The Fourteenth International Symposium on Artificial Life and Robotics 2009 (AROB 14th '09),
B-Con Plaza, Beppu, Oita, Japan, February 5 - 7, 2009

A Method for Expressing Human Posture as 3DCG Using Thermal Image Processing and 3D Model Fitting

T. Asada[1] and Y. Yoshitomi[2]

1: Dept. of Environmental Information Graduate School of Human Environment Science Kyoto Prefectural University, Shimogamo Sakyo-ku Kyoto 606-8522 JAPAN, E-mail: t_asada@aoi.kpu.ac.jp
2: Div. of Environmental Sciences Graduate School of Life and Environmental Sciences Kyoto Prefectural University Shimogamo Sakyo-ku Kyoto 606-8522 JAPAN, E-mail: yoshitomi@kpu.ac.jp

Abstract: An imitation of human motion has been used as a promising technique for a development of robot. Some techniques such as motion capture systems and data-gloves are used for analyzing a human motion. However, since those methods need (a) environmental restrictions such as preparation of two or more cameras and strict control on brightness condition and (b) physical restrictions such as wearing of markers and/or data-gloves, they are far from the method for recognizing a human motion on the natural condition. In this paper, we propose a method that makes a 3-dimensional CG (3DCG) by transforming a feature vector of human posture on a thermal image into a 3DCG model. The 3DCG models as training data are made with manual model fitting. Then, the human models synthesized by our method are geometrically evaluated in CG space. The average error in position is about 10 cm. Such a relatively small error might be acceptable in several cases of both the 3DCG animation generation and the imitation of human motion by a robot. Our method needs neither the physical nor environmental restrictions. The rotation-angles at each joint obtained by our method can be used for imitating a human posture by a robot.

Keywords: Human posture, Thermal image, Computer graphics, Model fitting, Feature vector**.**

I. INTRODUCTION

In recent years, various computer graphics (CG) animations have been appearing on television or in movie. Such a CG animation is usually made through tough analyses on a human motion. Some techniques such as motion capture systems and data-gloves are used for analyzing a human motion. There have been many researches for acquiring the information on a human posture from an image [1-4]. However, since those methods need (a) environmental restrictions such as preparation of two or more cameras and strict control on brightness condition and (b) physical restrictions such as wearing of markers and/or data-gloves, they are far from the method for recognizing human motion on the natural condition.

In order to overcome the issues coming from these restrictions, the method with thermal images was studied [5]. However, the difficulty in making a 3-dimensional (3D) information on posture from a 2-dimensional (2D) one has been still kept in the reported method [5].

In this paper, a method for expressing a human posture as a CG with human thermal image taken under no special restrictions on a person is proposed. Based on a geometrical relationship among the feature vector of human posture on the test image and those of the training images, the 3DCG models for test images are produced from some 3DCG models, which have been produced by manual model fitting beforehand through referring each human posture on a training image. Then, the human models synthesized by our method are geometrically evaluated in CG space. Our method needs neither the physical nor environmental restrictions. Our method might be useful at least in making a 3DCG for several limited human postures. Moreover, the rotation-angles at each joint obtained by our method can be used for imitating a human posture by a robot.

II. BASIC ALGORITHM

A basic algorithm for producing a CG for expressing a human posture is as follows.

<Step1> Several thermal images expressing one human posture per image are collected as training images. Through referring each training image expressing one human posture, the 3DCG model expressing the corresponding human posture is made by manual model fitting and stored in a computer. Then, the feature vectors for human postures of training images are calculated and stored in the computer.
<Step 2> The test image is inputted into the computer and then the corresponding feature vector for human posture on the image is calculated.
<Step 3> The human model corresponding to the test image is made from those of training images, according to the geometrical relationship in the feature vector

space among the feature vector of the test image and those of the training images.

For each motion in gesture on thermal image, a 3-dimensional CG (3DCG) animation can be made by changing continuously the rotation-angles at joints in a 3DCG model corresponding to the test image at the start time into those at the end time [6].

III. THERMAL IMAGE GENERATION

In this study, thermal images having a human posture per an image are produced by a Thermal Video System (TVS-700) (Nippon Avionics) with infrared rays and these images are recorded in digital videotape as a Mpeg1 file. Then, several thermal images with BMP format having a spatial resolution of 380×460 pixels and a gray level of 256 are obtained from the Mpeg1 file.

IV. HUMAN MODEL

The human model has 10 joints. All rotation-angles at 10 joints are stored in a file. Fig.1 shows the human model, the standard posture and its structure.

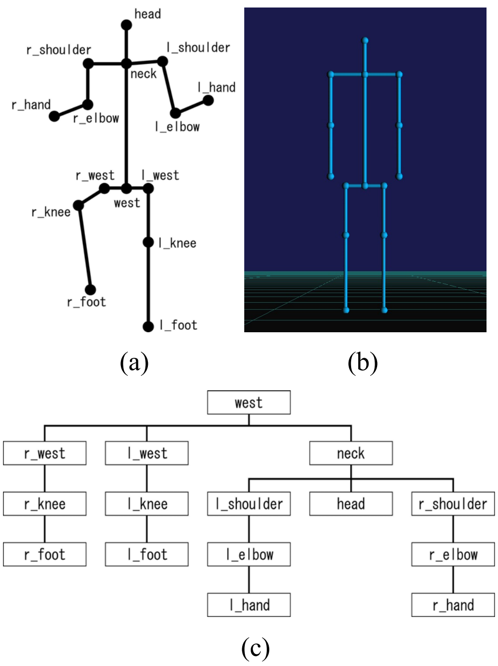

Fig.1. Human model (a), standard posture (b) and structure of human model (c)

V. MODEL FITTING

Fig.2 demonstrates the process of manual model fitting for making a human model corresponding to a human posture on a thermal image. Through the manual model fitting, all rotation-angles at 10 joints are stored in a file

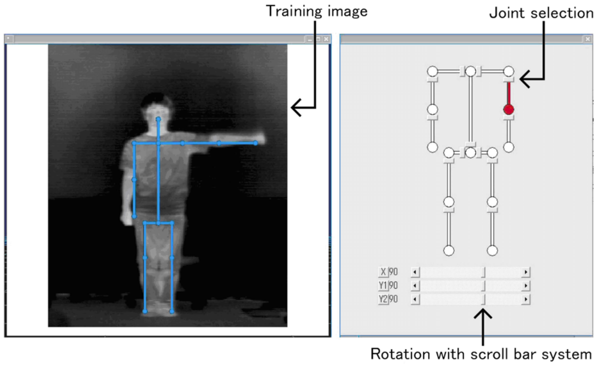

Fig.2. Manual model fitting

VI. FEATURE VECTOR GENERATION

After processing for standardizing the position and the size, the mosaic image is generated from the training image after subtracting the gray level of background image from the training image (Fig.3). The gray levels of mosaic image having 14×14 blocks, each of which has the average gray level on 25×25 pixels, are used as the components of feature vector.

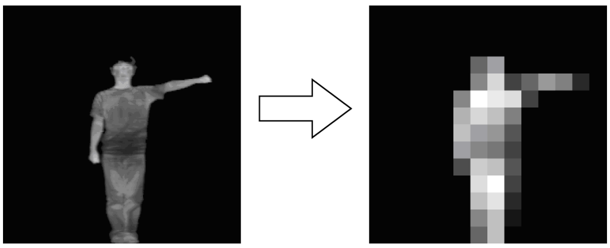

Fig.3. Mosaic image for making feature vector

VII. HUMAN POSTURE ESTIMATION

We assume $\{\boldsymbol{X}_m\}(m=1,2,\cdots,M)$ to be a set of posture pattern made from training images. $\boldsymbol{X}_m$ indicates a N-dimensional feature vector having gray levels of mosaic image as components.

Assuming the feature vector of test image, the feature vector of the class c to be $\boldsymbol{f}=(f_1,f_2,\cdots,f_N), \boldsymbol{f}_c=(f_{c,1},f_{c,2},\cdots,f_{c,N})$ respectively. In the conventional way, the test image is recognized as class j when the Euclid distance d_c, defined by the equation (1) as a discriminant function, gives the minimum value for class j among all classes.

$$d_c = \left\{ \sum_{i=1}^{N} \left(f_i - f_{c,i} \right)^2 \right\}^{1/2} \qquad (1)$$

However, this conventional method for pattern recognition needs various kinds of training images for recognizing many test-images with good accuracy. Therefore it is not practical to produce the CG model corresponding to the test image, with this conventional pattern recognition.

In this study, for transforming the test image into the corresponding 3DCG model, a stepwise approximation to the feature vector of test image is performed in the following way. Hereinafter, the shortest distance among those of all pairs of feature vectors coming from training images is referred as R .

<Step1> The approximate feature vector is obtained as that having the shortest distance (hereinafter referred as $d_{\min}$) to the feature vector of the test image among those of all training images. When $d_{\min}$ is under the half of R , go to the step 5, getting from the database all rotation-angles at 10 joints on the model corresponding to the class whose training image gives $d_{\min}$ to the test image. Otherwise, go to the step 2.

<Step2> In the feature vector space, every length of the perpendicular drawn from the position of the test image to the straight line passing a pair of the positions of feature vectors of training images is calculated. The approximate feature vector is obtained as the position vector of intersection of the shortest perpendicular, which is expressed as H_0 , from the position of the test image on the straight line passing the positions of a pair of training images. The selected two feature vectors of the training images for making the approximate feature vector are expressed as $f_{a,0}$, $f_{b,0}$. When the geometrical relationship of the approximate feature vector for the test image to two selected feature vectors of training images is inside linear interpolation and the length of H_0 is under the half of R , go to the step 5, getting all rotation-angles at 10 joints on the models corresponding to the compounded feature vector obtained by the inside linear interpolation with its ratio and all rotation-angles at 10 joints on the models corresponding to two feature vector used for the inside linear interpolation (Fig.4). Otherwise, set $n = 0$ and go to the step 3.

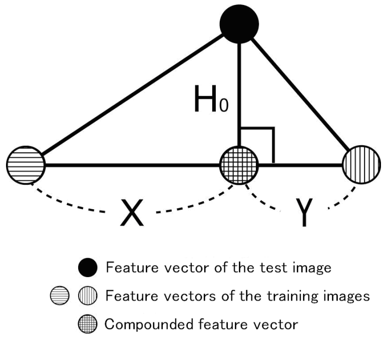

Fig.4. Schematic diagram showing inside linear interpolation for getting the second-step approximate feature vector

<Step3> The feature vector $f_{a,n+1}$ corresponding to the position vector of intersection of the shortest perpendicular from the position of the test image on the straight line passing the positions of a pair of $f_{a,n}$ and $f_{b,n}$ is generated. Then, every length of the perpendicular drawn from the position of the test image to the straight line passing a pair of the positions of $f_{a,n+1}$ and the feature vector of training image is calculated. The approximate feature vector is obtained as the position vector of intersection of the shortest perpendicular, which is expressed as H_{n+1}, on the straight line passing the positions of a pair of $f_{a,n+1}$ and the feature vector of training image. The selected feature vector for making the approximate feature vector is expressed as $f_{b,n+1}$. When the geometrical relationship of the approximate feature vector for the test image to $f_{a,n+1}$ and $f_{b,n+1}$ is inside linear interpolation and the length of H_{n+1} is under the half of R , go to the step 5, getting all rotation-angles at 10 joints on the model corresponding to the compounded feature vector obtained by the inside linear interpolation with its ratio and all rotation-angles at 10 joints on the models corresponding to two feature vector used for the inside linear interpolation. Otherwise, go to the step 4.

<Step4>
$n := n + 1$, then go to the step 3.

<Step5>
The model is produced with the rotation-angles finally obtained in the above step(s) at 10 joints.

Figs. 5 and 6 (1) demonstrate the above approximation of the feature vector of test image, using a simple case. Fig.5 (a) and (c) illustrate the training images while Fig.5 (b) illustrates the test image. Fig.6 (1) demonstrates the relationship among the values of components of feature vectors of the three images illustrated in Fig.5. In this case, the feature vector of the test image (b) is compounded by the inside linear interpolation with the feature vectors of the training images (a) and (c). We extend this idea of approximation to the more general case shown in Fig.6 (2), where (d) denotes the position of the feature vector of test image. Fig.6 corresponds to Fig.4 with which the second step is demonstrated.

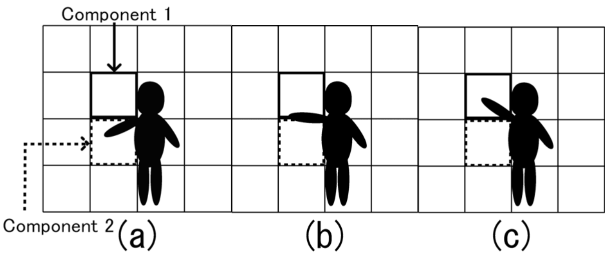

Fig.5. Schematic diagram on images; (a)& (c) training, (b) test

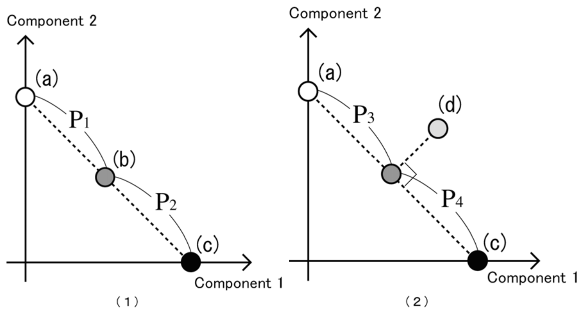

Fig.6. Schematic diagram on compound of feature vector, (1) simple, (2) more general

VIII. ESTIMATION OF POSITION ERROR IN CG SPACE

In the experiments, a personal computer, Dell Dimension 8300 (CPU: Pentium IV 3.2 GHz, main memory: 2.0 GB, OS: Windows XP), is used. For programming, Microsoft Visual C++ 6.0 is used. For the evaluation of our method, several marks are attached to the subject and the model in CG space (Fig.7). Then, the model is manually fitted to the human posture. For fitting the model to the human posture, we do not use the markers. The markers are used only for measuring the position-errors of markers on the human model (Fig.8). We use the motion capture system named as Radish (Library) for measuring the position of markers attached on the subject. We use 28 training-images shown in Fig.9 and 10 test-images shown in Fig.10. In the case of 8 test images (a-h), each image has very similar human posture on training images (No.1-8) respectively, while 2 test images (i, j) do not have very similar human posture on training images (No.1-24).

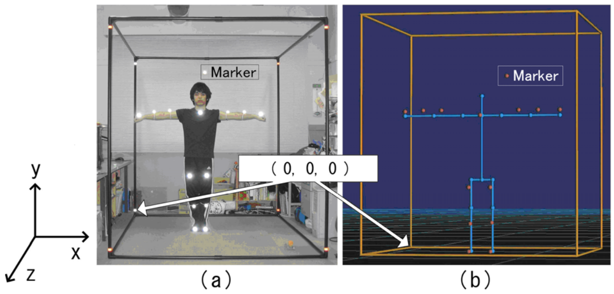

Fig.7. Scene in measuring three-dimensional positions (a), CG space corresponding to real space (b)

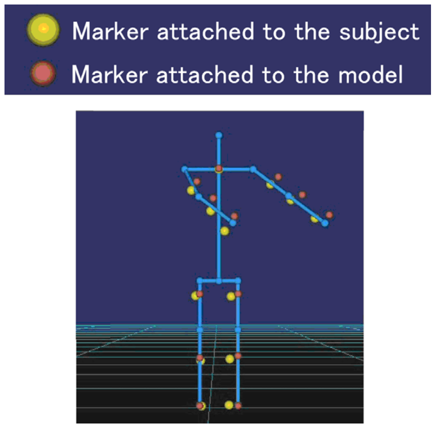

Fig.8. A model with markers

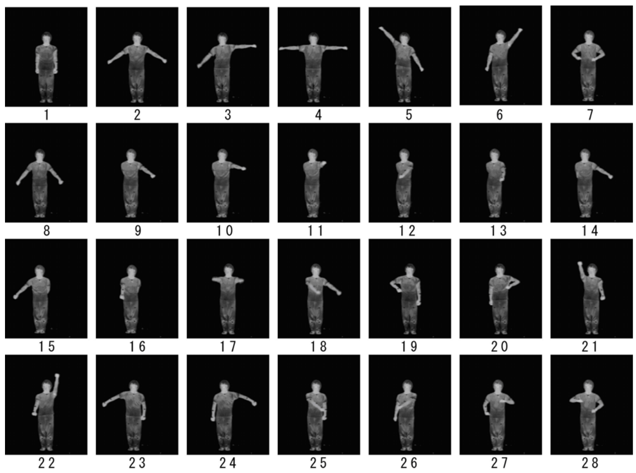

Fig.9. Training images for position measuring

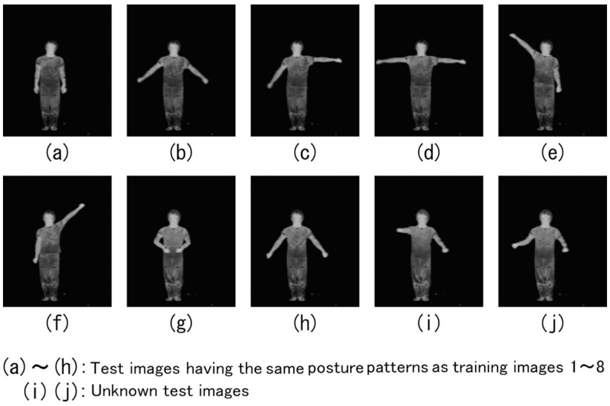

Fig.10. Test images for position measuring

Table 1 shows the position-errors of markers on the human model for the training images. As the total, the average and maximum position-errors of markers are 6.2 and 13.6 cm respectively. Tables 2 and 3 show the position-errors of markers on the human model for the test images. As the total, the average and maximum position-errors of markers are 7.5 and 20.4 cm for the images a-h respectively, while those are 8.9 and 24.3 cm for the images i,j respectively. As the total, both the average and maximum position-errors of markers for the images i,j, which do not have very similar image on the training images (No.1-28), are slightly bigger than those for the images a-h, each of which has very similar image on the training images (No.1-8). Since, the average position-errors of markers for test images are about 10 cm, we think that our method is useful for estimating a human posture roughly. Figs 11 and 12 show the human models for test images i, j respectively. Fig. 13 demonstrates the synthetic process of feature vector in the case of test image j. The process is terminated at the step 3. When using the conventional way, the test image j having the shortest distance in the feature vector space to the training image No.7 is recognized as the class of training image No.7. Since the human model shown in Fig. 12 is more similar to the human posture of the test image j than that of Fig.14, our method might have an advantage over the conventional way. It takes about one minute for making a human model with manual model fitting, while it takes about 0.6 second for estimating a human posture and making a human model for a test image.

Our method needs neither the physical nor environmental restrictions. The rotation-angles at each joint obtained by our method can be used for imitating a human posture by a robot.

Table 1. Position-errors of makers on human model (Training images No.1−8)

No.	1	2	3	4	5	6	7	8	Total
Average	4.4	7.5	5.4	5.5	7.3	5.4	7.4	6.4	6.2
Maximum	6.8	23	10	13	11	16	14	15	13.6

(cm)

Table 2. Position-errors of makers on human model (Test images a−h)

	a	b	c	d	e	f	g	h	Total
Average	5.0	6.6	7.4	7.1	8.4	5.4	11.9	8.2	7.5
Maximum	12.4	24.0	20.2	13.5	27.1	16.0	26.7	23.5	20.4

(cm)

Table 3. Position-errors of makers on human model (Unknown test images i,j)

Image	i	j	Total
Average	8.1	9.6	8.9
Maximu m	25.4	23.2	24.3

(cm)

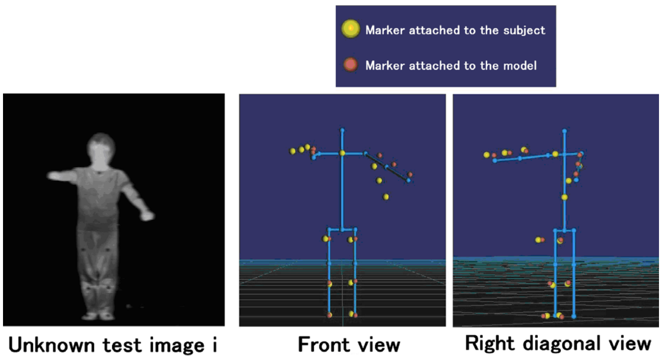

Fig.11. Unknown test image i and corresponding CG

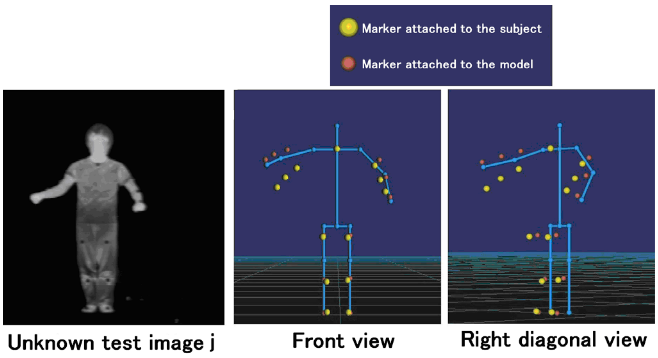

Fig.12. Unknown test image j and corresponding CG

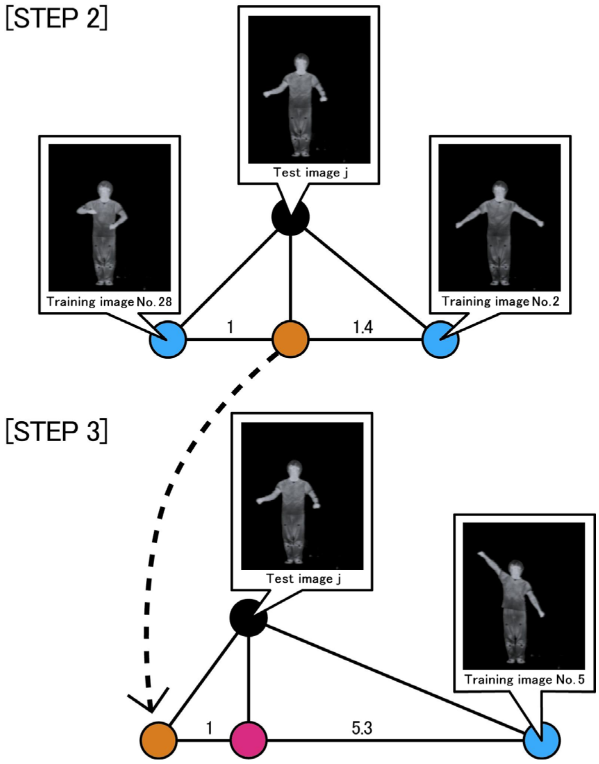

Fig.13. Synthetic process of feature vector for unknown test image j

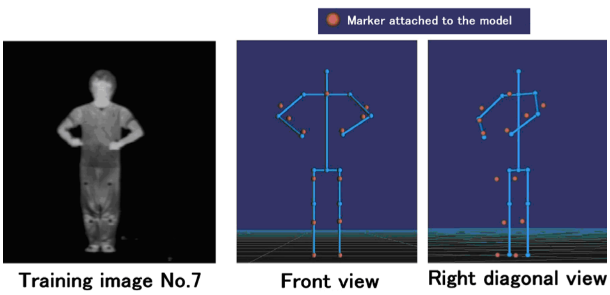

Fig.14. CG for training image No.7

IX. CONCLUSION

We propose a method that makes a 3DCG by transforming a feature vector of posture on a thermal image into a 3DCG model. The 3DCG models as training data are made with manual model fitting. The average error in position is about 10 cm. Such a relatively small error might be acceptable in several cases of both the 3DCG animation generation and the imitation of human motion by a robot. Our method needs neither the physical nor environmental restrictions.

REFERENCES

[1]Pentland A (1990), Automatic extraction of deformable part models, Int. Journal of Computer Vision 4(2):107-126
[2]Pentland A and Sclaroff S (1991), Closed-form solutions for physically based shape modeling and recognition. IEEE Trans. on Pattern Analysis and Machine Intelligence 13(7):715-729
[3]Ohya J and Kishino F (1994), Human posture estimation from multiple images using genetic algorithm. Proc. 12th ICPR:750-753
[4]Wren C, Azarbayejani A, Darrel T and Pentland A (1996), Pfinder, real-time tracking of the human body. Proc. 2nd Int. Conf. on Automatic Face and Gesture Recognition:51-56
[5]Iwasawa S, Ebihara K, Ohya J, Nakatsu R and Morishima S (1997), Real-time estimation of human body postures from monocular thermal images (in Japanese). Journal of the Institute of Image Information and Television Engineers 51(8):1270-1274
[6]Asada T, Uragami K, Ikeno Y, Tanijiri T, and Yoshitomi Y (2004), A method for synthesizing computer graphics animation by transforming feature vector of posture on thermal image into 3-dimensional model. Proc. 13th IEEE Workshop on Robot and Human Interactive Communication:325-330