Motion Retargeting for the Hand Gesture

Chunbao Ge[1]
Bejing 100020, China
hagcb@163.com

Yiqiang Chen[2]
Bejing 100080, China
YiqiangChen@ict.ac.cn

Changshui Yang[2]
Bejing 100080, China
Changshui@ict.ac.cn

Baocai Yin[1]
Bejing 100020, China
Ybc@bjut.edu.cn

Wen Gao[2]
Bejing 100020, China
WGao@ict.ac.cn

ABSTRACT

This paper presents a new technique for retargeting the sign language data captured from motion capture device to different characters with different sizes and proportions. Realistic and natural animations can be produced to express similar meanings to the original. The proposed method first defines many sensitive points on the human body and selects the key sensitive points through analyzing the importance of the sensitive points. Next a novel mapping method based on relative position is presented to adapt the original sensitive points to the target sensitive points. Finally we utilize an IK solver to realize the retargeting problem. Experimental results show that the proposed method dramatically improves the recognition rate about 30%.

Keyword

Character animation, motion retargeting, key sensitive points, IK, Chinese Sign Language

1. INTRODUCTION

Recently, motion capture has become one of the most promising technologies in character animation. Realistic motion data can be captured by recording the movement of a real actor with a motion capture system, and motion retargeting will adapt these motion data to new character. While the target character is different from the original one, the target character is likely to lose desire features of original motion. The problem can be solved through motion retargeting technology.

Many solutions to motion retargeting have been presented for different applications. Conventional retargeting techniques seldom consider the accurate meanings expressing in motion data. For example, in the sign language, very small changes in the hand postures will lead to wrong meanings. Our task is to retarget sign language motion data to new characters and guarantee the similarities of the meanings. The most important characteristics for the sign language are the precise position relations between the hand and the other parts of the human body, so we must analyze and acquire these important characteristics.

WSCG 2005 Conference proceedings ISBN 80-903100-7-9
WSCG'2005, January 31-February 4, 2005
Plzen, Czech Republic.

Figure 1. Sign language of "human", "eye", "rectangle"

Figure 1 shows several key-frames of sign language (This model was downloaded from Miralab, and we only use it for demonstration as a standard virtual human model according to VRML). The left shows "human". The middle shows "eye". The right shows "rectangle". For the left and the right the most important feature is the relative position between two hands. For the center the important feature is the relative position between the hand and the head, and the meaning will change if the forefinger points to other position.

---

1. Multimedia and Intelligent Software Technology Beijing Municipal Key Laboratory, Beijing University of Technology.
2. Institute of Computing Technology, CAS.

This paper presents a new technique for retargeting sign language data captured from motion capture device to different characters with different sizes and proportions. We can produce realistic and natural animations that express similar meanings to the original. We begin our discussion by providing a related work of our method in next section. Next we introduce the model of the human upper limb and reprocessing of the captured data. In section 3 we introduce feature analysis for the hand gesture. In section 4 we present a novel mapping method based on the relative position of the sensitive points. In section 5 we introduce our IK solver. Finally, we show several experimental results and conclude the paper.

2. RELATED WORKS

Several techniques have been proposed for reusing or altering existing motions. [Witkin95a] motion warping and [Bruderlin95a] motion displacement mapping discuss motion editing technique based on direct manipulation of data curves. [Bruderlin95a] and [Unuma95a] utilized signal-processing techniques for motion editing. [Wiley97a] proposed the interpolation synthesis algorithm that chooses and combines most relevant motions from the database to produce animation with a specific positional goal. Though some of the techniques above can be used for motion retargeting problem with user's extra efforts, they don't specifically address the motion-retargeting problem. [Boulic92a] presented the combined direct and inverse kinematics control technique for motion editing. The concept called coach-trainee metaphor is very similar to the motion retargeting problem formulation. A method, which is devoted to the motion-retargeting problem, was proposed by [Gleicher97a]. He used the space-time constraint method that minimizes an objective function $g(x)$ subject to the constraint of the form $f(x)=c$. Since the whole interval has to be integrated to find the optimal solution, the method is intrinsically an off-line process. [Choi00a] adopted the idea of inverse rate control to compute the changes in joint angles corresponding to those in end-effectors positions while imitating the captured joint angles by exploiting the kinematics redundancy.

When the virtual character and performer have different sizes and proportions, not all aspects of the motions can be preserved during mapping. At the lowest level, it is simply not possible to mimic both the locations of the end-effectors and the joint angles. A system must make choices to which aspects of the motion should be preserved and which should be allowed to change. Gleicher's space-time motion editing [Gleicher98a] and retargeting system [Gleicher97b] proposed the notion of preserving the important qualities of the motion by changing unimportant ones, where the important qualities were defined by constraints. Lee and Shin's hierarchical motion editing [Lee99a] provided similar results using a different underlying implementation. Popovic and Witkin demonstrated results that made the kinetic aspects of the original motion important to preserve [Popovic99a].

These methods mentioned above are all offline in that they examine the entire motion simultaneously in processing. Shin, etc [Shin01a] proposed an importance-based approach that retarget a performer to an animated character in real-time. They mapped as many of the important aspects of the motion to the target character as possible through importance analysis, while meeting the online, real-time demands. However, their approach addressed only the interaction between the end-effectors of a character and objects in the environment. In fact, there may also be interaction among the segments of a character, for example, sign language. In sign language we convey our meanings through the hand gesture, and small difference in the hand gesture may lead to wrong expressions. Due to the geometric difference between the character and the performer, the hand gesture of animation characters may deliver misleading meaning and even unrealistic motion through directly mapping. Existing algorithms don't deal with how to express precise meanings of the hand gesture. This paper presents a new retargeting method that can preserve accurate meanings of original motion for the upper limbs of the human body.

3. MODEL OF HUMAN UPPER LIMB AND REPROCESSING OF DATA

Our model has 18 joints and 32 degrees of freedom in each upper limb. The shoulder joint has three DOFs. The elbow and the wrist have two DOFs respectively. In order to perform complicated hand gestures there are three joints in each fingers and sum up to 25 DOFs in a hand (As shown in Figure 2).

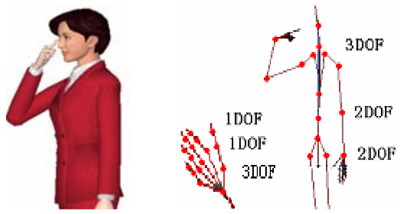

Figure 2. The controlled degrees of freedom for the dynamic model of the virtual human

We build our sign language motion library through data glove and location tracker device. Data glove is a kind of device for capturing the hand motion, and location tracker device can record the position of the shoulder, the elbow and the wrist. The captured data usually have many noises and cannot reflect the realistic motion in detail, so we must preprocess these captured data. A tool is developed to edit these original data through manual adapting, and finally we achieve a suit of standard sign language motion library for a fixed model proportional to the performer.

4. ANALYZE FEATURES OF HAND GESTURE

The most important features for the sign language are the precise position relations between the hand and other parts of the human body. The small change occurring in the end-effectors may lead to mistake in the meaning. This section discusses how to get the most important features from the hand gesture. In figure 1, the middle shows “eye” when the forefinger points to the eye. And if the forefinger points to the nose, the gesture means “nose”. In this example the end of the forefinger is very close to the eye, so it is important to express the meaning of the gesture. The relative positions between the end of forefinger and the eye may be selected as the most important information. Similarly, we may analyze the important information involving the relation between two hands (see the left and the right pictures in figure 1).

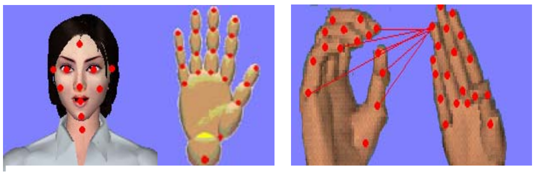

Figure 3. Feature points defined in the head and the hand

In order to find the feature information, we define three sets of special points called sensitive points for two hands and one head (see figure 3). Several notable points (for example eye, ear, nose, mouth, etc) in the head and two hands are selected as the characteristic points.

Usually, the importance is higher when the distance between two points is closer, and this feature must be preserved after retargeting. We call the first and the second closest points as key sensitive points and secondary key sensitive points defined as follows. Experiment shows that the main feature information can be got through selecting the key sensitive points and the secondary key sensitive points.

Here we only consider three parts including one head and two hands. We define three sets: ***H*** denotes sensitive points in the head; ***L*** and ***R*** denote the sensitive points in the left hand and the right hand respectively.

Let ***L***= $\{l_i, 0 \le i \le n\}$, ***R***= $\{r_i, 0 \le i \le n\}$, ***H***= $\{\mathrm{h}_i, 0 \le i \le m\}$. Here we define 22 points in each hand and 11 points in the head.

First we analyze the motion between two hands, and the motion between the head and the hand is similar. Given points r_i in ***R*** and l_j in ***L***, let $d_{ij}(t)$ be the Euclidean distance between them at t frame (We deal with each frame independently). Let ***D***= $\{(\mathrm{d}_{\mathrm{ij}}(\mathrm{t}), 0 \le i, j \le n\}$. We sort the ***D*** and get the two minimal values d_1 and d_2 ($d_1 \le d_2$). Suitable r_i and l_j corresponding to the d_1 and the d_2 are selected as key sensitive points and secondary sensitive points. For example, the left gesture in figure4, we select k_1 and k_2 as the key sensitive points, and select s_1 and s_2 as the secondary sensitive points. For the right gesture, there are only key sensitive points k_1 and k_2 (We consider the secondary sensitive points only if $d_2 \le 3$).

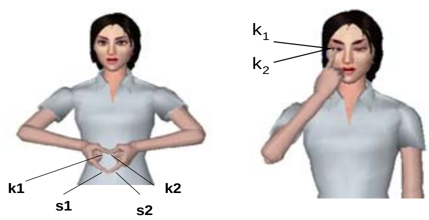

Figure 4. Analyze the key sensitive points and the secondary sensitive points: the left shows “heart” and the right means “eye”

5. MAP SENSITIVE POINTS

From above analysis, we know that the most important thing is the relative position between the hand and other parts of the body for understanding the sign language. If we map the relative position to target object when retargeting, these important aspects for understanding the meaning will be preserved. We have got several couples of important sensitive points through above analysis. Next, we will discuss how to map the relative position for each couple sensitive points.

Usually the relative position can be denoted with a vector linking to two sensitive points, and we must retain identical vector corresponding to the key sensitive points after retargeting.

Two kinds of situations need to be considered when we compute the relative positions of two sensitive points. The first condition is that one point is fixed and the other point moves to the target position, for example, motion occurring in the hand and the head. The other condition is that two points both move to the target positions, for example, motion occurring in two hands.

As shown in figure 5, here s_1 and s_2 are the sensitive points of the source object; t_1 and t_2 are the relative sensitive points of the target object without adapting. For first condition, we let s_2 fixed points, and then we transfer s_1s_2 to t_1t_2 until s_2 and t_2 are coincident. For the second condition, we transfer s_1s_2 to t_1t_2 until their centers are coincident. New t_1' and t_2' are the target position.

We denote it with two equations as followings.

$$t_1' = (s_1 - s_2) + t_2 \,,\; t_2' = t_2$$

Or

$$t_1' = \frac{t_1 + t_2}{2} + \frac{s_1 - s_2}{2} \,,\; t_2' = \frac{t_1 + t_2}{2} + \frac{s_2 - s_1}{2}$$

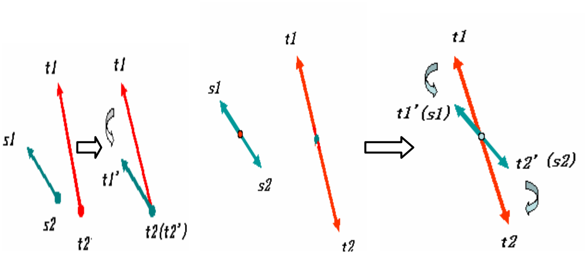

Figure 5. The left shows that one is fixed and the other move to the new position. The right shows that two points both move to the new positions

6. INVERSE KINEMATICS SOLVERS

Traditionally, inverse kinematics solvers can be divided into two categories: analytic and numerical solvers. Most industrial manipulators are designed to have analytic solutions for efficient and robust control. [Kahan83a] and [Paden86a] independently discussed methods to solve an inverse kinematics problem by reducing it into a series of simpler subproblems whose closed-form solutions are known. Korein and Badler [Korein82a] showed that the inverse kinematics problem of a human arm and leg allows an analytic solution. Actual solutions are derived by Tolani and Badler [Tolani96a]. A numerical method relies on an iterative process to obtain a solution. Girard and Maciejewski [Girard85a] addressed the locomotion of a legged figure using Jacobian matrix and its pseudo inverse. Koga [Koga94a] made use of results from neurophysiology to achieve an "experimentally" good initial guess and then employed a numerical procedure for fine-tuning. Zhao and Badler [zhao94a] formulated the inverse kinematics problem of a human figure as a constrained non-linear optimization problem. Rose et al. [Rose96a] extended this formulation to handle variation constraints that hold over an interval of motion frames.

Here we adopt the analytical method based on the geometrical constraints. According to the Stokoe's definition [Stokoe60a], each sign language can be broken into four parameters: hand shape, orientation, position and motion. These parameters as four important features play an important role in the sign language recognition. We build an objective function to satisfy these constraints (e.g. shape, orientation and position) for our special applications. Our method can find a suitable solution that maximizes the value of the objective function. Considering the IK chain displayed in the Figure 6 represents the human upper limb. This chain has three joints: the shoulder *S*, the elbow *E*, the wrist *W* and the end-effectors *F*. To guarantee orientation of the hand we select two points *N* and *M* in the hand together with the wrist joint *W* to define a plane in the hand, as shown in the Figure 6.

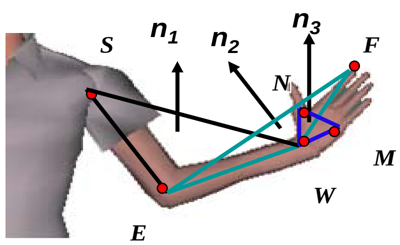

Figure 6. Joints chain of the upper limb and normal vectors n_1, n_2, n_3 defined by MNW, MEF, SEW

We give several constraints for the IK chain: the position of the key sensitive points (hard constraint), the normal vector $\boldsymbol{n_1}$ of the plane *MNW* determining the hand orientation, the normal vector $\boldsymbol{n_2}$ of the plane *EWF* and the normal vector $\boldsymbol{n_3}$ of the plane *SEW* determining the shape of the upper limb. In addition, each joint must meet the physiological constraints of the human body, so their activities should be restricted in a limited range. We give different weights for three orientation constraints according to their importance. The objective function is defined as in:

$$G = \alpha(n_1 \cdot n_1^{'}) + \beta(n_2 \cdot n_2^{'}) + \gamma(n_3 \cdot n_3^{'})$$

Here n_i and $n_i^{'}$ represent the original normal vectors and the target normal vectors respectively, and we specify α =0.8, β =0.15, γ =0.05.

There are two elbow circles O_1 and O_2 defined in [Tolani00a]. One is made up of the elbow, the shoulder and the end-effector, and the other is made up of the elbow, the wrist and the end-effector, as shown in the figure 7. Let their swivel angles are Φ and Ψ. When placing the end-effectors (F) at a desired point in space, there are an infinite number of solutions. When the swivel angles Ψ of the elbow circles O_1 is confirmed we can decide the position E, and hence we get another elbow circles O_2. Let its swivel angles is Φ, so we can get the position of W. Then we can get the length of EF and the DOF's values of the shoulder, the elbow and the wrist parameterized by Φ, Ψ. Furthermore we can get the reasonable range of Φ, Ψ and the length of EF according to the physiological constraints and the geometrical knowledge, and we can enhance the efficiency of our IK algorithm to a great extent. For example, let EF= d_2 , SF= d_1 , SE= l_1 , EW= l_2 , WF= h , we can deduce: $d_1 + l_1 \geq d_2 \geq d_1 - l_1$, $l_2 + h \geq d_2 \geq h - l_2$, and get the range of d_2 : $m \geq d_2 \geq n$. m=min { $d_1 + l_1$, $l_2 + h$ }, n=max { $d_1 - l_1$, $h - l_2$ }.

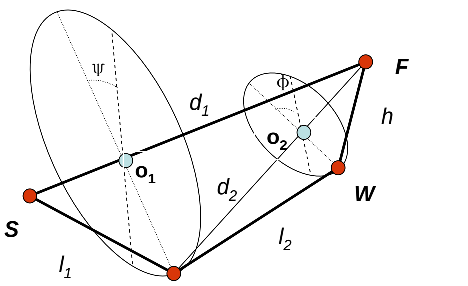

Figure 7. Two circles formed by the elbow and the wrist

Algorithm consists of two steps.

First we solve the DOF's values of the shoulder, the elbow, and the wrist according to the position of the key sensitive points. We preserve other DOF's values of joints in the hand if there is a lack of secondary sensitive points; otherwise we solve it again in a local joint chain in the hand formed by secondary sensitive points. It is described in the following steps.

(1)Get the valid ranges of Φ, Ψ, d_2.

(2)For each d_2 between m and n, we ascertain the elbow circle O_1

(3)For each Ψ we compute the position of E and get the elbow circle O_2

(4) For each Φ we compute the position of F and get the value of G

(5) Select suitable E and F values corresponding to the maximal G and gets the values of DOFs

As for secondary sensitive points, because we had ascertained the DOF's values of the shoulder, the elbow and the wrist, we can deal with it only in a finger chain.

7. EXPERIMENTAL RESULTS

Our method had been implemented on Chinese sign language synthesis system. It is a system using computer technology that translates text into animation of the virtual human in order to help hearing impaired people study sign language and communicate with the outsides conveniently. For pursuing more harmonious interaction between human and the machine and applying it for more fields (for example, TV news broadcasting, internet, communication and film) so as to improve their life's quality we build many virtual human models for the users' choice. Our task is to produce animations for different virtual human models that express the same meanings and can be readily understood by members of the deaf population.

Figure 8. Virtual human models: Joe, Yuxin, Jali, Lisa, Susan, Lili

Our systems have six virtual human models that one is the standard model and others are different from the standard in sizes and proportions (see figure 8). To evaluate the effectiveness of our work, two methods are adopted. First we conducted tests among deaf people for our retargeting results. In this experiment we selected 160 deaf people from four deaf schools and 100 typical examples of sign language including various kinds of contacting, crossing to test. Usually these sign languages will express wrong meanings if we don't retarget it. Test result is shown in the tabe1.

model	Lisa	Jali	YuXin	Susan	Lili
R.R	97.54%	98.33%	98.43%	97.35%	96.26%

Table 1. R.R means recognition rate

In addition, we invited some experts in sign language to examine all words for the five models. There are 3162 basic words in Chinese sign language. Test result is shown in the tabe2.

model	Lisa	Jali	Yu Xin	Susan	Lili
B.R.R	47.28%	65.03%	61.52%	52.43%	63.25%
A.R.R	96.54%	94.33%	96.43%	95.35%	95.26%

Table 2. B.R.R means recognition rate before retargeting, A.R.R means recognition rate after retargeting.

It is very effective to preserve original meanings and can be readily understood by deaf people after retargeting from the test result. Experimental results show that the proposed method dramatically improves the recognition rate about 30%. Our methods can produce animation for the Sign language in real-time. We show some results in figure 9. These snapshots show some key-frames for several typical words in sign language, and animation for retargeting results can be got from our application for Chinese Sign Language Synthesis System.

8. CONCLUSIONS AND FUTURE WORK

We have presented a new approach for motion retargeting that transforms the upper limbs motions of a performer to the virtual characters with different sizes and proportions. First we define many sensitive points on a human body and select key the sensitive points and the secondary sensitive points through analyzing the importance of the sensitive points. Then we propose a novel mapping method based on relative position that adapts the original sensitive points to the target sensitive points. Finally we utilize an IK solver to realize the retargeting problem. Our methods had been implemented on Chinese Sign Language Synthesis System.

Sign language, as a kind of most structured body language, is regarded as an indispensable means of everyday communication for deaf people. Research on sign language recognition will make for the communications between deaf people and common people. Conventional sign language recognition seldom utilizes the synthesis information of sign language. We can produce many suits of synthesis data for different models through our retargeting technique, and our future work is to implement the sign language recognition system based on synthesis information.

9. ACKNOWLEDGMENTS

This work has been supported by National Science Foundation of China (contract number 60303018), National Hi-Tech Development Program of China (contract number 2003AA114030), the Natural Science Foundation of Beijing of China (No.4011001), the Educational Committee of Beijing of China (No.01KJ-017, No. 2002KJ001)

.

10. REFERENCES

[Boulic92a] R. Boulic and D. Thalmann. Combined direct and inverse kinematic control for articulated figure motion editing. Computer Graphics Forum, 11(4): 189–202, 1992.

[Bruderlin95a] A. Bruderlin and L. Williams. Motion signal processing. InR. Cook, editor, Computer Graphics (SIGGR–APH '95 Proceedings), 97-104,August1995. ACM-0-89791-701-4.

[Choi00a] Kwang-Jin Choi and Hyeong-Seok Ko. On-line motion retargetting. Journal of Visualization and Computer Animation, 11:223-243, 2000.

[Gleicher98a] M. Gleicher. Retargeting motion to new characters. In SIGGRAPH 98 Conference Proceedings, Annual Conference Series, pages 33–42. ACM SIGGRAPH, Addison Wesley, July 1998. ISBN 0-89791-999-8.

[Gleicher97a] Michael Gleicher. Motion editing with spacetime constraints. In Proceedings of 1997 Symposium on Interactive 3D Graphics, 139-148, 1997.

[Girard85a] M. Girard and AA Maciejewski, "Computational modeling for the computer animation of legged figures," Computer Graphics, Vol. 19, No. 3, pp. 263-270, July 1985.

[Kahan83a] W. Kahan. Lectures on computational aspects of geometry. Unpublished manuscripts, 1983.

[Koga94a] Y. Koga, K. Kondo, J. Kuffer, and J. Latombe. Planning motions with intentions. Computer Graphics (Proceedings of SIGGRAPH 94), 28:395–408, July 1994.

[Korein82a] J. U. Korein and N. I. Badler. Techniques for g–enerating the goal-directed motion of articulated structures. IEEE CG&A, pages 71–81, Nov. 1982.

[Lee99a] Jehee Lee and Sung Yong Shin. A hierarchical approach to interactive motion editing for human likefigures. In Proceedings of SIGGRAPH 99, 39–48, 1999.

[Paden86a] B. Paden. Kinematics and Control Robot Manipulators. PhD thesis, University of California, Berkeley, 1986.

[Popovic99a] Zoran Popovic and Andrew Witkin. Physically based motion transformation. In Proceedings of SIGGRAPH 99, 11–20, 1999.

[Rose96a] C. Rose, B. Guenter, B. Bodenheimer, and M. F.Cohen. Efficient generation of motion transitions using spacetime constraints. Computer Graphics (Proceedings of SIGGRAPH 96), 30:147–154, August 1996

[Shin01a] Shin H. J, Lee. J, Gleicher, M, and Shin, S. Y. Computer Puppetry: An Importance-Based Approach. ACM Transactions on Graphics, Vol. 20, No. 2, April 2001, Pages 67-94.

[Stokoe60a] W. C. Stokoe, Sign Language Structure: An Outline of the Visual Communication System of the American Deaf. Studies in Linguistics: Occasional Papers 8 (Revised 1978). Buffalo, NY: Linstok, 1960.

[Tolani96a] D. Tolani and N. I. Badler. Real-time inverse kinematics of the human arm. Presence, 5(4): 393–401, 1996.

[Tolani00a] D. Tolani, A. Goswami, and N. Balder. Real-time inverse kinematics techniques for anthropomorphiclimbs. Graphical Models 62(5), Sept. 2000,335-388

[Unuma95a] K. A. Munetoshi Unuma and R. Takeuchi. Fourier principles for emotion-based human figure animation. In R. Cook, editor, Computer Graphics (SIGGRAPH ’ 95 Proceedings), 91–96, August 1995. ACM-0-89791-701-4.

[Witkin95a] A. Witkin and Z. Popovic. Motion warping. In R. Cook, editor, Computer Graphics (SIGGRAPH ’95 Proceedings), pages 105–108, August 1995. ACM-089791-701-4.

[Wiley97a] D. J. Wiley and J. K. Hahn. Interpolation synthesis of articulated figure motion. IEEE Computer Graphics and Applications, 39–45, November/December 1997.

[Zhao94a] J. Zhao and N. I. Badler. Inverse kinematics positioning using nonlinear programming for highly articulated figures. ACM Transactions on Graphics, 13(4): 3–13–336, 1994.

Figure 9. Some snapshots of typical sign language: The left shows standard sign language. The middle shows results without retargeting and the right shows results after retargeting. (They represent the meanings of “head”, “tongue”, “eye”, “fight”, “hesitate”, “bless” and “human” respectively.)