International Journal of Humanoid Robotics

Quantitative Evaluation Method for Pose and Motion Similarity based on Human Perception

Tatsuya Harada
Intelligent Cooperative Systems Lab., the University of Tokyo
7-3-1 Hongo Bunkyo-ku, Tokyo 113-8656, Japan
harada@ics.t.u-tokyo.ac.jp

Sou Taoka
Intelligent Cooperative Systems Lab., the University of Tokyo
7-3-1 Hongo Bunkyo-ku, Tokyo 113-8656, Japan
taoka@ics.t.u-tokyo.ac.jp

Taketoshi Mori
Intelligent Cooperative Systems Lab., the University of Tokyo
7-3-1 Hongo Bunkyo-ku, Tokyo 113-8656, Japan
tmori@ics.t.u-tokyo.ac.jp

Tomomasa Sato
Intelligent Cooperative Systems Lab., the University of Tokyo
7-3-1 Hongo Bunkyo-ku, Tokyo 113-8656, Japan
tomo@ics.t.u-tokyo.ac.jp

In the imitation research, "how much similar two motions are" is an essential problem. This paper tries to clarify the criterion by which human evaluates the similarity of poses and motions. First we proposed the quantitative pose similarity evaluation which follows human's intuition. This proposed evaluation uses the optimized parameters such as attention to each human's region and the nonlinearity of the distance perception from questionnaires of similarities for many pairs of human poses. Secondly by comparing the conventional similarity evaluation, we confirmed the validity of the quantitative pose similarity evaluation among different pairs of human poses. Lastly the quantitative pose similarity evaluation was applied to the motion similarity evaluation. Verification of that appropriateness revealed that the quantitative pose similarity evaluation is also useful for the motion similarity evaluation. Since the quantitative pose similarity evaluation reflects the human's perception, this evaluation clarifies the criterions by which human evaluates the pose or motion similarity. The clarified criterions for pose or motion similarity evaluation are as follows. 1) The distance perception is processed in three dimensional positions, 2) the relationship between the similarity and the sum of the Euclidean distances for each region is nonlinear, 3) human fixates poses symmetrically, 4) fingertips are paid particular attention, 5) human pays more attention to further regions from the hip in the upper half of the body, and 6) human pays more attention to nearer regions from the hip in the lower half of the body.

Keywords: Human Cognition, Human's intuition, Fixation, Attention, Motion Capture.

1. Introduction

Though robots have been required mainly to do tasks human gives, robots in future are required to carry out complex tasks in various fields. Because it is impractical to program robots for such all tasks, robots in future need to learn tasks autonomously. For the autonomous learning, imitation is effective learning method[1 2], since imitation is one of the most important mechanism whereby knowledge and skills are acquired between individuals. Imitation learning from humans who succeed tasks enables robots to carry out the tasks certainly and safely as the teaching human does. Such imitation learning needs recognition of human behaviors. If robots cannot recognize human behaviors at all, robots cannot imitate humans either.

The researches about recognition of human behavior have been studied in many fields. It is known that the Hidden Markov Models enable systems to recognize American Sign Language [3] or gestures [4] with good accuracy. However since there is the considerable gap between the visual information and the recognition process[5], it is necessary to carry out basic studies of the human visual perception for human motions such as human's attention to human motions, human's similarity metric between human motions and so on.

In the field of cognitive science, it is known that point light displays of only the major joints and head of human motions are enough for human to recognize the gender of the walker[6 7] and to recognize what kind of behavior the moving human does[8]. It is also known that face recognition and biological-motion perception depend on orientation of motions[9]. Prior information about display orientation of motion improves biological motion perception[10]. In brain science, it is known that the perception of goal-directed hand action and the perception of signs conveyed by expressive body movements implicate different human parietal systems and the amygdala[11]. There is the existence of neural mechanisms which analyze the kinematics defining biological motion[12 13]. Observing to recognize motions activates memory-encoding structures while observation to imitate motions activates in the regions involved in the planning and in the generation of actions[14]. Many researches about gender recognition, motion distinction, influence of orientation for motion recognition, or specifying brain activity regions have been studied, but researches about similarity metric of two motions which matches human's intuition are rarely studied. In the imitation research area, it is known that human fixates end-point when trying to imitate arm movements[15]. However how much human pays attention to each body region has not been clarified quantitatively yet.

The aim of this paper is the construction of the quantitative evaluation method for the pose and motion similarity based on the human perception. Then we try to clarify the criterion by which human evaluates the similarity of poses and motions through constructing the quantitative evaluation method for the pose and motion similarity.

One of the reasons why it is difficult to clarify how human evaluates the similarity of two motions is that motions include the element of time. Expansion or contraction of time must be considered when comparing motions. Such consideration makes the problem too complex. So in this research, we deal mainly with the similarity of poses which are instantaneous cross-sections of motions. First we try to construct the quantitative pose similarity evaluation which follows human's intuition in chapter 2 to chapter 5. We call this acquired quantitative pose similarity evaluation as 'the optimal quantitative pose similarity evaluation' in this paper. Secondly we check the validity of the optimal quantitative pose similarity evaluation in chapter 7 by comparing with the conventional pose similarity evaluation. Finally we try to verify the appropriateness of adaptation of the optimal quantitative pose similarity evaluation to motion similarity evaluation in chapter 8. Since the optimal quantitative pose similarity evaluation reflects the human's perception, this evaluation clarifies the criterions by which human evaluates the pose or motion similarity.

2. Basic Quantitative Pose Similarity Evaluation

In order to obtain the quantitative pose similarity evaluation which follows human's intuition, we mainly consider two points.

(1) Pose description methods
(2) Weights of attention to body regions on comparing two poses

First, we list candidates of human's pose describing methods on joint angle space or three-dimensional position space. Secondly, we calculate distances of poses considering weights of attention to body regions under each method of describing poses. These calculated similarities become a basis of the quantitative pose similarity evaluation. We choose the best method of describing poses and adjust weights of attention so that these distances agree with human's intuitive similarities. The human's intuitive similarities are obtained through questionnaires for a set of pairs of poses by some subjects. The selection of the best pose describing method and the adjustment of optimal weights will construct the optimal quantitative pose similarity evaluation. We go into the details in the following.

2.1. *Selection of Pose Description Method*

Quantitative pose similarity evaluation needs description of poses numerically. This research tries to select the pose description method and clarify how human describe poses when evaluating the pose similarity. We investigate the following 4 pose description method.

- 3-dimensional joint angle representation
- 3-dimensional position representation
- 2-dimensional position representation
- 2-dimensional joint angle representation

Fig.1 shows above 4 methods graphically. The prepared set of pairs of poses is represented differently through 4 description methods. The joint angle representation means that poses are defined by the link-joint model and described in the joint angle space. The position representation means that poses are described as the relative positions of each body region to a hip position. The 2-dimensional representations means that poses are projected on a display and described in the display coordinate. In the 2-dimensional representation, the same poses are differently described according to the projection directions.

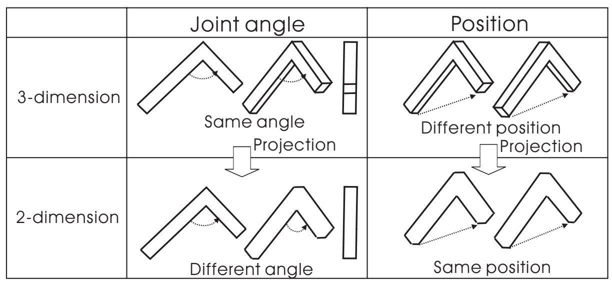

Fig. 1. 4 pose description methods are shown. The upper left shows 3-dimensional joint angle representation. The upper right shows 3-dimensional position representation. The lower left shows 2-dimenshonal joint angle representation. The lower right shows 2-dimentional position representation. Projection of 3-D objects reduces 1 dimension and makes them 2-D objects. Provided that a 3-D object remains constant, the projected 2-D object changes according to the direction in which the 3-D object is projected. If two projected 2-D objects are the same, two original 3-D objects are not necessarily the same.

2.2. *Distances Calculation between Poses with Weights of Attention to Body Regions*

It is hard to think that human watches each body regions with equal attention. So we consider weights of attention to body regions in distance calculation between poses. The optimized weights are thought to clarify how much attention human pays to each region when evaluating the pose similarity.

The distance between poses A and B is represented as

$$D(A,B) = \sum_k W_k P_k(A,B) \tag{1}$$

where W_k is the weight of the body region k and $P_k(A,B)$ is calculated differently in each pose description method. $P_k(A,B)$ for each pose description method is written as

$$P_k(A,B) = \begin{cases} 1 - \|q_{Ak} \cdot q_{Bk}\| & \text{3D joint angle} \\ (X_{Ak} - X_{Bk})^2 + (Y_{Ak} - Y_{Bk})^2 + (Z_{Ak} - Z_{Bk})^2 & \text{3D position} \\ (X'_{Ak} - X'_{Bk})^2 + (Y'_{Ak} - Y'_{Bk})^2 & \text{2D position} \\ |\delta\theta_k(A,B)|^2 & \text{2D joint angle} \end{cases}$$

where q_{Ak} is the quaternion [16] of the body region k of the pose A. X_{Ak} is the x-axis of the body region k of the pose A. X'_{Ak} is the x-axis of the body region k of the projected pose A. $\delta\theta_k(A, B)$ is calculated as

$$\begin{aligned} \delta\theta_k(A, B) &= \theta_k(A) - \theta_k(B) \\ &\quad if \;\; \delta\theta < -\pi \;\; then \;\; \delta\theta = \delta\theta + 2\pi \\ &\quad if \;\; \delta\theta > \pi \;\; then \;\; \delta\theta = \delta\theta - 2\pi \end{aligned}$$

where $\theta_k(A)$ is the angle of the body region k of the projected pose A. Fig. 2 shows the names of the body regions in this paper.

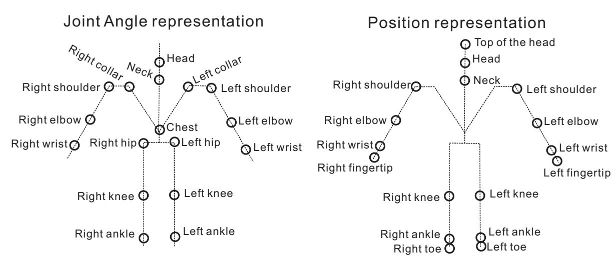

Fig. 2. The names of body regions are shown. Body regions dealt with each pose description method are different. The body regions of the left body are dealt with in joint angle representations. The body regions of the right body are dealt with in position representations. The number of body regions is always 17.

In each set of pairs of poses, the pose distances are standardized. The standarized quantitative similarity Q_i of ith pairs of poses A_i and B_i is written as

$$Q_i = \frac{D(A_i, B_i) - \mu}{\sigma}$$

where μ and σ are the mean and the standard deviation for the set of pairs of poses respectively.

2.3. *Human's Intuition for Pose Similarity*

The human's intuition for the pose similarity is acquired by questionnaires for the prepared set of pairs of poses from ten or more subjects. We showed subjects two human figures which take a pair of poses side by side horizontally on the PC display. The surface skins were attached on the stick models in order that subjects can recognize poses easily. Metallic skins were chosen so as not to let subjects imagine the human figure's sex, age, race and so on. Figures face front to be seen by subjects easily. Fig. 3 shows 2 examples of poses actually shown to subjects.

Subjects evaluate the similarity of pairs of poses as 4 ranks evaluation, 'approximately same', ' relatively similar', 'relatively dissimilar' and 'dissimilar'. Each rank is expressed as the numerical similarity from 1 to 4 respectively. About Fig. 3, the

Fig. 3. Two examples of pairs of poses shown to subjects on the display. Human judges the right pair of poses similar and the left pair of poses dissimilar.

left pair of poses was evaluated as 'dissimilar' while the right pair of poses was evaluated as 'approximately same'.

In one set of pairs of poses, the similarities of each subject are standardized. The intuitive similarity I_{ij} of ith pair of poses by the subject j is

$$I_{ij} = \frac{E_{ij} - \mu_j}{\sigma_j}$$

where E_{ij} is the similarity of ith pair of poses, μ_j is the mean of the similarities for the set of pairs of poses and σ_j is the standard deviation for the set of pairs of poses by the subject j. We regard the mean of the standardized similarities ($I_i = \frac{1}{n}\sum_j^n I_{ij}$) for ith pair of poses among n subjects as the human's intuition for pose similarity of the ith pair of poses.

2.4. *Selection of Set of Pairs of Poses*

We prepare a set of pairs of poses to evaluate similarities. The set has 200 pairs of poses including exceptions as symmetric poses said later. Whole body poses were obtained with the mechanical motion capture (Gypsy, META Motion Corp.) and described in the BVH format. The poses are represented by stick models with 57 degrees of freedom. The parameters of the stick model are lengths of links and Euler angles of joints.

The poses are selected under certain conditions. The first condition is that the physical proportion of the poses is fixed. For every poses, the ratio of height and the length of each links are constant (the height is $1.75[m]$). Second condition is that the translation and the rotation of the root of the body are ignored.

To make sure to avoid bias in the set of pairs of poses, the provisional similarities (D_p) are calculated with Eq. (1) under the conventional conditions. The conventional conditions are

- Every weight of a region is the same value ($W_k = const$).

- The pose description method is 3-dimensional joint angle representation.

We classified pairs of poses with the logarithm of the provisional similarity into 5 classes. The pose group was chosen so that the number of pairs of poses in each class is equally 40.

There are some symmetrical relationships between two poses such as point symmetry and line symmetry. There is a possibility that human's evaluation for the pose similarity is influenced by such symmetry. In this paper, we picked line symmetry because point symmetry of poses is included in the rotation of poses. Fig. 4 shows an example of a line symmetric pair of poses. The prepared pose set includes 50 symmetric pairs of poses.

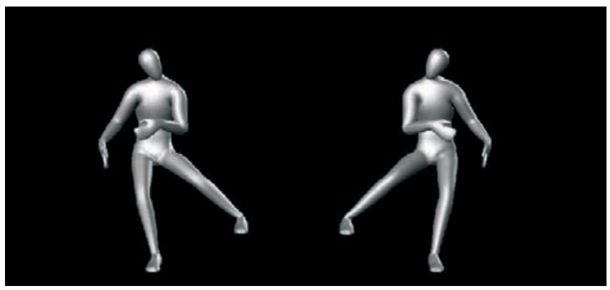

Fig. 4. Left pose and right pose are an example of symmetric pair of poses.

If human evaluates the similarity of a symmetric pair of poses specially, human compares not only two poses but also symmetric poses of them. Therefore on considering the symmetry of a pair of poses, a symmetric pose should be introduced in the calculation of the pose similarity. We adopted Minimum Mirror-Body Distance and Minimum Mirror-Region Distance to consider the symmetry.

Given a pair of poses A, B and A' which is a symmetric pose of A, the Minimum Mirror-Body Distance $D_{mb}(A, B)$ between A and B is calculated as follows:

$$D_{mb}(A, B) = \min \{D(A, B), D(A', B)\}$$

The Minimum Mirror-Region Distance $D_{mr}(A, B)$ between A and B is calculated as follows:

$$D_{mr}(A, B) = \sum_k W_k \min \{P_k(A, B), P_k(A', B)\}$$

As a result of adaptation of Minimum Mirror-Body Distance and Minimum Mirror-Region Distance, there existed 3 kinds of the distances between two poses such as $D(A, B)$, $D_{mb}(A, B)$ and $D_{mr}(A, B)$. Because there are four methods of describing pose, similarity of a pair of poses are calculated in $3 \times 4 = 12$ ways. One of them was selected so that agreement between the human's intuition for pose similarity and the quantitative pose similarity evaluation becomes the highest.

The selected distance is thought to be the distance which human uses mainly when evaluating the pose similarity.

2.5. *Correlation Coefficient between Intuitive Similarities and Quantitative Similarities*

Agreement between the intuitive pose similarity and the quantitative pose similarity is measured by the correlation coefficient between these similarities. The high correlation coefficient means that the quantitative similarity well accords with the intuitive similarity. The correlation coefficient ρ is written as

$$\rho = \frac{\sum_{i=1}^{m}(I_i - \bar{I})(Q_i - \bar{Q})}{\sqrt{\sum_{i=1}^{m}(I_i - \bar{I})^2}\sqrt{\sum_{i=1}^{m}(Q_i - \bar{Q})^2}}$$

where Q_i is a quantitative similarity, I_i is a intuitive similarity of ith pair of poses, m is the number of pair of poses, $\bar{Q}$ is the mean of the quantitative similarity ($\bar{Q} = \frac{1}{m}\sum_{i=1}^{m} Q_i$) and $\bar{I}$ is the mean of the intuitive similarity ($\bar{I} = \frac{1}{m}\sum_{i=1}^{m} I_i$). A correlation coefficient always satisfies $-1 \leq \rho \leq 1$.

2.6. *Optimization of Parameters in Quantitative Pose Similarity Evaluation*

We employed an exponential simulated annealing[17] to optimize the parameters of the quantitative pose similarity evaluation. The purpose of the simulated annealing is to optimize the parameters so that the energy becomes the lowest. The simulated annealing starts under some state of parameters $W(1)$ which includes weights of body regions and continues polling until stopping criterion meets. In nth polling, one parameter in the state $W(n)$ is chosen randomly and changed. The state $W'(n+1)$ whose one parameter is changed is a candidate for new state $W(n+1)$. If $\rho\,(W'(n+1))$, which is the correlation coefficient between intuitive similarities and quantitative similarities under the state of parameters $W'(n+1)$, is higher than $\rho\,(W(n))$, the state accepts the change of one parameter $(W(n+1) = W'(n+1))$. If not, the state accepts the change with the probability a_{n+1}, which has variables such as the difference between the correlation coefficients under each state and the temperature $t(n+1)$. The last state $W(N)$ includes optimal parameters.

In this research, the starting state $W(1)$ is the weights of body regions $\{\forall k;\ W_k = const\}$. In nth polling, the body region k is chosen randomly. Then $W_k(n)$ is changed randomly. After $\rho\,(W(n))$ and $\rho\,(W'(n+1))$ are calculated, the acceptance probability a_{n+1} is calculated as

$$a_{n+1} = \max\left[\exp\left\{\frac{\rho\,(W_{n+1}) - \rho\,(W_n)}{t(n+1)}\right\}, 1\right]$$

The temperature $t(n+1)$ is γ^{n+2}, where γ is the temperature reduction factor and is 0.9999 to cool the temperature slowly enough. Stopping criterion is $n = 100000$.

2.7. *Result for Basic Quantitative Pose Similarity Evaluation*

Table 1. The numbers are the correlation coefficient between the calculated similarities and the human's similarities. The leftmost row show what distance was used in calculation of the similarities. The first column show what methods of describing pose was used in calculation of the similarities. (*1: 3-Dimensional Joint Angle Representation, *2: 3-Dimensional Position Representation, *3: 2-Dimensional Position Representation, *4: 2-Dimensional Joint Angle Representation)

	*1	*2	*3	*4
Normal Distance $D(X,Y)$	0.710	**0.805**	0.768	0.765
Minimum Mirror-Body Distance $D_{mb}(X,Y)$	0.705	0.762	0.763	0.784
Minimum Mirror-Region Distance $D_{mr}(X,Y)$	0.713	0.732	0.726	0.766

Table 1 shows the correlation coefficients between the calculated similarities and the human's similarities when parameters are optimized severally. This result reveals that the correlation coefficient is the highest when 1) the pose description method is 3-dimensional position representation and 2) the normal distance is used in calculation of the similarities. Such pose description method and distance of each body region between poses is regarded to be used by human mainly when evaluating the pose similarity. We use 1) and 2) for quantitative similarities after this.

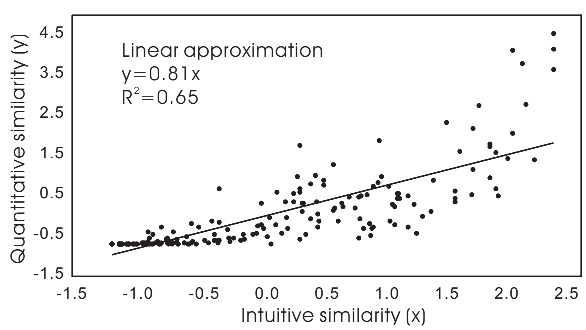

Fig. 5. The intuitive similarities and the quantitative similarities with optimal weights of regions about the pose group in which 50 of 200 pairs of poses are symmetric. The quantitative similarities tend to be low about the pairs of poses whose intuitive similarities are low. But approximation straight line cannot approximate the relationship of the quantitative similarities and the intuitive similarities well. There are thought to be two reasons: 1) The quantitative similarities aren't linearly dependent of the intuitive similarities. 2) Some quantitative similarities are too high to be said to be within the suitable range.

Fig. 5 shows the relationship between the quantitative similarities and the intuitive similarities. Fig. 5 reveals that the relationship between those similarities is nonlinear. It is desirable for the relationship to be linear. The reason of this nonlinear relationship is thought that the Euclidean distances $\sqrt{P_k(A,B)}$ are squared invariably in Eq. (1). Human categorizes the pairs of poses whose distances exceed a certain upper limit as one group. Therefore the human's intuition for pose similarity has an upper limit. On the other hand, the basic quantitative pose similarity evaluation does not have an upper limit. The upper limits are also thought to affect the nonlinear relationship. So we should add new parameters to the weights of body regions in the calculation of the pose similarity.

3. New Parameters for Quantitative Pose Similarity Evaluation

3.1. *Power of Euclidean Distance and Upper Limit of Distance*

We add new parameters in the calculation of the new distance of poses $D'(A,B)$ as shown in Eq. (2). New parameters are r and M. r is the power of the Euclidean distance. M is the upper limit of the weights of body regions.

$$D'(A,B) = \sum_k \min\left\{ \left(W_k \sqrt{P_k(A,B)}^{\,r} \right), M \right\} \tag{2}$$

Optimized r means how human recognizes the Euclidean distance on evaluating the pose similarities. Optimized M means the upper limit of the distance perception over which human judges pairs of poses dissimilar.

3.2. *Result of New Parameters r and M*

Employing simulated annealing, we optimized the parameters $W_k,\ r$ and M so that the correlation coefficient $\rho(W,\ r,\ M)$ between the similarities became the highest. In one polling, one of W_k, r and M was changed randomly. The set of pairs of poses is same as the set used in section 2.

Optimal parameters and the correlation coefficient are shown in Fig. 6. A dotted line shows the stick model. A center of a circle means position of the body region. A diameter means the weight of the body region. Fig. 6 reveals that the weight of each region is almost symmetric in a pose.

4. Symmetric Weights of Body Regions

From the result of Fig. 6, though the weights of the body regions are independent of each other, we can obtain the symmetric weights for fingertips, wrists, elbows, shoulders, knees, ankles and toes. We apply the symmetric weights for the symmetric body regions. For example, weights for left-fingertip and right-fingertip are combined into one weight as fingertips. As a result, the number of weights becomes 10 from 17.

4.1. *Optimization of Symmetric Weights*

We optimized parameters (symmetric weights of regions W_k, the power of the Euclidean distance r and the upper limit M) so that the correlation coefficient between

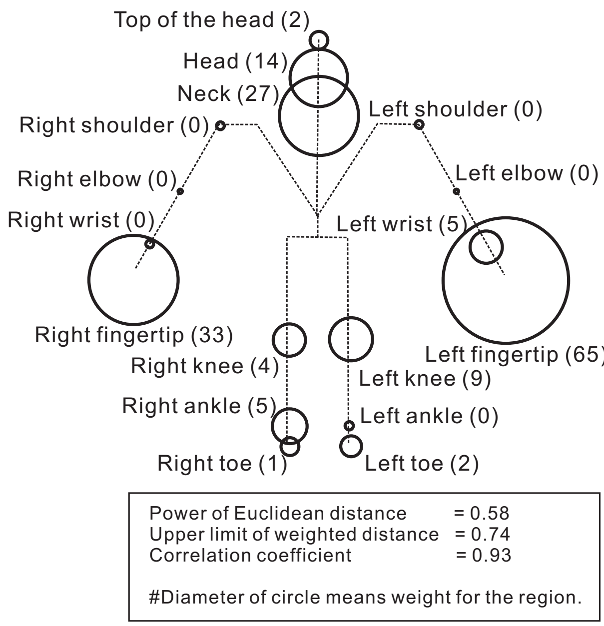

Fig. 6. The optimal parameters and the correlation coefficient between the quantitative similarities and the intuitive similarities about the pose group in which 50 of 200 pairs of poses are symmetric. Parameters are the weights of regions W_k, the power of the Euclidean distance r and the upper limit of the weights M. Each weight of a region is represented as the length of the diameter of the circle whose center is at the region. The name of each region is written around the region. The number written after the name of the region shows the rounded-off optimal weight of the region.

the similarities also becomes the highest. The set of pairs of poses is same as the set used in section 2 too.

As a result of optimization, the correlation coefficient between the similarities was 0.924. The decrease of the correlation coefficient was no more than 0.001 though the number of weights lessened from 17 to 10. This result guarantees validity of symmetric weights and reveals that human pays attention to body regions of a pose symmetrically.

Fig.7 shows the relationship between the similarities. Comparing Fig.5, the relationship between the similarities becomes linear. This result demonstrates the effectiveness of the adoption of new parameters r and M. However about some pairs of poses, the similarities are mismatched (Human judges the pairs of poses relatively similar. The quantitative pose similarity evaluation judges them relatively dissimilar). It is thought that we overlook one or more important factor. Fig. 8 shows such pairs of poses whose similarities are mismatched. Fig. 8 reveals most of them (except for only one pair of poses at the upper left) are the pairs of poses which are symmetric each other.

Fig. 9 shows the relationship between the similarities about only symmetric 50 pairs of poses which are included in the prepared pose set. Fig. 9 reveals that human

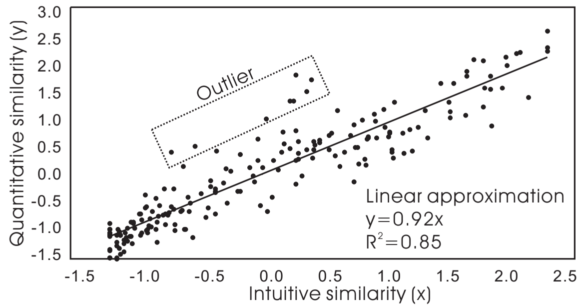

Fig. 7. The relationship between the intuitive and the quantitative similarities with optimal parameters as symmetric weights of regions, the power of the Euclidean difference and the upper limit of weights about the pose group in which 50 of 200 pairs of poses are symmetric. About most of pairs of poses the plotted dots are around the approximation straight line. But about some pairs of poses the plotted dots are far from the approximation straight line.

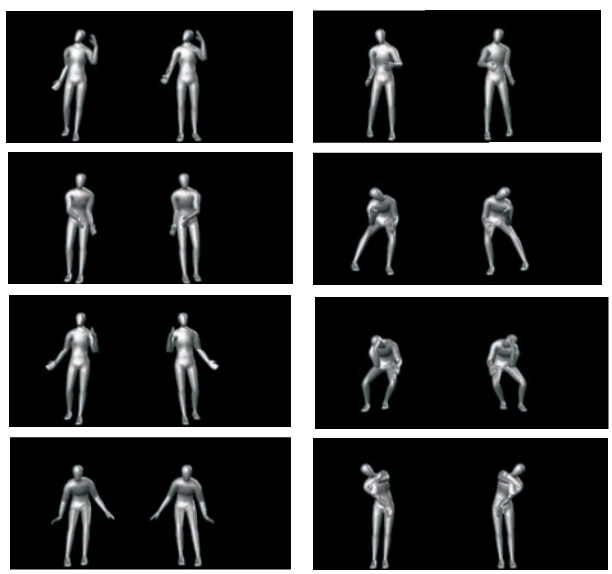

Fig. 8. Pairs of poses of which calculated similarities and human's similarities are mismatched. 7 of 8 pairs of poses are symmetric.

tends not to judge 'dissimilar' to completely symmetric pairs of poses. However because the similarities of 43 pairs of poses out of 50 (except 7 pairs of poses shown in Fig. 8) are evaluated properly with the optimal parameters, it is thought to be hard for the optimal quantitative pose similarity evaluation to contain those exceptions.

Fig.10 shows the optimal parameters and the correlation coefficient of the result for symmetric weights. Some points are read in Fig.10. 1) The weight of fingertips

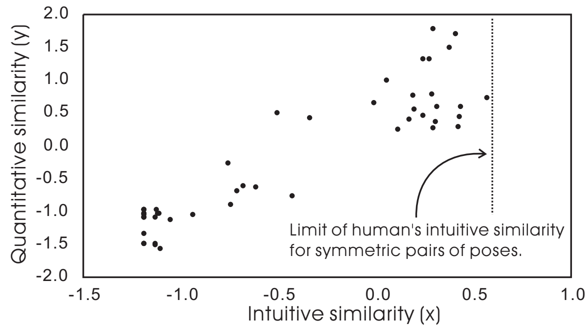

Fig. 9. The calculated similarities and the human's similarities about completely symmetric pairs of poses. Because the calculated similarities are linearly dependent of the human's similarities about the whole pairs of poses, some pairs of poses which the quantitative pose similarity evaluation judges similar are judged similar by human, too. But the other poses which the quantitative pose similarity evaluation judges dissimilar are judged not dissimilar by human. Thus human doesn't judges dissimilar about symmetric pairs of poses.

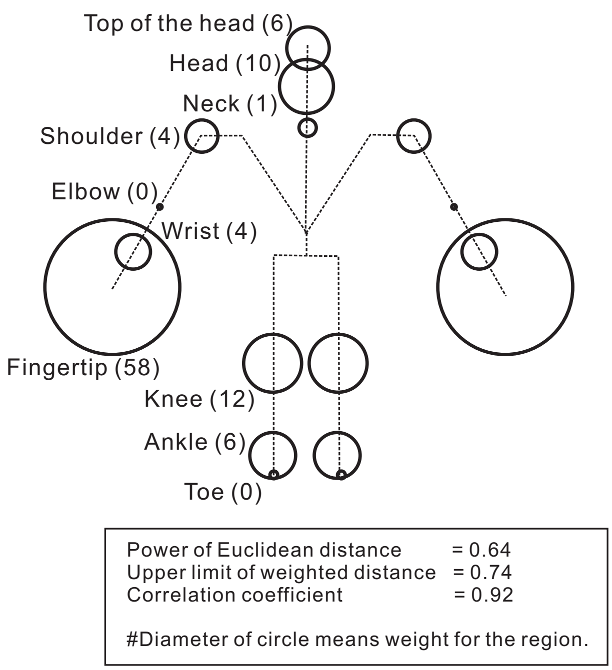

Fig. 10. The optimal parameters and the correlation coefficient between the quantitative similarities and the intuitive similarities about the pose group in which 50 of 200 pairs of poses are symmetric. Parameters are the symmetric weights of regions W_i, the power of the Euclidean distance r and the upper limit of the weights M.

is heavy. 2) The weights of farther regions from the hip are heavier in the upper half of the body. On the other hand the weights of nearer regions from the hip are heavier in the lower half of the body. 3) The relationship between the weights of the head and the top of the head is different from the relationship in Fig. 6. About 3), there is a possibility that parameters are inaccurately optimized in this experiment.

There is a possibility that the weights of some body regions that have small influence for pose similarities are not assigned appropriately if small number of body regions have large influence for the similarities. Fig. 10 may show such effects respectively. 1) The influence of fingertips hided the influence of wrists and elbows. 2) The influence of upper half of the body hided the influence of lower half of the body and vice versa. 3) The influence of head hides the influence of the top of the head and vice versa. So another experiment is necessary to clarify the effect of joints that have small influence for pose similarities.

5. Appropriate Weights Allocation for Each Body Regions

To avoid the problem that some weights are not assigned appropriately, it is necessary to minimize the correlation coefficient between Euclidean distances of the body regions that have the dependent relationship for the pose similarities among the set of pairs of poses. Pairs of body regions that have such dependent relationship are fingertips and wrists, fingertips and elbows, head and top of the head and the upper half of the body and the lower half of the body. By utilizing the pairs of poses that are selected to lower the correlation coefficient, the weight of the focused body region can be clarified without the effects of other body regions.

5.1. *New Set of Pairs of Poses*

To allocate appropriate weights for each body region, a new set of pairs of poses is prepared which is selected among 5 different classes. In the first class, pairs of poses have the identical upper halves of the body and the randomly selected lower halves of the body. In the second class, pairs of poses have the identical lower halves of the body and the randomly selected upper halves of the body. In the third class, the Euclidean distances of fingertips between two poses are less than 6[`cm`]. In the fourth class, the lower halves of the body were identical and the Euclidean distance of fingertips between two poses are less than 6[`cm`]. In the fifth class, there is no restriction. The new set of pairs of poses contains 200 pairs of poses where 40 pairs of poses were selected from each class. In addition to those conditions, those 200 pairs of poses are selected so that the correlation coefficient between the Euclidean distances of head and top of the head regions among the pairs of poses becomes low.

5.2. *Result*

We acquire the human's intuition for pose similarities about the new set of pairs of poses in the same way as section 2. Parameters such as the symmetric weights, the power of the Euclidean difference and the upper limit of the weights are optimized to maximize the correlation coefficient between the similarities. Fig.11 shows optimal parameters and the correlation coefficient about the new set of pairs of poses.

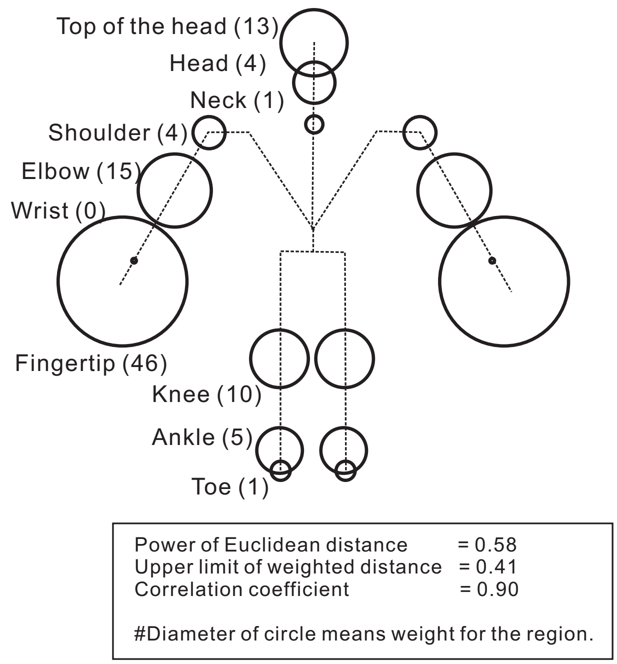

Fig. 11. The optimized parameters and the correlation coefficient between the calculated similarities and the human's similarities about the pose group in which there are separated regions. Parameters are the symmetric weights of regions W_i, the power of the Euclidean distance r and the upper limit of the weights M.

The weights of body regions which have dependent relationship are made to be clear. The similar results between Fig.10 and Fig.11 for the power r, the upper limit M and the weight W_i, suggests that these results are independent of the set of pairs of poses. If another set of pairs of poses is given, the quantitative pose similarity evaluation following these results should be able to evaluate the similarities which follow human's intuition. Therefore we construct the optimal quantitative pose similarity evaluation following these experimental results.

6. Optimal Quantitative Pose Similarity Evaluation

We defined the optimal quantitative pose similarity evaluation in this paper according to the results from section 2 to section 5. The characteristics of the optimal quantitative pose similarity evaluation are as follows.

- The pose description method is 3-dimensional position representation.

- Symmetricity of pairs of poses is not considered.
- The power of the Euclidean distance r is 0.6.
- The upper limit of the weights M is 0.5[m] converted the fingertips' Euclidean distances.
- The weights of body regions are symmetric and defined mostly following the result in section 5.2. The weights of body regions are shown in Table 2.

Table 2. The weights of regions in the optimal quantitative pose similarity evaluation. The contents in the first row mean the name of the region. The contents of second row mean the weight of the region.

knee	ankle	toe	shoulder	elbow	wrist	fingertip	neck	head	headtop
10	5	1	4	15	0	50	1	5	10

The correlation coefficient between the similarities is 0.916 among the original set of pairs of poses and 0.902 among the new set of pairs of poses. These correlation coefficients are thought to be high enough because the mean of the correlation coefficients between each subject's similarity and human's intuition for pose similarity is 0.849.

7. Validity of Optimal Quantitative Pose Similarity Evaluation

In this section, we try to verify the validity of the optimal quantitative pose similarity evaluation. First we prepare a different set of pairs of poses from the sets in section 2 and section 5. Secondly intuitive similarities are acquired about this set of pairs of poses. Thirdly the similarities are calculated by the optimal quantitative pose similarity evaluation and the conventional quantitative pose similarity evaluation respectively. Fourthly we compare the correlation coefficients between the similarities between the human's intuition for pose similarities and the quantitative similarities by the optimal evaluation or the conventional evaluation. If the correlation coefficient between the similarities from the optimal evaluation is far higher than the conventional evaluation, the validity of the optimal quantitative pose similarity evaluation can be verified.

7.1. *Set of Pairs of Poses*

The set of pairs of poses for this experiment also contains 200 pairs of poses. To make sure to avoid biases in the set of pairs of poses, we classified pairs of poses into 10 classes according to the similarities by the optimal quantitative pose similarity evaluation. Pairs of poses in this set are chosen so that the number of pairs of poses in each class is equally 20.

7.2. *Result for Validity of Optimal Quantitative Pose Similarity Evaluation*

We calculated the similarities by the optimal quantitative pose similarity evaluation and the conventional quantitative pose similarity evaluation. The experimental result of the correlation coefficient between the similarities by the optimal and conventional quantitative evaluations and human's intuition is shown in Table 3.

Table 3. The correlation coefficients between the intuitive similarities and the quantitative similarities by the optimal quantitative pose similarity evaluation and the conventional quantitative pose similarity evaluation about the new pose set. The optimal quantitative pose similarity evaluation agrees more with human's intuition than the conventional quantitative pose similarity evaluation.

	Correlation Coefficient
Optimal Quantitative Pose Similarity Evaluation	0.897
Conventional Quantitative Pose Similarity Evaluaion	0.692

Table 3 shows the correlation coefficient between the human's intuition and the quantitative similarities of the optimal evaluation is far higher than that of the conventional evaluation. Moreover the correlation coefficient from the optimal quantitative pose similarity evaluation is thought to be high enough because the mean of the correlation coefficients between each subject's similarities and the human's similarities is 0.854. These results verify the validity of the optimal quantitative pose similarity evaluation.

8. Adaptation of Optimal Quantitative Pose Similarity Evaluation to Quantitative Motion Similarity Evaluation

In this chapter we tried to verify the appropriateness of adaptation of the optimal quantitative pose similarity evaluation to quantitative motion similarity evaluation. Unlike the pose, the motion includes element of time. It is surmised that human doesn't recognize motions as only a set of poses. Because the motion has velocity, acceleration, contexts, a position and a pose of a whole body. Based on above things, this experiment calculates the quantitative similarities considering only poses and compares them with the intuitive similarities. If a pose is an important basic element of a motion, these two similarities have high correlation except for other elements of a motion. The purpose of this experiment is to verify the appropriateness of adaptation of the optimal quantitative pose similarity evaluation when evaluating similarities of human's imitating motion.

8.1. *Set of Pairs of Motions*

We obtain 25 original motions and 9 imitating motions to each original motion written in BVH form from our previous results[18]. There are 250 pairs of motions. Every motion is 5 seconds long, whose sampling time is [33msec]. All original motions are picked from everyday life motions in a study. Intuitive similarities are acquired through questionnaires by 5 ranks evaluation from 8 subjects[18].

8.2. *Alignment of Time between Motions*

Alignment of time is necessary to adapt the optimal quantitative pose similarity evaluation to the motion similarity. It is known that Hidden Markov Models[19] and DP matching[20] can align time. We align time based on DP matching since the pose similarity can be easily implemented into the DP matching. We introduced the expansion DP matching[21] and the continuous DP matching[22] to calculate the motion similarity.

8.3. *Result*

We calculated the correlation coefficients between the human's intuition and the quantitative similarities by the optimal and conventional evaluations. The experimental results are shown in Table 4.

Table 4. The correlation coefficients between the intuitive similarities and the quantitative similarities by the optimal quantitative pose similarity evaluation and the comparison quantitative pose similarity evaluation. The calculated similarities are calculated using expansion DP matching and continuous DP matching.

	Optimal Quantitative Pose Similarity Evaluation	Comparison Quantitative Pose Similarity Evaluation
Expansion DP Matching	0.731	0.693
Continuous DP Matching	0.733	0.539

Table 4 reveals the correlation coefficient between the intuitive intuition and the quantitative similarities by the optimal evaluation is higher than that of the conventional evaluation for both DP matchings. It is thought that this result verifies the appropriateness of adaptation of the quantitative pose similarity evaluation to quantitative motion similarity evaluation.

9. Conclusion

In this paper we construct the quantitative pose similarity evaluation which follows human's intuition and clarify the criterion by which human judges the similarity of a pair of poses. We also verify the appropriateness of the adaptation of the quantitative pose similarity evaluation to the quantitative motion similarity evaluation. The quantitative pose similarity evaluation is constructed by the optimization of the parameters and the selection of the pose description methods to maximize the correlation coefficient between the human's intuition for pose similarities and the quantitative pose similarity evaluation. The candidate methods of describing poses are 3-dimentional joint angle representation, 3-dimentional position representation, 2-dimentional joint angle representation and 2-dimentional position representation. The parameters are weights of attention to body regions, the power of the Euclidean distance and the upper limit of the weights of body regions. The acquired quantitative pose similarity evaluation follows considerably human's intuition. The correlation coefficients between the quantitative similarities and human's intuition are approximately 0.9. This is higher than the mean of correlation coefficients between the intuitive similarities and each subject's similarities. Moreover we verified the appropriateness of adaptation of the quantitative pose similarity evaluation to quantitative motion similarity evaluation with comparing to the conventional pose similarity evaluation.

The acquired quantitative pose similarity evaluation clarifies the tendencies of human on evaluating the similarities of poses. The tendencies are as follows.

- Fingertips are fixated most.
- Human represents poses as 3-dimensional positions of body regions.
- The relationship between the human's intuition and the Euclidean distance of two poses is nonlinear.
- Each region is paid attention to symmetrically.
- Human pays more attention to farther regions from a hip in the upper half of a body and nearer regions from a hip in the lower half of a body.
- Complete symmetric pairs of poses are not judged dissimilar.

In this paper, we removed some elements from poses and motions such as physical proportion, rotation of whole body and so on. However to evaluate the similarity of poses precisely and wide-usefully, we should add such parameters to the quantitative evaluation. Moreover we should also add parameters which have specific significance such as the gazing directions. We verified the adaptation from pose similarity to motion similarity. However the actual quantitative motion similarity evaluation needs the characteristic of parameters to the motion such as velocity, acceleration and so on.

References

1. Stefan Schaal. Is imitation learning the route to humanoid robots? *Trends in Cognitive Science*, Vol. 3, pp. 233–242, 1999.
2. A. Alissandrakis, C. Nehaniv, and K. Dautenhahn. Learning how to do things with imitation. In *AAAI Fall Symposium on Learning How to Do Things*, pp. 1–6, 2000.
3. Thad Starner, Joshua Weaver, and Alex Pentland. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 12, pp. 1371–1375, 1998.
4. Aaron F. Bobick and Andrew D. Wilson. A state-based approach to the representation and recognition of gesture. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 12, pp. 1325–1337, 1997.
5. M. Kawato. Nou no Keisanriron. Sangyoutosho, 1996.

6. Lynn T. Kozlowski and James E. Cutting. Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychopysics*, Vol. 21, No. 6, pp. 575–580, 1977.
7. James E. Cutting. Generation of synthetic male and female walkers through manipulation of a biomechanical invariant. *Perception*, Vol. 7, No. 4, pp. 393–405, 1978.
8. Winand H Dittrich. Action categories and the perception of biological motion. *Perception*, Vol. 22, No. 1, pp. 15–22, 1993.
9. Nikolaus F Troje. Reference frames for orientation anisotropies in face recognition and biological-motion perception. *Perception*, Vol. 32, No. 2, pp. 201–210, 2003.
10. Marina Pavlova and Alexander Sokolov. Prior knowledge about display inversion in biological motion perception. *Perception*, Vol. 32, No. 8, pp. 937–946, 2003.
11. Eva Bonda, Michael Petrides, David Ostry, and Alan Evans. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, Vol. 16, No. 11, pp. 3737–3744, 1996.
12. E. Grossman, M. Donnelly, R. Price, D. Pickens, V. Morgan, G. Neighbor, and R. Blake. Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, Vol. 12, No. 5, pp. 711–720, 2000.
13. E.D. Grossman and R. Blake. Brain activity evoked by inverted and imagined biological motion. *Vision Research*, Vol. 41, pp. 1475–1482, 2001.
14. J. Decety, J. Grezes, N. Costes, D. Perani, M. Jeannerod, E. Procyk, F. Grassi, and F. Fazio. Brain activity during observation of actions influence of action content and subject's strategy. *Brain*, Vol. 120, pp. 1763–1777, 1997.
15. M.J. Matarić and Marc Pomplun. Fixation behavior in observation and imitation of human movement. *Cognitive Brain Research*, Vol. 7, No. 2, pp. 191–202, 1998.
16. James Kuffner. Effective Sampling and Distance Metrics for 3D Rigid Body Path Planning. Proc. 2004 IEEE Int'1 Conf. on Robotics and Automation (ICRA 2004), 2004.
17. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, Vol. 220, No. 4598, pp. 671–680, 1983.
18. T. Harada, H. Takamatsu, S. Taoka, T. Mori, T. Sato. Difficulties of Actions and Essential Observable Body Parts on Imitation. The 21st Annual Conference of THE ROBOTICS SOCIETY OF JAPAN, (2003), 2J25.
19. L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, Vol. 77, No. 2, pp. 257–286, 1989.
20. Robert A. Wagner and Michael J. Fischer. The string-to-string correction problem. *Journal of the ACM*, Vol. 21, No. 1, pp. 168–173, 1974.
21. H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 26, No. 1, pp. 43– 49, 1978.
22. Takuichi Nishimura, Hiroaki Yabe, and Ryuichi Oka. A method of model improvement for spotting recognition of gestures using an image sequence. *New Generation Computing*, Vol. 18, No. 2, pp. 89–101, 2000.