even human eyes can not discern them exactly.

### 4. Genetic algorithms for matching silhouette with parametric model

Genetic algorithm is a stochastic optimization technique which can be considered as a multi-point search. The result is less likely to fall into a local optimal solution. In our application to match human model, a layer method is proposed to take the advantage of human model in order to reduce the searching dimension.

#### 4.1. Steps of GA

The process of GA is summarized as following:

**Initialization:** Initialize the values of parameters  $N$  (size of chromosome pool),  $P_{c0}$  (initial crossover rate) and  $P_{m0}$  (initial mutation rate). Take potential parameters as genes and create  $N$  initial chromosomes randomly from the genes.

**Step 1:** Evaluate the fitness of each chromosomes in the chromosome pool.

**Step 2:** Compute  $P_c$  and  $P_m$  with different statistics attributes of chromosome pool. Selects chromosomes to perform crossover and mutation operation from chromosome pool according to the fitness, the crossover ratio and mutation ratio.

**Step 3:** Crossover and mutation operators are applied to the selected chromosomes.

**Step 4:** New chromosomes are evaluated by their fitness function value.

**Step 5:** New chromosomes are inserted into chromosome pool and bad chromosomes in fitness are eliminated through selection.

**Step 6:** If the chromosome is not convergent then go to Step 2, otherwise output the best chromosome and the optimum solution.

In our application, we do not search in the whole 24-D

parametric space. This search will be time-consuming and is most likely not to converge. We can take advantage of connecting tree of human body to construct a layer search. That is, to search the trunk rectangular first. A trunk is taken away when it is found. Then search the head from the remainder image. Finally search four limbs. After all parts are found, fitness of whole body is also computed to judge the success of the search. If the fitness is lower than a threshold, we restart the search process. ( See figure 4)

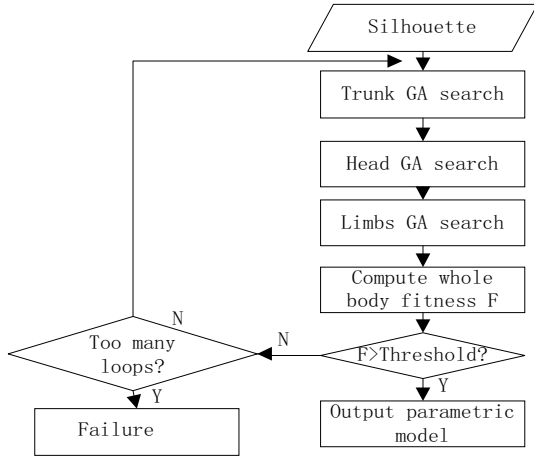


Figure 4. Layer search of human body part

## 4.2. Chromosome coding

We use potential rectangular parameters as genes. Then a set of 5 genes, which is the minimal subset for representing a independent rectangular, is taken as a chromosome. A chromosome of trunk is  $(x, y, w, h, \alpha)$ . Because the position is given by trunk and scale relationship of upper and lower limb is assumed, the chromosome of each of four limbs is  $(w, h, \alpha 1, \alpha 2)$ . And the chromosome of head is composed of only three genes, that is  $(w, h, \alpha)$ .

## 4.3. Evaluation of chromosome

We pose a quantitative operator  $S(I_1, I_2)$ , which measures the shape similarity between two binary image  $I_1, I_2$ . It is difficult to design such a shape similarity operator approximately reflecting our psychological feelings about shapes, however, special operators for specified purposes still exist and can work perfectly well. Area and space distribution, edge curvature and requirements for error control are generally considered factors. Here we introduce a similarity operator which only

consider the area difference between two shapes.

The factors affecting the similarity between two shapes  $I_1$  and  $I_2$  we consider here is the ratio of positive error  $p$  and the negative error  $n$  between shapes, defined as  $p = \Omega(I_1 \cap I_2^c) / \Omega(I_1)$  denotes the positive error;  $n = \Omega(I_2 \cap I_1^c) / \Omega(I_2)$  denotes the negative error.

Here  $I^c$  denotes the complement set of  $I$ ,  $\Omega$  denotes the operator counting non-zero elements of its argument set. We assume that  $I_1$  is model and  $I_2$  is silhouette to let us know the meaning of  $p$  and  $n$  clearly.

According to these definitions, we can simply define  $S(I_1, I_2)$  as:

$$S(I_1, I_2) = e^{-[\alpha p + (1-\alpha) \cdot n]} \cdot [\beta(1-p) + (1-\beta) \cdot (1-n)] \quad (4.1)$$

$\alpha, \beta \in [0, 1]$  are coefficients reflecting the degree of emphasis on the errors and in our silhouette recognition, which show our credibility to the result of segmentation.  $S(I_1, I_2)$  takes value between  $[0, 1]$ , and is used as fitness function.

## 4.4. Genetic Operator

**Crossover:** The crossover operation creates new chromosomes (children) from two existing chromosomes (parent) by exchanging some genes. We apply inner linear interpolation function to compute children chromosome. The chromosome is selected to perform crossover operator with probability value which is in proportion to its fitness. Therefore, fitter chromosome will have more chances to reproduce its genes. In genetic algorithms, crossover operator makes the chromosome pool tend to convergence.

**Mutation:** Mutation introduces new genes into chromosome. The chromosome is selected to perform mutation operator with the same probability. Mutation operator can prevent the chromosome pool from falling into a local optimum.

## 4.5. $\alpha - \beta - \gamma$ estimation of body parts for GA initialization

Considering the smoothness and continuity of human body's motion, a  $\alpha - \beta - \gamma$  filter is a natural choice to predict the motion of each part of human body. We apply the general framework given in [11]. But we make a constant acceleration assumption instead of constant velocity assumption, and use a  $\alpha - \beta - \gamma$  filter to predict the motion of each part of human body.

This estimation is done after the matching in every frame, then the initial value and search range formed from the result can be used in next matching. Thus the search space in GA is reduced remarkably. The motion we now face, after the normalization of the silhouette, is the continuous variance of the  $\alpha$  parameter of each rectangle.

To our experience, when a good recognition result is gotten, the  $\alpha-\beta-\gamma$  filtering can accurately predict the motion for several dozens frames, during which the matching results are very stable, and GA converges much faster than at the beginning. If great shape error occurs at some time, i.e. the fitness is very low, then the estimation of  $\alpha-\beta-\gamma$  filtering becomes invalidate, and new search is done in original parameter space.

## 5. Experiment result and discussion

There are two important issues we would like to discuss here:

1. The parameter selection in GA. The parameters of GA affect the computation speed and convergence performance. There are 6 parameters to control GA performance. They are the initial population, crossover rate, mutation rate,  $\alpha$  and  $\beta$  in fitness function and the termination fitness and generation. The parameters other than  $\alpha$  and  $\beta$  are relatively stable according to our experience. But  $\alpha$  and  $\beta$  is sensitive to the silhouette quality. If much noise appears in silhouette, we must decrease  $\alpha$  and increase  $\beta$ . If much noise such as shadow appears out of the silhouette,  $\alpha$  should be increased and  $\beta$  should be decreased. In other word,  $\alpha$  and  $\beta$  are adjusted to repress noise. In our experiment, parameters other than  $\alpha$  and  $\beta$  are only affected by motion estimation from  $\alpha-\beta-\gamma$  filter and change not much if motion estimation does not fail, but  $\alpha$  and  $\beta$  should be adjusted according to silhouette quality in each frame.

2. When occlusion occurs. In our framework, occlusion problem is not devised to solve independently. In the layer GA process, the overall fitness determines whether the search result is valid. The fitness of every part determines whether the search result of this part is valid. A low fitness means the occlusion of this part. And before the motion estimation this part is thought as falling in a default angle.

Figure 5 is a tentative experiment using a real video data. In this video a man is walking through scene of the camera in different directions. In figure 5, images in line 1 and 4 are original frames; line 2 and 5 are silhouettes extracted by our method; and line 3 and 6 are the

parametric models searched by GA. Models can always be searched out in this sequence.

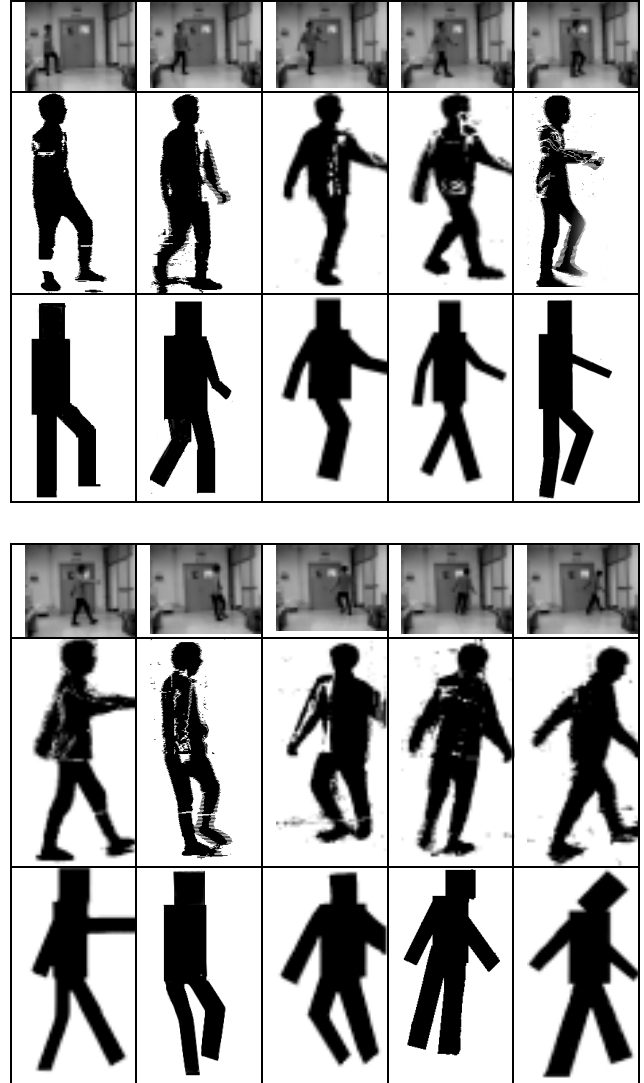


Figure 5. A real video sequence ; its silhouette and model matching result

## 6. Conclusion

In this paper, we present a framework to extract human moving silhouette and then match it to parametric model using genetic algorithm for human posture recognition. Also an  $\alpha-\beta-\gamma$  filter estimating the motion of body parts is effectively applied to reduce search space. Tentative experimental results show that this method is robust enough to extract human model from video.

We plan to extend our work in two different directions:

First, we will develop an automatic tool for human motion video indexing based on techniques proposed in this paper. In this application, we must deal with global motion to reconstruct a panorama background. Second, not only intensity but also multi-cues such as color, motion, anthropometric constraints, should be applied in silhouette extraction. And how to introduce knowledge to system must be considered seriously too.

### Acknowledgements

The first author would like to thank Tao Feng and Hongzan Sun for their help and discussion.

### References

- [1] John Daugman, Face and Gesture Recognition, *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vol. 19, No.7, Jul. 1997.
- [2] C. Cedras, and M. Shah, Motion-based recognition: A survey, *Image and Vision Computing*, 13(2):129-155, Mar. 1995.
- [3] C. R. Richard, A. Azarbajejani, T. Darrell, A. P. Pentland, Pfunder: Real-Time Tracking of Human Body, *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vol. 19, No.7, Jul. 1997.
- [4] J. K. Aggarwal, Q. Cai. Human Motion Analysis: A Review, *Computer Vision and Image Understanding*, 73(3), pp428-440, Mar.1999
- [5] M. K. Leung and Y.H. Yang, First sight: A human body outline labeling system. *IEEE Trans. on Pattern Recognition and Machine Intelligence*, 17(4), pp.359-377, 1995
- [6] S.A. Niyogi and E. H. Adelson, Analyzing and recognizing walking figure in XYT. *Proc. Computer Vision and Pattern Recognition*, pp. 469-474, Seattle, WA, 1994
- [7] Hideo Saito, Masayuki Mori, Application of genetic algorithms to stereo matching of images, *Pattern Recognition Letters* 16(1995) 815-821
- [8] Goldberg. D.E, *Genetic algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, 1989 .
- [9] Pal, S.K.. D. Bhandari and M.K. Kundu, Genetic algorithms for optimal image enhancement *Pattern Recognition Letter*. 15(3) 261-171, 1994
- [10] D. Lu, *Stochastic process*. Tsinghua University Press, 1986.
- [11] K. Rohr, Toward Model-Based Recognition of Human Movements in Image Sequence, *CVGIP: Image Understanding*, Vol. 59, No. 1, pp. 94-115, Jan. 1994.
- [12] D. M. Gavrilu, The Visual Analysis of Human Movement: A Survey. *Computer Vision and Image Understanding*, v 73, n 1, Jan. pp.82-98, 1999.
- [13] J. Yamato, J. Ohya, K. Ishii, Recognition Human Action in Time-Sequential Images Using Hidden Markov Model, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp.379-385, 1992.
- [14] S.X.Ju,M.J.Black and Y.Yacoob. Cardboard people: A parameterized model of articulated image motion. *Proc. International Conf. on Face and Gesture*, Vermont, pp.561-567,1996.
- [15] Vladimir I. Pavlovic, Rajeev Sharma, and T. S. Huang, Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE Trans. on Pattern Recognition and Machine Intelligence*, Vol. 19, No.7, Jul. 1997.