

A Human Action Recognition Scheme in Temporal Spatial Data for Intelligent Web Browser

Kyungeun Cho[†]

ABSTRACT

This paper proposes a human action recognition scheme for Intelligent Web Browser. Based on the principle that a human action can be defined as a combination of multiple articulation movements, the inference of stochastic grammars is applied to recognize each action. Human actions in 3 dimensional (3D) world coordinate are measured, quantized and made into two sets of 4-chain-code for xy and zy projection planes, consequently they are appropriate for applying the stochastic grammar inference method. We confirm this method by experiments, that various physical actions can be classified correctly against a set of real world 3D temporal data. The result revealed a comparatively successful achievement of 93.8% recognition rate through the experiments of 8 movements of human head and 84.9% recognition rate of 60 movements of human upper body. We expect that this scheme can be used for human-machine interaction commands in a web browser.

Keywords: Temporal Multi-Stream Data Processing, Multimedia Data Processing, Motion Data Processing

1. INTRODUCTION

Internet-based services are increasingly provided to users in these days. Many applications of Web3D technologies such as Virtual Reality (VR) panorama are widespread. During all the years of user interface evolution, the user environment that they presented has remained very unfamiliar to novice. These trends induce to begin building novel user interfaces with virtual user interfaces for web browsing systems, VR navigation on Internet or controlling a hands-free web browser for physically-handicapped person. An intelligent web browser tracks human head and upper body motion in real time. For example, as a person sits in front of a computer with webcam, his upper body motion controls the navigation in a VR Web browser[1,2].

This paper presents a recognition scheme to an-

※ Corresponding Author : Kyungeun Cho, Address : (100-715) Pildong 3 ga 26, Chunggu, Seoul, Korca, TEL : +82-2-2260-3834, FAX : +82-2-2265-8742

E-mail : ckc@dongguk.edu

Receipt date : May. 6, 2005, Approval date : May. 24, 2005

[†] Dept. of Computer and Multimedia Engineering Dongguk University.

alyze human actions on 3D temporal data of video where an unsupervised inference procedure is introduced to stochastic grammars to apply for a user friendly web based interfaces. This scheme is based on the principle that human body actions are defined as a combination of multiple articulation movements which is built up from multiple mutually-synchronized temporal data. Human actions are considered as a stochastically predictable sequence of states. Therefore, we apply a mixed statistical-syntactic approach to recognition of action. In addition, we use a mechanism to infer stochastic grammars, which deals with learning the production probabilities for a set of given grammars, each of which represents single human action. Thus, the propriety of the recognition has been confirmed by showing that various physical movements as the real world 3D temporal data can be classified correctly through the stochastic regular grammar inference method. In our work, we use 60 types of head and upperbody movements for recognition.

The remainder of this paper is organized as follows: Section 2 summarizes the related work. Section 3 describes the general stochastic grammar inference

method. Section 4 presents an overview of the recognition system architecture, the preprocessing phase, and the stochastic regular grammar inference method applied to human action recognition. Section 5 presents experimental results. Section 6 concludes the paper with a brief outline of the future works.

2. RELATED WORK

2.1 Human-machine Interfaces Using Action Recognition

The development of a system responsible to human action requires new paradigm for user-interface design and new recognition strategies for motion understanding. Human action recognition plays one of the important roles to smart human-machine interaction, and is one of the major goals of the next computing evolution. Such approaches improve the quality of human-machine interaction by making interfaces more user-friendly. They assist a person in his work by diminishing his effort to communicate with machines. Efficient and easy human-machine interfaces using action recognition are already established. Digital Office[3] which has a user interface for a whiteboard scanner recognizes human gestures. Perceptual Browser[3] tracks human head motion in real time. Virtual Environment (VE) navigation system[4], and Object Selection System[5] recognize hand gestures for the control of window based user interface. Robot controlling system TOAST[6], and Body-based interfaces[7] are other examples.

2.2 Human Action Recognition

There has been a tremendous growth in the amount of researches on computer vision aimed at understanding human action. Medical analysis of human gait movement, nonverbal communication in social psychology, VR using avatar control, automatic man-machine interaction, development of surveillance systems, sign-language recognizer,

choreographic analysis of dance and ballet, and gymnastic movement—all of them belongs to the application area of automatic action recognition. In some areas, action recognition systems are already established such as the Chinese Sensei system[8] analyzing Tai-Chai recognizing and translating sign language. Especially, there is a growing tendency nowadays to conduct the automation research of physical gestures and hand-movements that convey most of human communications[9]. If physical movements can be recognized automatically, the efforts that people makes to analyze manually can be automated and used effectively as convenient communication media between computer and human.

Practically, there are manifold phases in recognizing actions including tracking of physical actions, separation of bodies from the background, identification of body parts, and recognition of action patterns. Physical movements are composed of various kinds of bodily actions that constitute human bodies such as hands, lower-arms, shoulders, head and so on. In other words, we analogize a pass that moves of each body part in the 3D space to express human actions after combining them[10]. There are various approaches to the recognition of physical motions: Dynamic Time Warping (DTW) [11], syntactic method[12], template matching method [13], fuzzy method[14], Hidden Markov Model (HMM)[12].

In a previous research, Stochastic Context Free Grammar was used to the recognition of physical actions[12]. The system consists of an HMM bank and a probabilistic Earley-based parser. This caused them to make the grammar inference difficult in the framework of stochastic parsing. In this paper, we will show that a stochastic regular grammar inference method can resolve this problem because it has no much limit in inferring the grammar.

3. INFERENCE OF STOCHASTIC GRAMMAR

In this section, we apply an inference of stochas-

tic grammar scheme to recognize human actions. We describe concisely the appropriateness of our method for motion analysis, data representation method for motion sequences and finally stochastic grammar inference method.

3.1 Syntactic Pattern Recognition for Motion Analysis

Syntactic pattern recognition is a method focused on the structure of pattern. This method dismantles objective pattern into simple sub-patterns to recognize and explain relations that exist among the sub-patterns as a grammar theory of formal language. In general, when the patterns to be considered can determine each grammar, they can be recognized correctly.

However, a subpattern can more frequently appear than other subpatterns, and in some occasions it can include severe transformations by errors. It is possible to limit the number of pattern class, and each class is represented as a distinctive grammar.

But there are occasions when some errors or languages generated by them may not select a grammar class. If we can reflect probability value approaching to any class, the method that probability applies to grammar can overcome ambiguous classification problems. If physical movements include transformations, noises, and incomplete feature extractions, it is very difficult to express all physical movements in simple grammar although they have the same meaning. Thus, stochastic grammar can reflect these features most effectively[15].

Previous researches have been conducted to recognize all the actions directly after the composition into some patterns and to produce each grammar class. However, they need intricately huge labors to extract every recognition object and to grammaticize distinctive ingredients artificially. Even if the same movements, it also has some difficulties because the pattern sequence can be expressed so differently. Thanks to a preservation technique, however, computers can construct stand-

ard patterns and accumulate knowledge automatically. A research has been actively performed on how an unknown pattern is recognized through accumulated knowledge. Syntactic recognition through grammatical inference is applicable to this case.

A method that satisfies the grammatical inferring and the stochastic grammar is called the stochastic inference method. Stochastic grammar inference has already been implemented in normal or abnormal chromosome recognition study[15].

3.2 Data Representation for Human Motion Sequence

To generate input data for recognizer, we used STABIL++ system[16] which detects and tracks color markers and outputs 3D world-coordinates of each articulations. Fig. 1 shows an example of the video data sequence. We obtain 3D data of 11 color marks on each articulation, trunk, and head as in Fig. 2. We call each color marks as a node. As three marks of neck, breast, and belly on the body have only a little position movement, they do not have much influence on the change of actions like the marks attached on each articulation.

Practical body composition data are the pair of x, y, z position (x,y,z) . (x,y,z) s to the articulation are used after their quantization at intervals of 3 cm. After converting quantized data into projec-

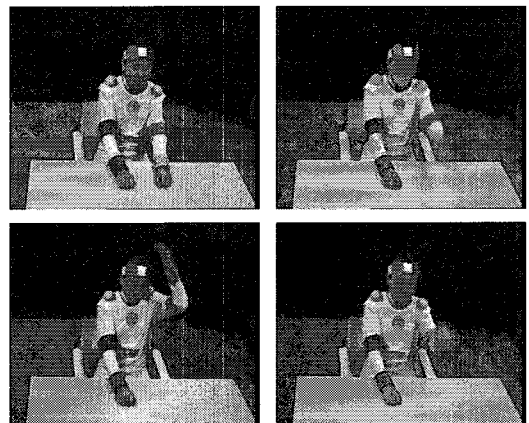


Fig. 1. Sequence of video data.

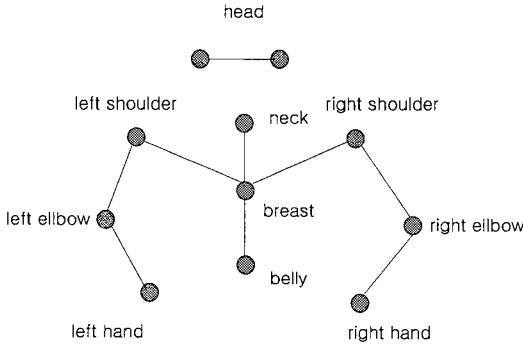


Fig. 2. Position of color marks for upper body.

tion value in each xy, zy plane, we represent them in 4 directional codes as a sub-pattern sequence. A sub-pattern sequence such as (left 12 up 11 right 4 up 5 right 9 down 13 left 2 down 1 left 5 down 2 left 3 up 13 left 1 up 1), shows an example of the left hand when the left arm is turned to the front and thereafter they are projected to xy plane and zy plane. The format of subpattern sequence will be indicated as [code code_count]*. These codes mean directional codes. The number of the codes refers to the frequency of the codes.

3.3 Theoretical Foundation for Stochastic Grammar Inference

An act to seek the probability value of each production rule with given grammar and learning patterns is called the inference of stochastic grammars. The given learning patterns are composed of patterns that produce different grammar respectively and probability value of each grammar. This section explains the method of estimating the production probabilities.

Consider a M -class problem characterized by the stochastic grammars $G_{sk} = (N_k, \Sigma_k, P_k, D_k, S_k)$ for $k=1, 2, \dots, M$, where N_k is a finite set of non-terminals, Σ_k is a finite set of terminals, P_k is a finite set of productions, D_k is a set of probability values of the production regulation to be assumed, and S_k means the starting symbol. To estimate the production probability set D_k , the probability P_{kij} associated with the production $A_i \rightarrow \beta_j$ in grammar

G_{sk} must be obtained for each learning pattern set X . It can be approximated by the relation ;

$$\text{estimated } P_{kij} = \hat{P}_{kij} = \frac{n_{kij}}{\sum_r n_{kir}} \tag{1}$$

where n_{kij} means the total average number of times when a production rule of $A_i \rightarrow \beta_j$ in grammar G_{sk} is used to all the learning patterns, and can be obtained by the following equation :

$$n_{ij} = \sum_{x_h \in X} n(x_h) p(G_{sk} / x_h) N_{kij}(x_h) \tag{2}$$

In the equation (2): $n(x_h)$ is the frequency of all patterns occurred in the learning pattern set, $N_{kij}(x_h)$ is the number of times that $A_i \rightarrow \beta_j$, a production regulation of grammar G_{sk} is used when a pattern of x_h is parsed. Although the probabilities of the grammars corresponding to the production rules are not known, the parsing steps are proceeded because the production rules are known. $p(G_{sk}/x_h)$ means the probability with which a pattern of x_h is produced from the grammar G_{sk} . This probability has to be produced during the step when each string is trained. $\sum_r n_{kir}$ in the equation (1) is the summation in the case of $A_i \rightarrow \beta_j$. The summation is obtained by the above-mentioned method for all the production rules that have the same left part nonterminal A_i in grammar G_{sk} . If the set of D_k is obtained, the learning step, i.e., the inferring step of the stochastic grammar is completed, and the test step can be performed. For each input test pattern we seek the probability corresponding to all the grammatical classes. The input pattern is classified into the class with the highest probability value[15].

4. IMPLEMENTATION OF ACTION RECOGNIZER

Action recognizer is largely composed of a pre-processor, a parser, and a stochastic grammar inference processor. The preprocessor quantizes measured actions in 3D world-coordinates and makes

two sequences of 4 directions code for xy and zy projection planes. The stochastic grammar inference processor takes learning data and produces 16 probability tables by using the regular grammar. When test data are given, the parser seeks a class with the highest stochastic value through the related stochastic information of the table. Fig. 3 shows our overall recognition system architecture. The motion ranges for the calculation of probability value can be reasonably defined and limited to the same level motions. This processing step is performed by the AMNC (Approximate Movement Node Classifier).

In this section, we explain a recognition method of human action which uses the inference of stochastic grammars. In section 4.1, the preprocessing steps to the input data for the stochastic grammar are presented. Section 4.2 describes the regular grammar generation method of the stochastic regular grammar inference system and the stochastic table production for the recognition.

4.1 Preprocessing Step for Learning and Testing Data

In order to generate input data of recognizer, we

use STABIL++ system [16] to detect and track color markers and output 3D world-coordinates of each articulations. The coordinates of (x,y,z) to the articulation as a component of the physical body are used after their quantization with the interval of 3 centimeters. They are quantized because very small motion shall be disregarded.

4.1.1 Plane Projection of Quantized Data and 4 Directional Coding

After converting quantized data into projection values in each xy, zy plane, we code them in 8 directional codes. The meaning can be different depending on the projected planes. That is, the 8 directional coding of the projected data to xy plane has the meaning of left, left-up, up, right-up, right, right-down, down in turn. This has the same meaning as the direct observance of physical motions. In fact, the subpattern sequences divide all the codes as the result of the 8 directional coding into two codes. 4 directional coding is originated from the simplification of the grammar because too many production rules may create a dangerous result of decreasing the recognition rate because of the high frequency of the transfers to the next conditions.

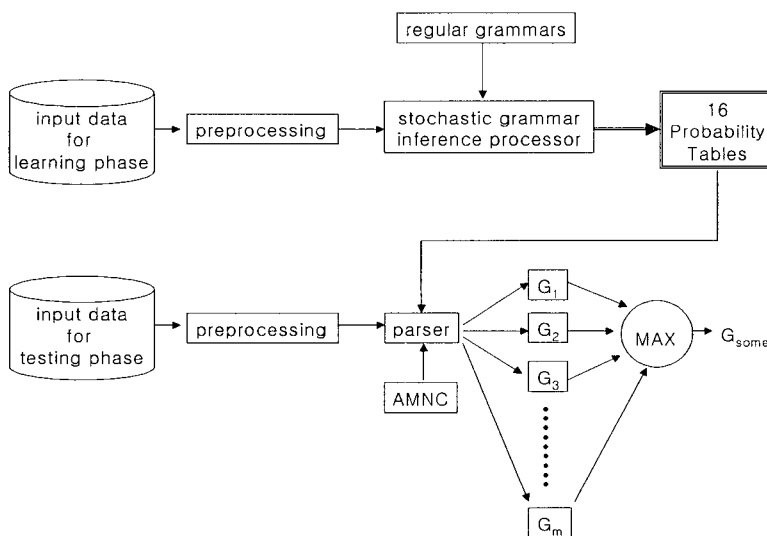


Fig. 3. The overall recognition system architecture.

An example of the subpattern sequence is shown in Table 1.

$$N = \{S, B, F, U, D\}$$

$$\Sigma = \{\text{backward, forward, up, down}\}$$

4.2 Stochastic Regular Grammar Inference Processor

In this subsection, we describe the implementation of the recognition processor in detail.

4.2.1 Specification of Stochastic Regular Grammar

When test data are given to the parser, it seeks a class with the highest stochastic value by referring to the related stochastic information in the probability tables. The input data processed in the preprocessing step shall be applied to the recognizer. An example of the state transition diagram of the finite automata and the regular grammar are shown in Fig. 4(a) and Fig. 4(b).

We estimate a stochastic value for each production rule of the regular grammar. Thus, M grammars that classify M human physical motions are expressed as follows. This is an example of the grammar for the learning pattern projected to xy plane.

$$G_{s1} = (N, \Sigma, P, D_1, S)$$

$$G_{s2} = (N, \Sigma, P, D_2, S)$$

...

$$G_{sM} = (N, \Sigma, P, D_M, S)$$

$$N = \{S, R, L, U, D\}$$

$$\Sigma = \{\text{right, left, up, down}\}$$

Only the following items are changed in the grammar of the learning pattern set projected on the zy plane.

4.2.2 Generation of Probability Tables

The probability value for each production rule is obtained as shown in Table 2. The production rules are replaced with the Confrontation Rules defined to follow the method of P_{kij} . Here, the subscript "k" means the position of the grammar class, the subscript "i" means the subscript on the left hand side, and the subscript "j" means the subscript on the right hand side. Some examples of the probability value estimated practically are shown in the column D_{44} of Table 2. D_{44} means the probability value of the production rules estimated after the learning the motions of the turning left arm. This example shows the estimated probability value to the learning patterns obtained through the projection on the xy plane of the motions of left hands when the left arm is turned.

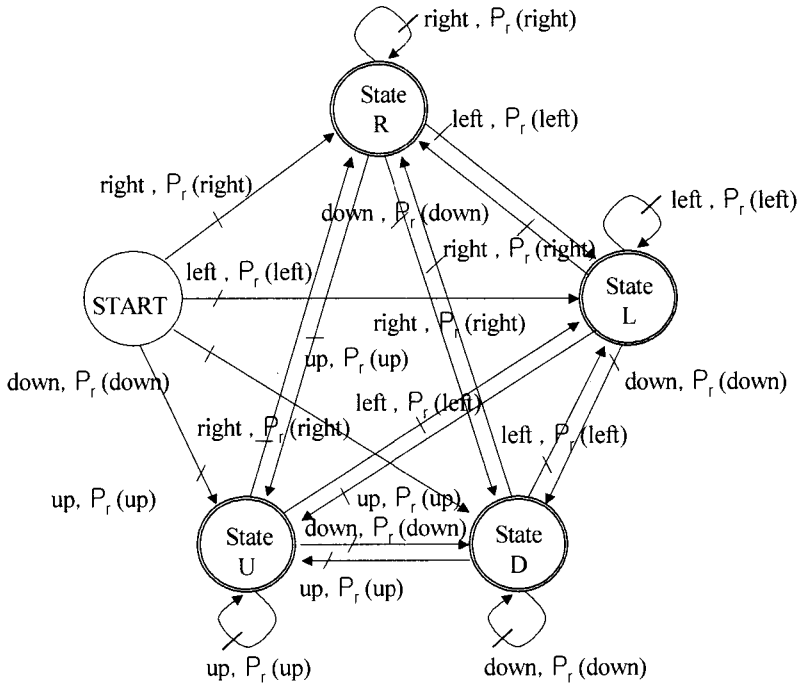
The inference steps of stochastic regular grammar are applied to the nodes used for the motion recognition and the projection on the xy and zy plane. Here, node means two positions of head and articulation points. We use 8 points : two head positions, two shoulder positions, two elbow positions and two ankle positions. Therefore, the probability tables derived from the inference result of the stochastic regular grammar, can be produced up to 16 types.

4.2.3 Testing Phase

Classification of any action pattern into the cor-

Table 1. An input example of stochastic regular grammar inference system : subpattern sequences from movement of the left hand with the left arm turning forward

node = left hand plane=xy	left 12 up 11 right 4 up 5 right 9 down 13 left 2 down 1 left 5 down 2 left 3 up 13 left 1 up 1 #
node=left hand plane=zy	backward 23 up 16 forward 23 down 14 backward 6 down 2 backward 8 up 7 backward 2 up 1 backward 6 down 2 backward 15 up 3 backward 2 up 5 backward 2 forward 2 down 2 forward 4 down 10 backward 24 up 2 backward 2 up 13 forward 3 up 2 forward 22 down 12 forward 2 down 5 #



(a) State diagram of finite automata M_{xy}

$G_{xy} = \{S, R, L, U, D\}, \{\text{right, left, up, down}\}, P, S\}$ with productions

- | | |
|-------------------------------------|-------------------------------------|
| Start \rightarrow right R_state | U_state \rightarrow right R_state |
| Start \rightarrow left L_state | U_state \rightarrow left L_state |
| Start \rightarrow up U_state | U_state \rightarrow up U_state |
| Start \rightarrow down D_state | U_state \rightarrow down D_state |
| R_state \rightarrow right R_state | U_state \rightarrow right |
| R_state \rightarrow left L_state | U_state \rightarrow left |
| R_state \rightarrow up U_state | U_state \rightarrow up |
| R_state \rightarrow down D_state | U_state \rightarrow down |
| R_state \rightarrow right | D_state \rightarrow right R_state |
| R_state \rightarrow left | D_state \rightarrow left L_state |
| R_state \rightarrow up | D_state \rightarrow up U_state |
| R_state \rightarrow down | D_state \rightarrow down D_state |
| L_state \rightarrow right R_state | D_state \rightarrow right |
| L_state \rightarrow left L_state | D_state \rightarrow left |
| L_state \rightarrow up U_state | D_state \rightarrow up |
| L_state \rightarrow down D_state | D_state \rightarrow down |
| L_state \rightarrow right | |
| L_state \rightarrow left | |
| L_state \rightarrow up | |
| L_state \rightarrow down | |

(b) Derived regular grammars G_{xy}

Fig. 4. State diagram of finite automata M_{xy} and derived regular grammars G_{xy} .

Table 2. Probability corresponding to Productions

	Productions	Replaced Productions	D_1	D_2		D_{44}		D_m
1	S → right R	A1 → β1	P_{111}	P_{211}		0.12217		P_{m11}
2	S → left L	A1 → β2	P_{112}	P_{212}		0.56045		P_{m12}
3	S → up U	A1 → β3	P_{113}	P_{213}		0.30227		P_{m13}
4	S → down D	A1 → β4	P_{114}	P_{214}		0.01511		P_{m14}
5	R → right R	A2 → β1	P_{121}	P_{221}		0.69052		P_{m21}
6	R → left L	A2 → β2	P_{122}	P_{222}		0.03151		P_{m22}
...					
13	L → right R	A3 → β1	P_{131}	P_{231}	...	0.00829	...	P_{m31}
14	L → left L	A3 → β2	P_{132}	P_{232}		0.74862		P_{m32}
...					
21	U → right R	A4 → β1	P_{141}	P_{241}		0.07366		P_{m41}
...					
29	D → right R	A5 → β1	P_{151}	P_{251}		0.00736		P_{m51}
30	D → left L	A5 → β2	P_{152}	P_{252}		0.05008		P_{m52}
...					
[Confrontation Rules]								
A1 = S			β1 = right R		β6 = left			
A2 = R			β2 = left L		β7 = up			
A3 = L			β3 = up U		β8 = down			
A4 = U			β4 = down D					
A5 = D			β5 = right					

rect action class is done in the testing phase. An arbitrary action is classified into the class where the matching probability is the highest. When test data are given, a subpattern sequence is obtained for each node and projection plane after preprocessing. The parser takes these subpattern sequences as input, and inspect the moved nodes of the test patterns by using the node classifier estimating the similar motions. The final probability value of each motion cannot be compared reasonably if two nodes are moved in an arbitrary action and if eight nodes are moved in another arbitrary action. So, we do not obtain the maximum value among the probability values of all the motions, but obtain the maximum value among the finally calculated probability values within only the set of the similar motions (AMNC).

Through the judgment of similar movement nodes, only the data that can be available as the learn-

ing patterns with reasonable critical (threshold) value are selected to be the candidate set. This is possible after comparing the average motional differences between the moved nodes of the present test input and the moved nodes of each class learned on the training table. In our experiments, we consider the critical (threshold) value as 1. A node with the motional difference of 1 is a candidate to be compared for the reasonable working result.

5. EXPERIMENTAL RESULT AND ANALYSIS

In this chapter, we explain some actions for experiment. We show the correct recognition rate, analysis the incorrectness of data and describe the applicable method of our research.

5.1 Experiment Data

Physical movements of human upper body are

recorded in the system of STABIL++ for experiments. 60 movements gestured by three persons are recorded. The next are some movements chosen from them.

- [head] hang down head and to the former place
- [head] raise head upward
- [head] turn head to the right and to the former place
- ...
- [body] Seeing backward to the right with shoulder and to the former place
- [body] bend body to the left and to the former place
- ...
- [right arm] contacting head with right hand
- [right arm] turn right arm backward
- ...
- [left arm] contact head with left hand
- [left arm] contact body (stomach) with left hand
- ...
- [both arms] raise and lower shoulders upward and down a few times
- [both arms] turn both shoulders forward (moving arms slightly)
- ...
- [compound] shaking head and folding arms at the same time
- [compound] collecting hands on head leaning body backward
- ...

Practically, the data for recognition to analyze movements are composed of the movements of head, bodies, arms, and of other compound motions. These kinds of movements are selected with the reference to the data of movement for human motion analysis studies[17]. Input data for the experiments are 900 in all, of which 500 odd data are used for learning, and 400 odd data were used for testing.

5.2 Experiment Result

Table 3 shows the results that have been obtained from the experiments with the recognizer. Experiment 1 is the result obtained from the head motion recognition. Experiment 2 is the result obtained from 60 Head and body movements before using AMNC, and experiment 3 is the result obtained from 60 Head and body movements after using AMNC.

The experimental result shows the recognition rate of 93.8% recognition rate through the experiments of 8 movements of human head without using the AMNC. For head movement recognition, the AMNC procedure is not used, because head movements have all the same movement degree. It means that they are already predefined as a similarity movement group. The experimental result through the experiments of 60 movements of human upper-body showed the recognition rate of 64.6% before the AMNC was used, and 84.9% after it was used.

5.3 Error Analysis

Table 4 shows some incorrect results of head movement and one arm movement. We understand that minute difference of the stochastic value may cause the incorrect recognition. The data analysis result caused by the minute stochastic value, can be classified into two parts. The one may be caused from the accuracy of the movement and we may have difficulty in recognizing the case. The test pattern 6 in Table 4 corresponds to this case. It seems that the case cannot be improved practically. Considering the data corresponding to an-

Table 3. Experimental results

	Experimental Contents	Correct Recognition number / Total number = correct recognition rate
Exper. 1	8 Head movement	60/64 = 93.8 %
Exper. 2	60 upper body movements before using AMNC	265/410=64.6 %
Exper. 3	60 upper body movements after using AMNC	348/410=84.9 %

other case, the incorrect recognition of them seems to be improved compared with the size of movements of each hand. It can be added to the post step of the treatment of the stochastic inference method. The error data corresponding to the case will be the test pattern 29 in Table 4.

5.4 Definition of Gestures for User Interface Commands

We explain a case how head movements can control the scrolling of text in a web browser. Head actions are defined as control commands of a web application. First, the Action recognizer gets the motion data and analyzes its action. Then, the command interpreter from the application receives

the recognized gesture, and executes the appropriate command. Table 5 shows an example of commands which can be used to interact with a browser.

6. CONCLUSIONS

In this paper a recognition model for upper body motions by using stochastic grammatical inferring method that processes 3D temporal data in web applications was presented. 3D sequence data of human motions are encoded into 4 directional codes, and projected to xy and zy plane. These sequences are processed as the input of the stochastic recognizer. They are used in the learning phase for building

Table 4. Example of the case when a movement is incorrectly found

Movement	Test Pattern	Candidates
Head Motions	6 : Turn to the left and then to the former place	4: Turn to the left. And then to the former place : 9.926396e+305
		6: Bend to the left. And then to the former place : 7.489708e+305
		7: Revolve to the right : 1.249047e+305
		2: Raise backward. And then to the former place : 9.928185e+304
		1: Hang down. And then to the former place : 6.681826e+304
		8: Revolving to the left : 3.573108e+304
		3: Turn to the right. And then to the former place : 1.918542e+304
		5: Bend to the right. And then to the former place : 7.918658e+303
One Arm Motions	29 : Back the right arm upward and down	25: Turn in the front of the right shoulder (Moving the arm slightly) : 1.035340e+234
		29: Raise and lower backward the right arm : 7.030614e+231
		30: Turn the right arm to the front : 1.224607e+230
		28: Raise and lower the right arm to the side : 3.255068e+225
		27: Raise and lower the right arm to the front : 3.80903e+224
		26: Turn the right shoulder backward (Move the arm slightly) : 4.592945e+223
		31: Turn the right arm backward : 9.315756e+222
		9: Turn the right arm backward with the shoulder. And then to the former place. : 5.482774e+220
58: Touch the head with the right hand shaking : 1.580669e+217		

Table 5. Example of human-machine interaction command

Action	User Interface Commands
[head] hang down head and to the former place	Scroll window down
[head] raise head upward	Scroll window up
[head] turn head to the right and to the former place	Scroll window right
[head] turn head to the left and to the former place	Scroll window left

the probability tables. These tables are used in the recognition phase for classifying each action into the right class of human action.

Some contributions of this paper are summarized as follows. First off, to understand human actions with stochastic grammars, we devise an inference method, which has been left unexplored and referred to as the further study. As a result of our approach, autonomous learning from predefined human action patterns has been revealed as possible. Secondly, our proposed scheme is suitable not only for simple actions composed of single articulation movement, but also for complex actions composed of several articulation movements.

In our experiments, 93.8% recognition rate of 8 movements of human head and 84.9% recognition rate of 60 movements of human upper body were achieved. These actions are expected to be utilized as human-machine interaction commands for a web browser.

7. REFERENCES

- [1] C. Bouras, A. Konidaris, and A. Sevasti, "The Transition from Contemporary to Virtual User Interfaces for Web-Based Services," *Computer Science Papers NEC Research Institute CiteSeer*, 1999.
- [2] C. Bouras, A. Konidaris, and A. Sevasti, "Virtual Environments in Educational Networks," *Computer Science Papers NEC Research Institute CiteSeer*, 1999.
- [3] M. Black, F. Berard, A. Jepson, W. Newman, E. Saund, G. Socher, and M. Taylor, "The Digital Office: Overview," *AAAI Spring Symposium on Intelligent Environments*, Stanford, March, 1998.
- [4] J. Chung, "Recognition for Realistic Interaction in the Virtual Environment," *VRRC(Virtual Reality Research Center)*, Korea, TR 2002-53
- [5] P. Rico, R. Kjeldsen, and J. Kender, "Interaction with On-Screen Objects using Visual Gesture Recognition," *Int'l workshop of CVPR'97*, 1997.
- [6] D. Verna, "Action Recognition: How Intelligent Virtual Environments Can Ease Human-Machine Interaction," *VSMM(Virtual Systems and Multimedia)' 2000, Intelligent Environments Workshop*, pp. 703-713, 2000.
- [7] C. Cho, H. Yang, G.J. Kim, and S. Han, "Body-based Interfaces," *VRRC(Virtual Reality Research Center)*, Korea, TR 2002-62.
- [8] D.A. Becker, "Sensei: A Real-Time Recognition, Feedback and Training System for T'ai Chi Gestures," *MIT Media Lab Perceptual Computing Group TR 426*, 1997.
- [9] J.M. Carroll, *Human-Computer Interaction in the New Millennium*, Addison Wesley, 2002.
- [10] S. Frey, *Das Berner System zur Untersuchung nonverbaler Interaktion, Methoden der Analyse von Face-to-Face-Situationen*, Stuttgart, Metzler, pp. 203-236, 1981.
- [11] D.M. Gavrila and L.S. Davis, "3-D Model-based recognition of human movement by dynamic time warping," *Computer Vision Lab.*, Univ. of Maryland, 1995.
- [12] Y.A. Inanov, "Application of stochastic grammars to understanding action," *MIT thesis*, 1998.
- [13] A. Sutherland, "Real-time Video-based Recognition of Sign Language Gestures using Guided Template Matching," *Gesture Workshop*, pp. 31-38, 1996.
- [14] M. B. Waldron and S. W. Kim, "Isolated ASL Sign Recognition System for Deaf Persons," *IEEE Trans. On Rehabilitation Eng.*, Vol. 3, No. 3, Sep., 1995.
- [15] R.C. Gonzalez and M.G. Thomason, *Syntactic Pattern Recognition*, Addison-Wesley Publishing Company, pp. 177-270, 1978.
- [16] O. Munkelt et al., "A model driven 3D image interpretation system applied to person detection in video images," *14th ICPR 98*, pp. 70-73
- [17] P.E. Bull, *Posture and Gesture*, Pergamon Press, pp. 163-187, 1990.



Cho, Kyungeun

1989. 3~1993. 2 Computer Science, Dongguk University (BS)

1993. 3~1995. 2 Computer Engineering, Dongguk University (MS)

1995. 3~2001. 8 Computer Engineering, Dongguk University (Ph.D)

2002. 3~2003. 2 Dept. of Digital Media, Fulltime Lecturer, Anyang University

2003. 3~2003. 9 Dept. of Game Engineering, Fulltime Lecturer, Youngsan University

2003. 9~Present Dept. of Computer Multimedia Engineering, Fulltime Lecturer, Dongguk University

Research Interest : Computer Game Algorithm, Game AI, Computer Graphics, Computer Vision