〔非公開〕

ＴＲ－Ｃ－０１４９

遺伝的アルゴリズムを用いたマルチカメラ画像からの人物の姿勢推定法の検討

大谷　淳　　岸野　文郎
Jun OHYA　　Fumio KISHINO

１９９６　３．１５

ＡＴＲ通信システム研究所

遺伝的アルゴリズムを用いたマルチカメラ画像からの人物の姿勢推定法の検討

(A Study of Human Posture Estimation from Multiple Camera Images Using A Genetic Algorithm)

大谷 淳　　岸野 文郎

Jun OHYA and Fumio KISHINO

ATR Communication Systems Research Laboratories
2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-02, Japan

Abstract

A new method for estimating human postures from multiple images using a Genetic Algorithm is proposed. In our algorithm, the posture parameters to be estimated are assigned to the genes of an individual in the population. For each individual, its fitness to the environment evaluates to what extent the human multiple images synthesized by deforming a 3D human model according to the values of the genes are registered to the real human multiple images. Natural selection based on the fitness chooses parents who generate children in the next generation, and for the new individuals, crossover and mutation are performed at random in the mating pool. After a certain number of repetitions for these processes, the estimated parameter values are obtained from the individual with the best fitness. Experiments using synthesized human multiple images show promising results for estimating 17 joint angle values for each degree of freedom of the joints and also the three translational and three rotational degrees of freedom of a human.

1 Introduction

Recently, non-rigid and articulated objects including humans have been considered as important and challenging research targets in the computer vision community. Actually, toward the realization of a variety of monitoring systems and visual communication systems [1], it has become important to estimate human motion and posture by automatic and passive ways.

Existing methods related to human posture estimation involve active methods and image analysis (passive) methods. In some active methods, small bulbs or color markers are attached to the human body and are tracked visually in the images acquired by a TV camera. Other methods use Polhemus magnetic sensors attached to the human body to detect the six pose parameters of the sensor positions [1]. These active methods are appropriate for real-time measurement of human motion, but their applications are strongly limited, because wearing these tools is unrealistic in human monitoring systems and encumbers the activities of the users in visual communication systems.

In image analysis methods, human posture description from a still scene has been studied [2],[3], [4],[5]. In this case, human images are segmented into parts, and each part is fitted by deformable models [2],[3]; unfortunately, this strategy requires range images, which are difficult to be acquired in many actual applications. Similarly, superquadrics have been fitted to 2D human doll images [4]; on a human, some reference points need to be given manually. Human body part models have been matched to a monocular human silhouette image[5]; the contour based matching used in the method is sensitive to image noise, and accurate parameter estimation for arbitrary postures is a difficult task for monocular image analysis. Additionally, human motion in a monocular image sequence has been analyzed in a model-based manner using constraints and knowledge on human movements [6],[7], [8],[9],[10], [11],[12], [13],[14],[15],[16]. Although these methods can deal with 3D human movements, since they require constraints and knowledge, they are not easily applied to arbitrary human postures. Moreover, in some of these methods, an image sequence over time is necessary for estimating a human posture at a time instant.

In this paper, a new passive method for estimating a human posture at a time instant from mutiple images is proposed. The proposed method does not need any constraints and knowledge on human movements nor a sequence of images; instead, it uses a 3D human model of a person whose postures are to be estimated, where the joints of the 3D human model can be rotated in the same manner as real human joints. Of course, humans can take a variety of postures, and the number of combinations of the joint angle values is infinite. Therefore, finding the set of joint angles of a person is a combinatorial optimization problem. In this paper, to solve this type of combinatorial optimization problem, a Genetic Algorithm [17],[18] is used.

The Genetic Algorithm is a search algorithm based on Darwinian evolutionary pro-

cesses: the mechanics of natural selection and natural genetics. From other search techniques such as simulated annealing, the Genetic Algorithm is different in the sense that the Genetic Algorithm searches from a population of points, not a single point. That is, the Genetic Algorithm tries to find many peaks in a large (high-dimensional) search space in parallel; thus, the probability of finding a false peak can be reduced drastically compared with the other techniques.

In the proposed method, a person is observed by multiple TV cameras so that human images from different viewing positions can be acquired. The parameters to be estimated are converted to bit strings and are assigned to the genes of an individual in the population, which has a certain number of individuals. The initial values of the genes of all individuals are given at random. The values of the genes of one individual are applied to the 3D human model, and the model is deformed according to the values. The deformed human model is observed by virtual multiple TV cameras whose geometries and camera parameters are the same as those of the real multiple cameras, and human images are synthesized. Then, the fitness which evaluates to what extent the real and synthesized human images are registered is calculated. After the fitness is obtained for all of the individuals, parents are selected at a probability in proportion to the fitness. In the mating pool, parents are mated to bear children (individuals in the next generation), where genetic operations (crossover and mutation) are performed. After this process is repeated for sufficient generations, the parameter estimations are obtained from the genes of the individual with the best fitness.

In Section 2, posture parameters to be estimated in the upper half of the body are illustrated. In Section 3, details of the proposed method are explained. Section 4 describes the experimental conditions. In Section 5, the experimental results are shown, and discussions are presented. Finally, Section 6 concludes the paper.

2 Parameters to be estimated

A human is an example of an articulated object, and can change its posture, position and orientation over time. The posture of an articulated object is specified in terms of joint angle values for each degree of freedom of each rotational joint. The position and orientation of an articulated object are specified in general via three translational and three rotational degrees of freedom with respect to the reference point of the articulated object. In this paper, human posture estimation includes (1) estimating joint angle values for each degree of freedom of the joints, and (2) estimating the six pose parameters (three translational and three rotational degrees of freedom) of a human.

This paper deals with the upper half of the human body. The rotational joints with their degrees of freedom in the upper half of the body and the six pose parameters are illustrated in Fig. 1. In this figure, C_1 is the reference point located at the center of the torso, and C_2, C_3 and C_4 are the joints of the shoulder, elbow and wrist of the left arm,

respectively; similarly $C_5 \sim C_7$ are defined for the right arm; C_8 is the center of the head. In this paper, the hands are considered to be rigid.

The notations for the joint angles are defined as follows. In Fig. 1, a 3D coordinate system is applied to the head, and the three rotation parameters about the three axes are defined: θ_1 (the head is moved in the vertical direction; e.g. nodding action), θ_2 (the head is moved in the left or right direction; shaking the head), and θ_3 (the head is tilted to the left or right side). The shoulder has the rotation θ_4 about the bone connecting C_2 and C_3, and the two rotations θ_5 and θ_6 at C_2. The elbow has the rotation θ_7 at C_3. The wrist has the rotation θ_8 about the bone connecting C_3 and C_4, and the rotations θ_9 and θ_{10} at C_4. The definitions of $\theta_1 \sim \theta_{10}$ are for the right arm, and similarly, $\theta_{11} \sim \theta_{17}$ are defined for the left arm. Therefore, 17 rotation parameters are to be estimated in this paper.

As described above, the six pose parameters of the human body need to be detected. Figure 1 shows the situation in which the reference point C_1 is located at the origin of the world coordinate $(X - Y - Z)$ system. The pose parameters include three translations along the X, Y and Z axes and three rotations α, β and γ about each axis.

3 Parameter estimation from multiple images using Genetic Algorithm

Humans can take a variety of postures. As described in Section 2, there are 23 parameters (17 parameters for the joints of the upper half of the body and six pose parameters) to be estimated. Although each joint has a movable range and is not completely independent of the movements of other joints, the number of combinations of the parameters is virtually infinite. That is, the problem to be solved is a combinatorial optimization problem. For such problems, the Genetic Algorithm [17],[18] is useful. This paper proposes a method exploiting the Genetic Algorithm for estimating the parameters.

The concept of the proposed method using a Genetic Algorithm is illustrated in Fig. 2. Since humans inherently have 3D structures and make 3D movements, a human is observed by multiple TV cameras $R_1, \cdots, R_N (N \geq 2)$, as shown in Fig. 2. If the number of cameras N is large, the probability of occlusions can be lowered. As described below, the proposed method can avoid 3D reconstruction, for example, by stereo matching; it is usually a difficult and unstable task. The human multiple images synchronously acquired by the cameras are converted to silhouette images, which are binarized images in which candidate regions corresponding to the human and background are discriminated. The silhouette images are obtained by thresholding the intensity differences between the known background images and the images acquired by the cameras $R_1, \cdots, R_N$.

In the proposed method, the 3D model of a human whose posture parameters are to be estimated is created in advance by the method described in Section 4. As shown in

Fig. 2, the genes $X_1, \cdots, X_n$ correspond to the parameters to be estimated. In this paper, each individual in the population has one chromosome and is asexual-haploid. The genes are 8-bit Gray coded integers. The initial values of the genes of each individual are given at random. The following processes are carried out for each individual.

The values of the genes are applied to the 3D human model, and each joint is rotated according to the parameters. The deformed human model is observed by the virtual multiple TV cameras $V_1, \cdots, V_N$ in Fig. 2, and the human multiple images are synthesized; the synthesized images are binarized so that areas for the background and regions corresponding to the human are discriminated. The virtual cameras are placed so that the geometrical relationships between any two virtual cameras are the same as those between the two corresponding real cameras, and the virtual camera V_i and the real camera $R_i (i = 1, \cdots, N)$ have the same camera parameters such as the viewing angle, focal length and viewing direction.

As shown in Fig. 2, the fitness of the individual to the environment is evaluated. In this case, the environment, in other words the target, corresponds to human multiple silhouette images from the real cameras. The fitness used in this paper evaluates how much the synthesized human images are registered to the target images. First, as shown in Fig. 3, the similarity $f_i (i = 1, \cdots, N)$ between the image from the virtual camera V_i and the image from the real camera R_i is calculated as follows.

$$f_i = \frac{S(A \cap B)}{S(A \cup B)}, \tag{1}$$

where A and B denote the candidate region for the human in the image from R_i and the human region in the image from V_i, respectively; $S(x)$ represents the area of region x (the number of pixels). In Eq. (1), f_i takes a value between 0 and 1, and if the regions A and B are registered completely, $f_i = 1$.

The fitness F of all of the cameras is defined as the average of f_i as follows:

$$F = (\sum_{i=1}^{N} f_i)/N. \tag{2}$$

The fitness F in Eq. (2) is calculated for all individuals in the population. Individuals with a higher fitness survive and reproduce (can be parents) at a higher rate, and vice versa. That is, natural selection is the process which chooses parents that can bear children in the next generation. The mechanism of natural selection is obtained from a biased roulette wheel [17] where each individual has a roulette slot sized in proportion to its fitness. Suppose that the population has P individuals with fitness $F_k (k = 1, \cdots, P)$. Then, the probability p_k for individual k to be selected is represented by

$$p_k = \frac{F_k}{\sum_{k=1}^{P} F_k}. \tag{3}$$

According to the probability p_k in Eq. (3), P individuals to be parents are selected and are entered into the mating pool, as shown in Fig. 2. In the mating pool, two individuals (parents) are mated and reproduce two new individuals (children). During the mating process, two genetic operations: crossover and mutation are performed, as illustrated in Fig. 4. In crossover, a position q along the bit string of an individual pair is selected at a probability of p_c (crossover rate), and the portions subsequent to q are swapped. In the example of Fig. 4, the portions a of Individual 1 and b of Individual 2 remain, while the portions a' and b' are exchanged. In mutataion, a bit is selected at a probability of p_m (mutation rate), and the value of the bit is changed from 0 to 1 (bit m in Fig. 4) or from 1 to 0 (m' in Fig. 4).

In this way, P new individuals are born, and the same processes are repeated. After a certain number of generations (repetition of the processes), the individual with the best fitness in the population is chosen, and the values of the genes $\hat{X}_1, \cdots, \hat{X}_n$ of that individual become the estimated values of the parameters to be estimated.

As explained above, the binarization employed to discriminate the human region and background is the only image processing in the proposed method. In addition, the fitness calculation in Eq. (2) is based on area information in images. Therefore, the proposed method is robust against image noise factors. Furthermore, it is obvious that the proposed method is useful not only for humans but also for articulated objects.

4 Experimental conditions

4.1 3D human model

Prior to the genetic operations, it is necessary to create a 3D model of the human whose posture is to be estimated. Separate models for each human body part: the head, body, and arms are used to create the human model. As shown in Fig. 5, we use the Cyberware Color 3D digitizer, which rotates around an object, projects laser stripes, and acquires the 3D coordinates and color information of each point on the surfaces of the parts [1]. By using a utility software, the acquired 3D coordinate data set is transformed to a 3D wire frame model, which consists of triangular patches and approximates the curved surface of each part. As shown in Fig. 5, the 3D model of each part is articulated with another so that the joint movements illustrated in Fig. 1 can be generated.

The wire frame model of the upper half of the human body used in the experiments (Section 5) is shown in Fig. 6(a), and the model mapped by color texture is shown in Fig. 6(b). The human wire frame model of Fig. 6 has approximately 5300 nodes, which could cause very slow computations. Therefore, the simple human model with 200 nodes in Fig. 7 is also used (Section 5). As described earlier, this paper does not deal with fingers, which have a much finer structure than other body parts. It is difficult to deal

with fingers together with the other body parts.

4.2 Human multiple images

In this paper, human multiple images synthesized using the models of Fig. 6 and Fig. 7 are used for target images in order to facilitate the comparison between the real and estimated posture parameter values. Three cameras are used for the multiple cameras, and images are acquired under orthographic projection so that position error analysis is facilitated. The cameras are placed on the three axes of the 3D coordinate system. When the origin of the coordinate system is located at the reference point C_1 of the human (Fig. 1), the three cameras observe the human from the front, side and top, respectively. The image size is 256 $\times$ 256 pixels. The image length (256 pixels) corresponds to 173 cm; therefore, 0. 68cm/pixels. For the computers, Silicon Graphics IRIS Workstations (Onyx Reality Engine 2, Crimson Reality Engine) are used.

5 Experimental results and discussion

5.1 Parameters of Genetic Algorithm

The influence of the parameters of the Genetic Algorithm on the accuracy of posture estimation is studied, where the parameters are crossover rate p_c, mutation rate p_m, and the number of generations. In Sections 5.1 and 5.2, the 17 parameters $\theta_1 \sim \theta_{17}$ are to be estimated, while the six pose parameters in Fig. 1 are constant ($X = Y = Z = 0$(cm), $\alpha = \beta = \gamma = 0$(deg)). Two synthesized human images from the three directions are used for the experiments. The images are displayed in Fig. 8(a) and (b), and the values of $\theta_1 \sim \theta_{17}$ for the human model in the two multiple images are listed in Table 1.

First, how the crossover rate p_c and mutation rate p_m affect the accuracy of posture estimation is investigated for the target images in Fig. 8(a) and (b). To p_c and p_m, the values of 0.1, 0.01, 0.001, 0.0001 and 0.00001 are given; therefore, there are 25 combinations. In the experiments, the population size P is 500, and the genetic operations are repeated until 500 generations. The obtained values of the fitness F in Eq. (2) for the two target multiple images are listed in Table 2, where the simple human model in Fig. 7 is used to accelerate the computations. As can be seen from Table 2, for the multiple images (a), $p_c = 0.01$ and $p_m = 0.0001$ give the best fitness $F = 84.0\%$, and for the images (b), $p_c = 0.01$ and $p_m = 0.001$ give the best fitness $F = 83.9\%$. If the average of the fitnesses of (a) and (b) is taken, $p_c = 0.01$ and $p_m = 0.0001$ give the best fitness.

Figure 9(a) and (b) show how much the real human region and the estimated human region are registered in the three images, where the white, green, red, and black areas are the registered areas, areas for the real human only, areas for the estimated human only, and other areas, respectively. As shown in Fig. 9, for (a), the right hand is occluded

behind the head in the image from the front, while for (b) the results look relatively good. The occlusion problem in (a) could happen, because the area of the hand is relatively small compared with the entire area for the human, and the fitness F does not get worse even if the hand is occluded. As a matter of fact, although the fitness is lower (82.7%) than that for $p_c = 0.01$ and $p_m = 0.0001$, in the result for $p_c = 0.01$ and $p_m = 0.00001$ the right hand is not occluded. Further study is needed for dealing with occlusions.

Results obtained using the precise human model in Fig. 6 are shown in Fig. 10, where $p_c = 0.01$ and $p_m = 0.00001$ for (a), and $p_c = 0.01$ and $p_m = 0.0001$ for (b). The fitness is increased to 87.9% for (a) and 90.1% for (b). The estimated values of the 17 posture parameters are listed in Table 3. The accuracies of the estimations are quite good except for θ_2, θ_8, θ_9, θ_{10}, θ_{15} and θ_{16} for (a) and/or (b). A possible cause for θ_9, θ_{10} and θ_{16} is that the angles are for the hands, and the hands are smaller than the other body parts; therefore, these angles do not cause large changes in the fitness. The cause for θ_2, θ_8 and θ_{15} is considered that the angles are about the central axes of the head and arms, and the rotations for θ_2, θ_8 and θ_{15} do not result in large changes in the head and arm images. The position errors at the eight joints $C_1 \sim C_8$ in Fig. 1 for the results of Fig. 10 are listed in Table 4, where the position error is defined as the Euclidean distance between the real and estimated positions in the world coordinate system. In Table 4, since the position of the torso is fixed (the six pose parameters are constant), the errors for C_1, C_2 and C_5 are zero. Most of the error values are smaller than 5 cm, which is less than 3% of the image length (256 pixels); even the maximum error is 9.6 cm, which is 5.5% of the image length.

The relationships between the fitness values and the number of generations in Fig. 10 are shown in Fig. 11, where the maximum, minimum and average fitness values are indicated. With respect to the maximum value, it increases sharply until approximately 200 generations, but after that, the increase is very slow and is almost saturated by 500 generations.

5.2 Improvement in the accuracy of posture estimation

As described in Section 5.1, the fitness is still not 100%. If the genetic operations are repeated further, the fitness could be increased, but from the increasing rate between 200 and 500 generations, the number of further repetitions could be extremely large.

To reach the real peak efficiently, improvement of the estimation accuracy is studied using the steepest ascent method for which the initial values are the values of the genes of the individual with the best fitness at the 500th generation. Then, if $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_{17})$, the problem is to obtain the maximum of the fitness $F(\boldsymbol{\theta})$. That is, θ is updated by

$$\theta_i^{m+1} = \theta_i^m + \rho \frac{\partial F(\boldsymbol{\theta})}{\partial \theta_i} \tag{4}$$

where ρ is a step width. If

$$|F(\boldsymbol{\theta}^{m+1}) - F(\boldsymbol{\theta}^{m})| \rightarrow 0 \tag{5}$$

is satisfied, $\boldsymbol{\theta}$ is the final estimation.

The fitness obtained from Eqs. (4) and (5) for Fig. 10 was improved by 0.7% (from 87.9% to 88.6%) for (a), and there was no improvement for (b). On the other hand, in some cases where the fitness by the genetic operations is lower, larger improvement was obtained; for example, the fitness values of 86% for (a) and 84% for (b) ($p_c = 0.0001$, $p_m = 0.001$, with the precise model) were improved to 90% and 86%, respectively. The estimated values by the genetic operations for (a) and (b) in Fig. 10 are quite close to the real peaks (values) in the search space, but by the steepest ascent method the values do not approach the real peaks. Further study using other techniques is necessary for the improvement of the estimation.

5.3 Estimation of the positional and orientational parameters

In Sections 5.1 and 5.2, the six pose parameters $X, Y, Z, \alpha, \beta, \gamma$ are constant, but in this section the six pose parameters are also estimated. The real values for the 23 parameters are listed in Table 5. The genetic operations are repeated for 500 generations, where the parameters for the operations are $p_c = 0.01$, $p_m = 0.0001$ and the population size=500. The results are shown in Fig. 12, and the estimated parameters are listed in Table 5. Although the fitness is 79.7%, the position parameters are estimated quite accurately (Fig. 12). However, the estimation for the right arm is very inaccurate. Since the dimension of the search space is higher than that in Sections 5.1 and 5.2, the population size may need to be larger for the estimation of the 23 parameters.

6 Conclusions

This paper has presented a method for estimating human postures from multiple images using a Genetic Algorithm. In our algorithm, the posture parameters to be estimated are assigned to the genes of an individual in the population. For each individual, its fitness to the environment evaluates to what extent the human multiple images synthesized according to the values of the genes are registered to the real human multiple images. Natural selection based on the fitness chooses parents who generate children in the next generation, and for the new individuals, crossover and mutation are performed at random. After a certain number of repetitions for these processes, the estimated parameter values are obtained from the individual with the best fitness.

Our method is a passive method, and can avoid 3D reconstruction from multiple images or an image sequence, which is generally a difficult task. The only image processing in our method is extracting human candidate regions in the multiple images by taking

the difference between the human images and known background images; therefore, our method is robust against image noise. Our method can be applied to articulated objects other than humans by replacing the 3D human models with the 3D models of the articulated objects.

The main experimental results using synthesized multiple images for the upper half of the human body are as follows:

1. The optimal values achieving the best fitness are found for the crossover rate and mutation rate. With the optimal values, the fitness values of approximately 90% are obtained from the genetic operations for 500 individuals after 500 generations. The 17 joint angles in the upper half of the body are estimated accurately, and the position errors are less than 6% of the image size.

2. To improve the fitness efficiently, the steepest ascent method is applied to the estimated parameters obtained from the Genetic Algorithm, and the fitness is improved by 2 or 3 %. To approach the value of 100% for the fitness, further study using other techniques may be necessary.

3. The case where the six pose parameters as well as the 17 parameters (for the joints of the upper half of the human body) are to be estimated has been studied, and a promising result was obtained.

In addition to the problems described above, the computation speed should be improved, for example, by parallel processing implementation so that much more data can be collected. The improvement in the computation speed will be useful also for improving the accuracy of the posture estimation by exploiting multiple images from more than three cameras and by increasing the number of individuals in the population. This paper dealt with the upper half of the human body, because it has enough degrees of freedom for posture estimation, but as the next target, the whole human body should be treated. As described earlier, our method is robust against image noise, but the robustness should be confirmed by conducting experiments using real human images as the targets.

参考文献

[1] J.Ohya, Y.Kitamura, H.Takemura, F.Kishino and N.Terashima, "Real-time reproduction of 3D human images in virtual space teleconferencing", Proc. of IEEE Virtual Reality Annual International Symposium (VRAIS'93), pp.408-414, Sep. 1993.

[2] A.Pentland, "Automatic extraction of deformable part models", International Journal of Computer Vision, 4, pp.107-126 (1990).

[3] A.Pentland and S.Sclaroff, "Closed-form solutions for physically based shape modeling and recognition", IEEE Tr. PAMI, Vol.13, No.7, pp.715-729 (1991).

[4] D.Terzopoulos and D.Metaxas, "Dynamic 3D models with local and global deformations: deformable superquadrics", IEEE Tr. PAMI, Vol.13, No.7, pp.703-714 (1991).

[5] Y.Kameda, M.Minoh and K.Ikeda, "Three dimensional pose estimation of an articulated object from its silhouette image", Proc. of Asian Conference on Computer Vision, pp.612-615 (1993).

[6] S.Tsuji, M.Osada and M.Yachida, "Tracking and segmentation of moving objects in dynamic line images", IEEE Tr. PAMI, Vol.PAMI-2, No.6, pp.516-522 (1980.11).

[7] J.O'Rourke and N.Badler, "Model-based image analysis of human motion using constraint propagation", IEEE Tr. PAMI, Vol.PAMI-2, No.6, pp.522-536 (1980.11).

[8] J.A.Webb and J.K.Aggarwal, "Structure from motion of rigid and jointed objects", Artificial Intelligence, 19, pp.107-130 (1982).

[9] D.Hogg, "Model based vision: a program to see a walking person", Image and Vision Computing, 1, pp.5-20 (1983).

[10] K.Akita, "Image sequence analysis of real world human motion", Pattern Recognition, Vol.17, No.1, pp.73-83 (1984).

[11] T.Tsukiyama and Y.Shirai, "Detection of the movements of persons from a sparse sequence of TV images", Pattern Recognition, Vol.18, Nos.3/4, pp.207-213 (1985).

[12] M.Yamamoto and K.Koshikawa, "Human motion analysis based on a robot arm model", Proc. of CVPR, pp.664-665 (1991).

[13] A.Pentland and B.Horowitz, "Recovery of nonrigid motion and structure", IEEE Tr. PAMI, Vol.13, No.7, pp.730-742 (1991.7).

[14] Z.Chen and H.Lee, "Knowledge-guided visual perception of 3-D human gait from a single image sequence", IEEE Tr. SMC, Vol.22, No.2, pp.336-342 (1992).

[15] S.Kurakake and R.Nevatia, "Description and tracking of moving articulated objects", Proc. of 11th ICPR, Vol.I, pp.491-495 (1992).

[16] K.Rohr, "Incremental recognition of pedestrians from image sequences", Proc. of CVPR, pp.8-13 (1993).

[17] D.E.Goldberg, "Genetic Algorithms in search, optimization, and machine learning", Addison-Wesley Publishing Company, Inc., 1989.

[18] J.R.Koza, "Genetic Programming", The MIT Press, 1992.

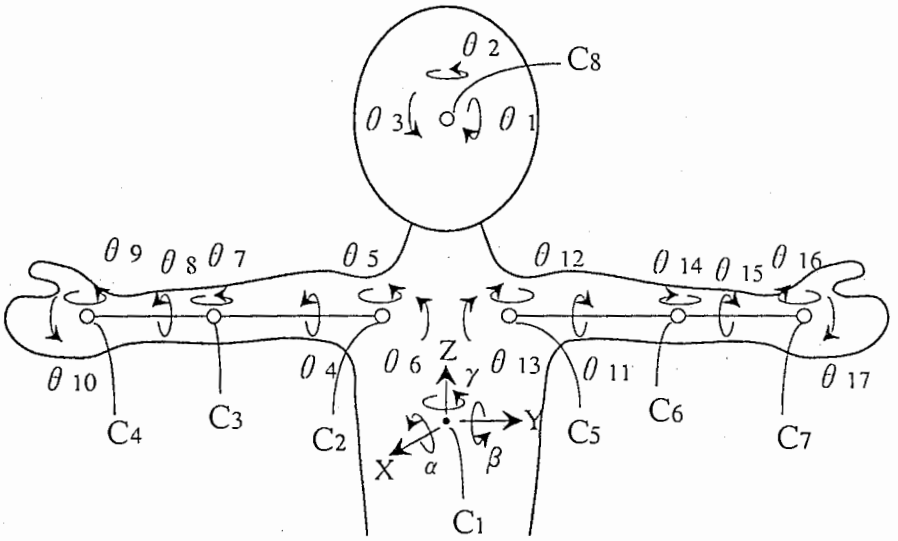


Figure 1: Parameters to be estimated

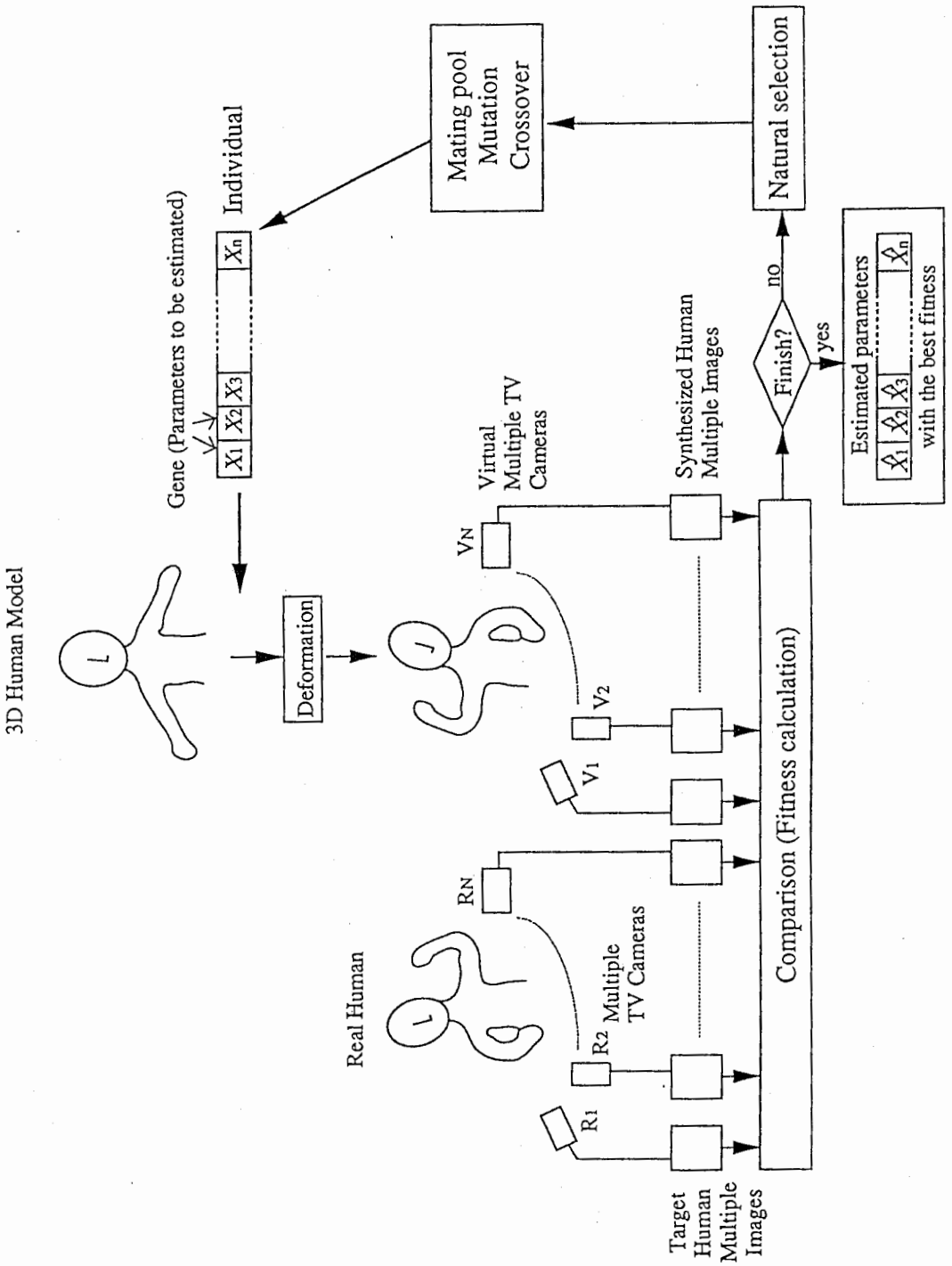


Figure 2: Principle of the proposed method

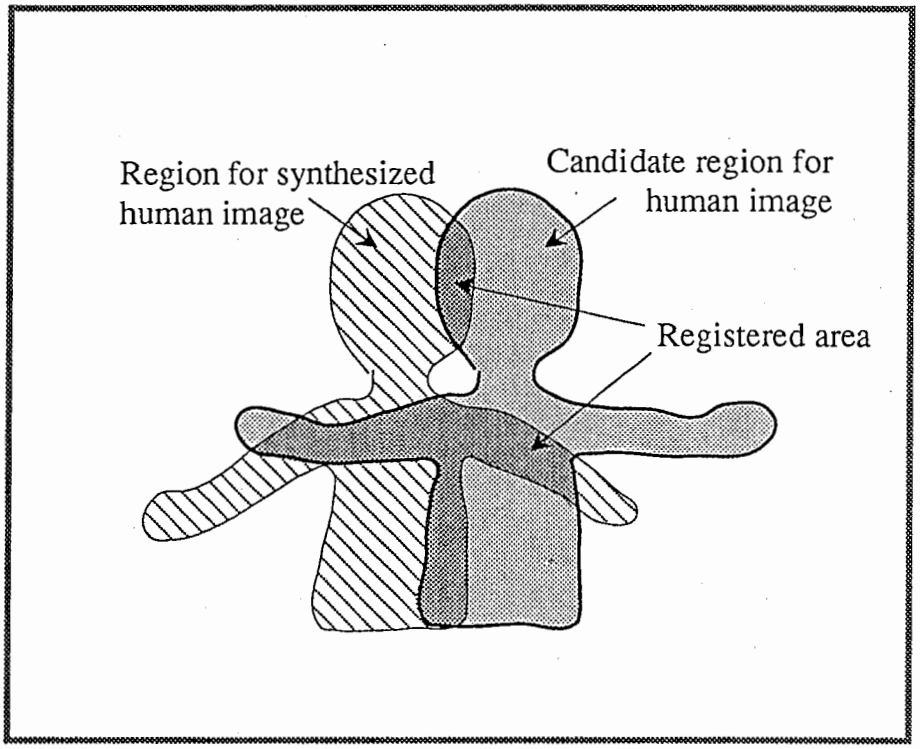
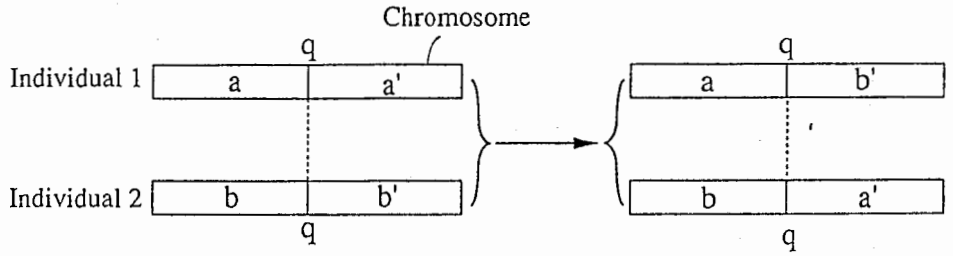


Figure 3: Similarity

Crossover



Mutation

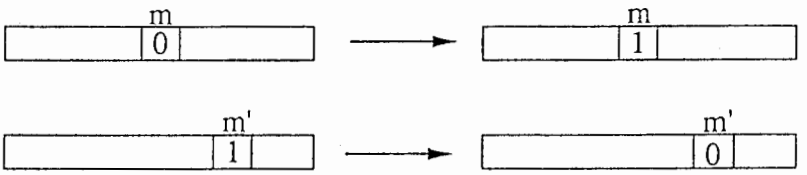


Figure 4: Crossover and mutation

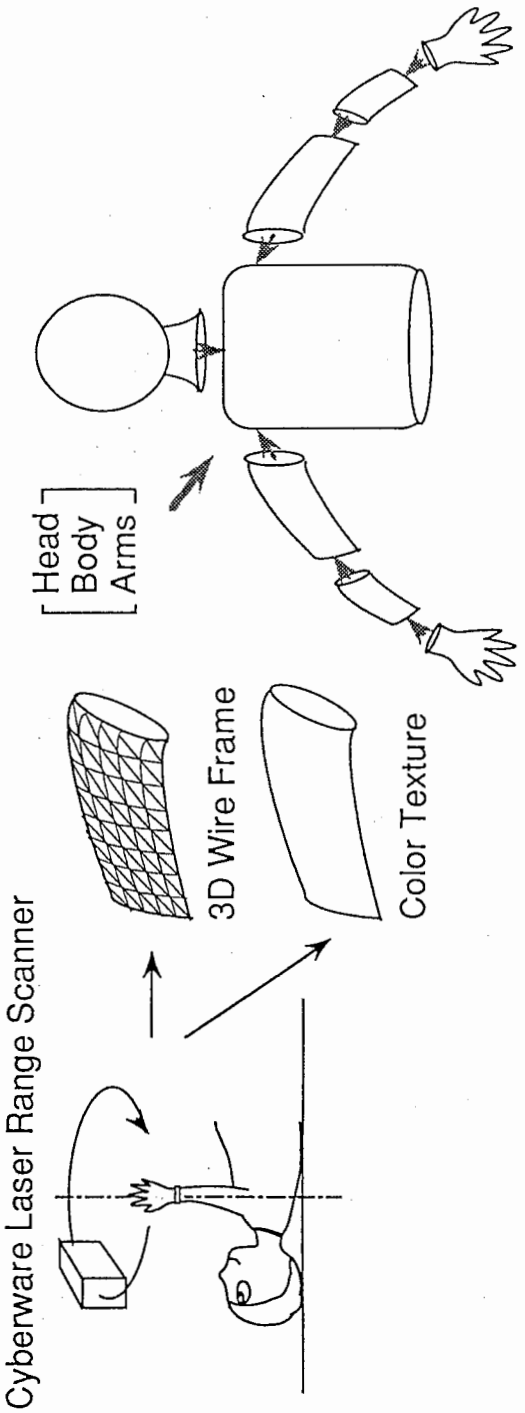
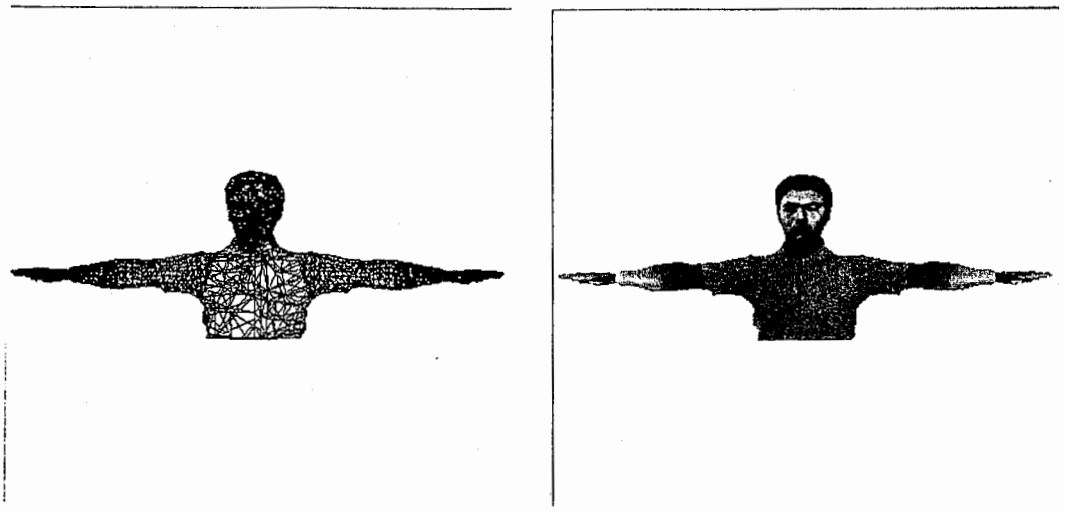


Figure 5: 3D human modeling



(a) (b)

Figure 6: 3D precise human model: (a) Wireframe model, (b) Mapped by color texutre

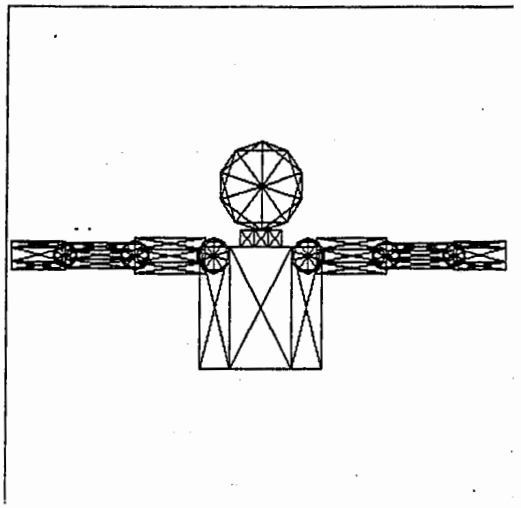


Figure 7: 3D simple human model

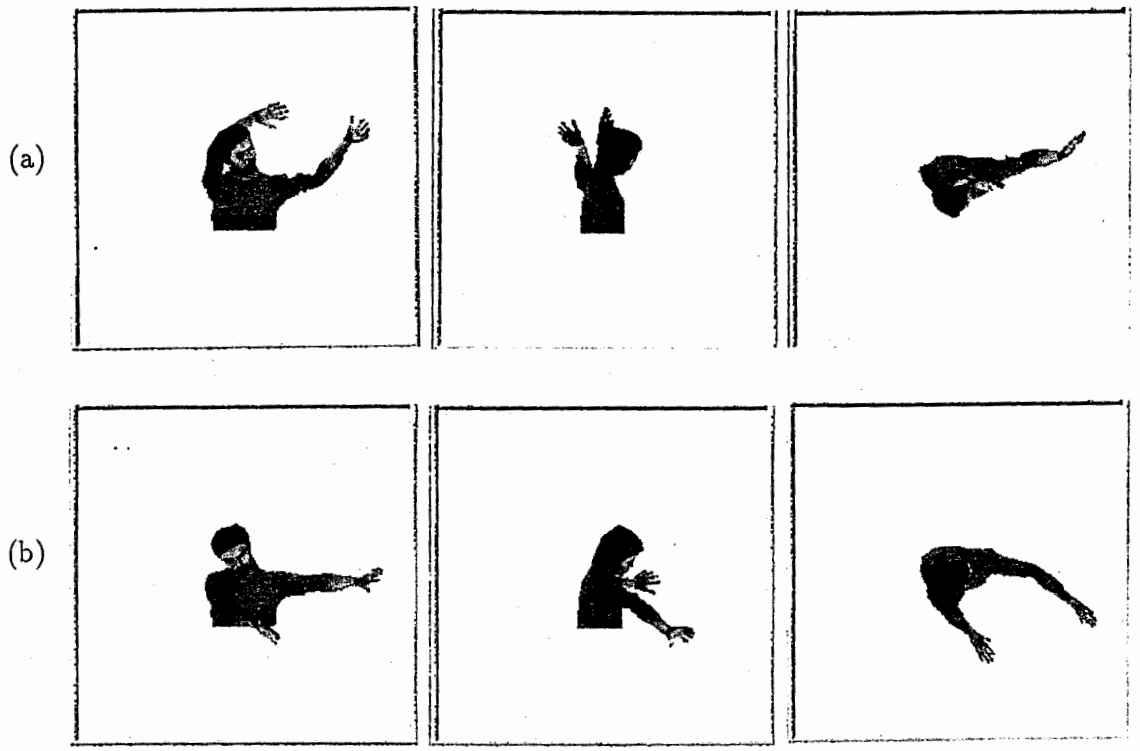


Figure 8: Target multiple images

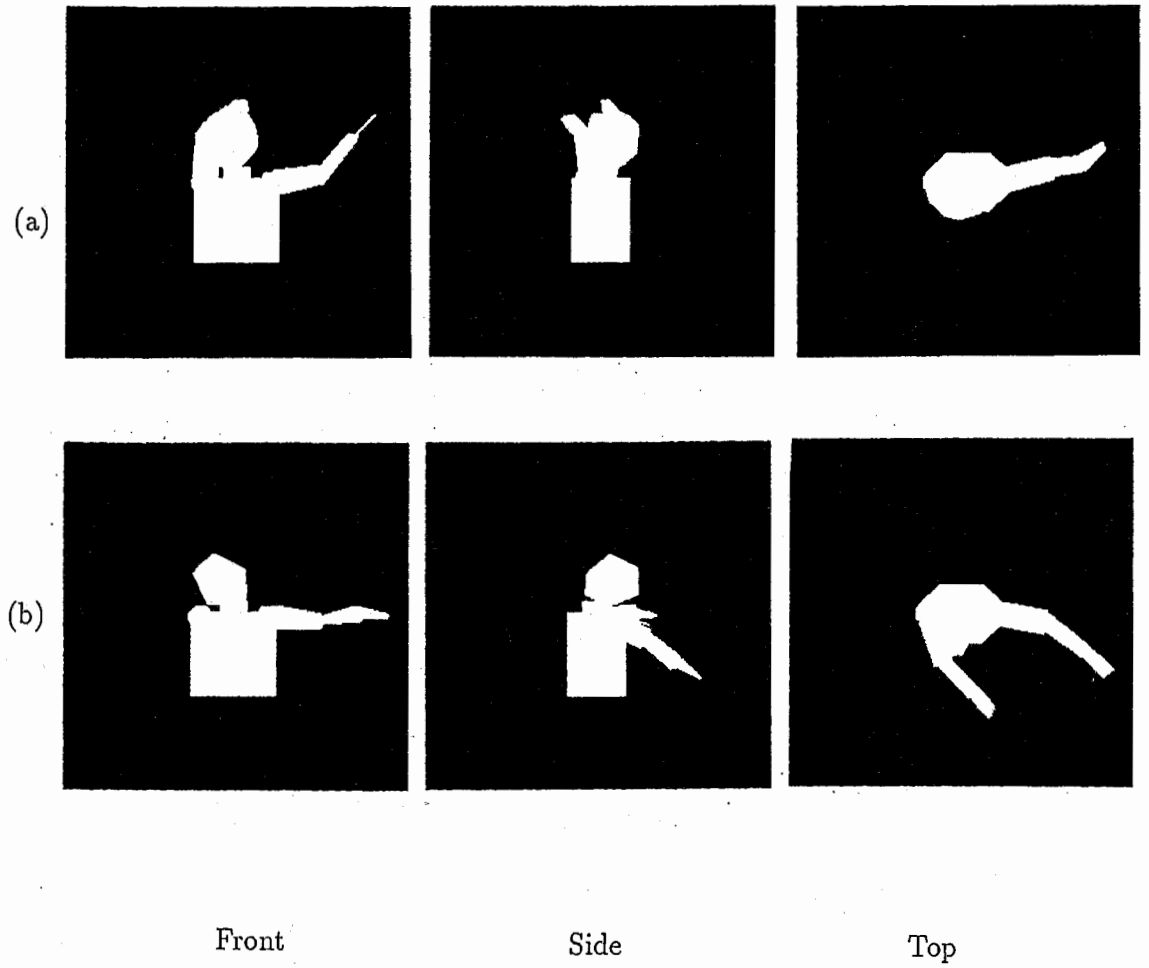


Figure 9: Estimated results using the simple model for (a) and (b); $p_c = 0.01$, $p_m = 0.0001$, 500 individuals, 500 generations. (The white, green, red and black areas are the registered areas, areas for the real human only, areas for the estimated human only and other areas, respectively.)

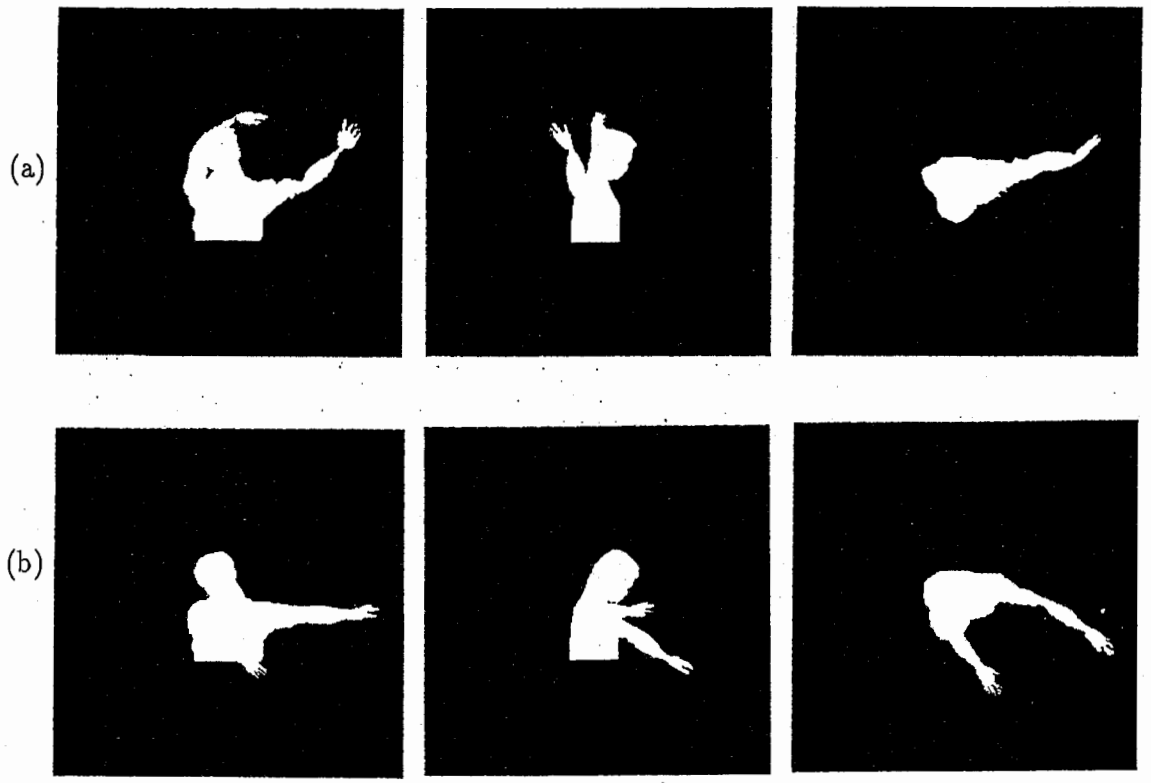
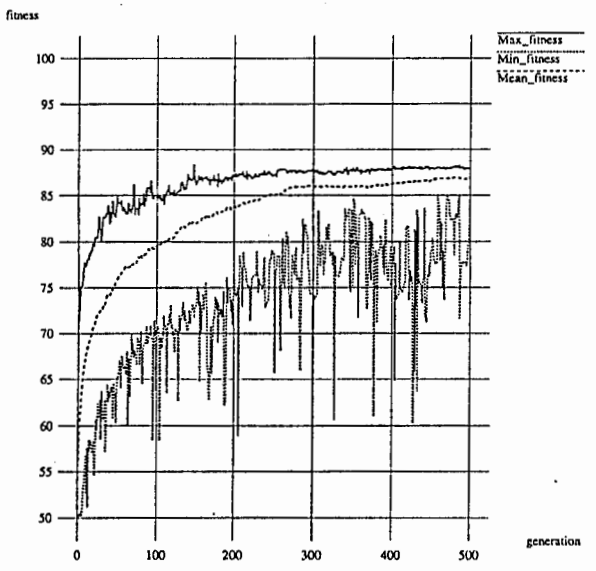
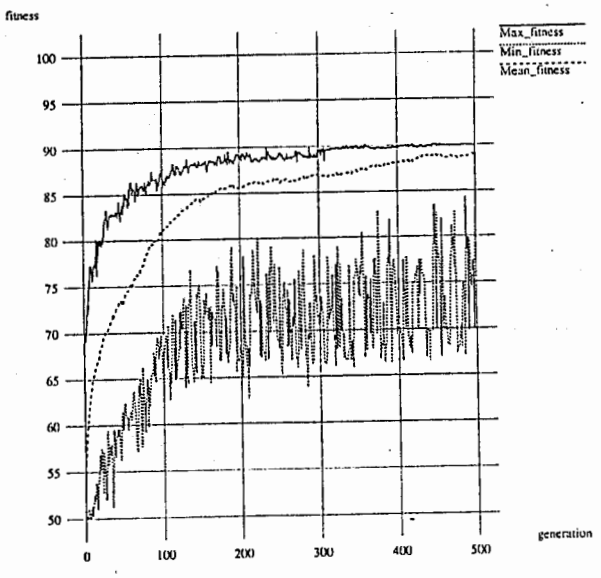


Figure 10: Estimated results using the precise model; $p_c = 0.01$, 500 generations, 500 individuals, for (a) $p_m = 0.00001$, and for (b) $p_m = 0.0001$.



(a)



(b)

Figure 11: Relationship between the fitness and generation

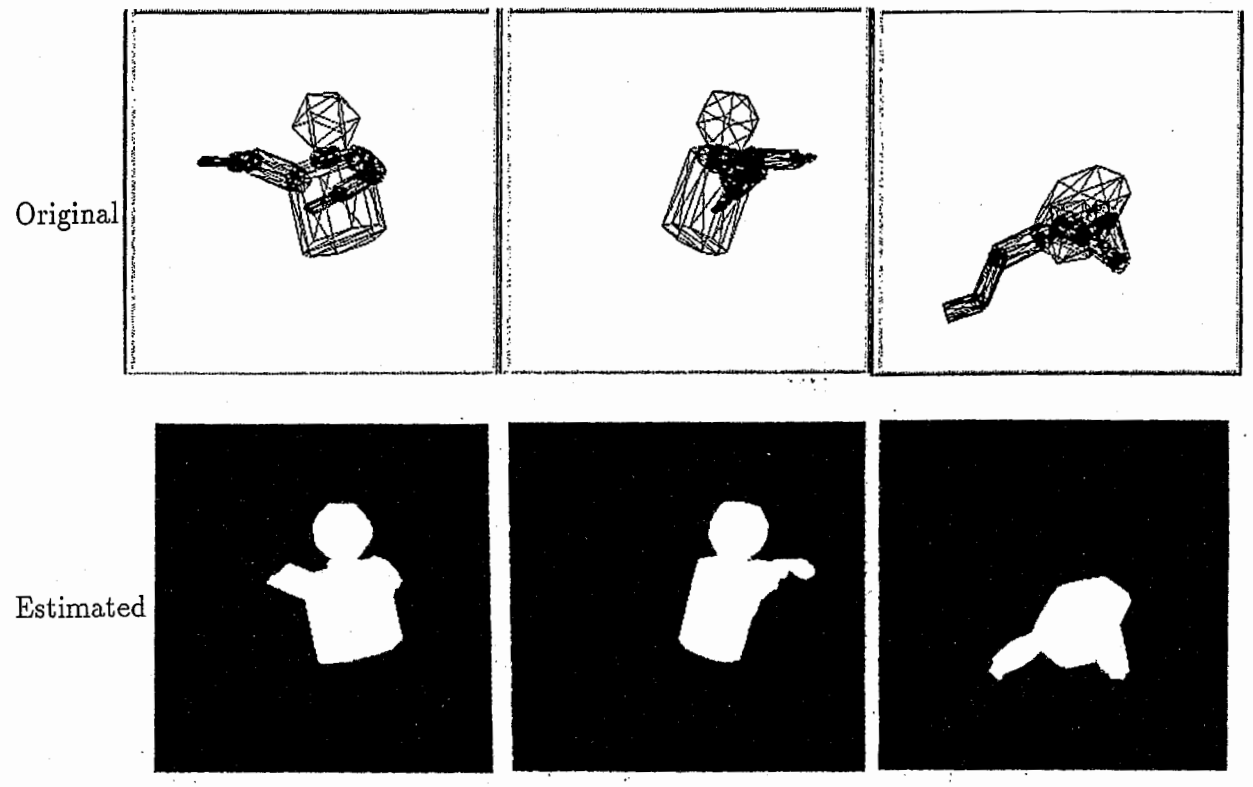


Figure 12: Estimation of the 23 parameters: (the simple model, $p_c = 0.01$, $p_m = 0.0001$, 500 individuals, 500 generations.)

Table 1: Real values of $\theta_1 \sim \theta_{17}$ (deg)

Target	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}
(a)	32	31	37	3	112	-75	49	47	36	6	-79	15	14	-38	6	17	26
(b)	35	-3	26	28	110	21	24	23	-12	-5	-3	-10	2	-28	55	0	8

Table 2: Fitness values (%) for (a) (upper in each box) and (b) (lower): (the simple model, 500 individuals, 500 generations).

Mutation rate p_m	Crossover rate p_c				
	0.1	0.01	0.001	0.0001	0.00001
0.1	69.4 65.2	75.9 66.0	75.3 70.7	71.9 68.6	72.6 68.1
0.01	79.0 76.3	78.7 77.8	78.9 77.0	78.4 76.6	79.0 77.4
0.001	83.2 83.5	82.7 83.9	83.0 83.6	81.1 79.8	80.1 81.6
0.0001	83.4 78.2	84.0 83.5	80.7 81.7	78.6 68.5	78.6 71.7
0.00001	81.6 77.8	82.7 82.5	80.1 73.7	79.1 74.9	72.1 68.0

Table 3: Estimated values of $\theta_1 \sim \theta_{17}$ (deg): (the precise model, $p_c = 0.01$, 500 generations, 500 individuals, for (a) $p_m = 0.00001$, and for (b) $p_m = 0.0001$.)

Target	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}
(a)	26	-23	5	-17	107	-72	45	-51	-29	56	-71	15	20	-18	-29	-36	25
(b)	37	-22	15	-10	118	26	5	-30	5	4	35	-19	4	-5	-73	-29	-12

Table 4: Position errors: with the precise model, $p_c = 0.01$, 500 generations, 500 individuals, for (a) $p_m = 0.00001$, and for (b) $p_m = 0.0001$.

Joint	Position error (cm)	
	(a)	(b)
C_1	0.0	0.0
C_2	0.0	0.0
C_3	1.5	3.3
C_4	4.4	2.1
C_5	0.0	0.0
C_6	2.6	4.4
C_7	4.0	2.4
C_8	9.6	4.1

Table 5: Real and estimated values for the 23 parameters: X, Y, Z (cm); $\alpha, \beta, \gamma, \theta_1 \sim \theta_{17}$ (deg)

	X	Y	Z	θ	β	γ	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8
Real values	25	25	25	20	20	20	-15	-50	-15	7	-6	-39	47	37
Estimated values	22	27	24	13	8	13	4	34	-12	-41	34	-11	108	-31

	θ_9	θ_{10}	θ_{11}	θ_{12}	θ_{13}	θ_{14}	θ_{15}	θ_{16}	θ_{17}
Real values	-20	47	24	-98	8	-125	-31	2	9
Estimated values	-117	11	11	-90	51	-122	34	-12	19