

# Long-Term Evaluation of a Modular Gesture Interface at Home for Persons with Severe Motor Dysfunction

Ikushi Yoda<sup>1</sup>(✉), Kazuhiko Ito<sup>2</sup>, and Tsuyoshi Nakayama<sup>2</sup>

<sup>1</sup> National Institute of Advanced Industrial Science and Technology (AIST),  
Tsukuba, Japan

`i-yoda@aist.go.jp`

<sup>2</sup> Research Institute, National Rehabilitation Center for Persons  
with Disabilities (NRCD), Tsukuba, Japan

`{itoh-kazuyuki-0923,nakayama-tsuyoshi}@rehab.go.jp`

**Abstract.** We have researched gesture interfaces for persons with motor dysfunction who could not use normal interface switches. The users have cerebral palsy, quadriplegia, or traumatic brain injury, and they have involuntary movement, spasticity, and so on. The purpose is to provide these users with easy and low-priced interfaces for operating PCs. To develop a low-priced interface, we used a commercially available image range sensor and developed a non-contact and non-constraint interface. For this purpose, we have collected various gestures of many persons with motor dysfunction and classified them into three types of users. We have developed seven types of gesture interfaces for users and applied three different modules to three subjects. We describe the basic structure of the gesture interfaces and their long-term evaluations.

**Keywords:** Gesture interface · Gesture recognition · Alternative input device · Persons with motor dysfunction

## 1 Introduction

Some individuals with severe motor dysfunction are unable to use existing computer interfaces due to spasticity, involuntary movements, limited range of arm motion, or diminished muscle strength. The interfaces available to these individuals, if any, are limited to customized switch interfaces, which makes it impossible for them to operate a computer with any degree of ease [9]. In their daily life, they are reliant on their caregivers to perform all but the most basic operations with household appliances such as televisions and air conditioners. Providing an environment in which these individuals can operate a personal computer with ease and use home appliances without restriction is essential for allowing them to live fulfilling and rewarding lives. Yet for disabled individuals who are only able to use simple switch-based devices, these more sophisticated operations are all but impossible. It is also very expensive to develop an interface capable of responding to changes in the individual's movements caused by physical deterioration as a result of disease progression and aging.

In the present study, we conducted research and development on a gesture interface to enable simpler operation of computers and home appliances among the many individuals with motor dysfunction who have difficulty operating a standard keyboard and mouse. Specifically, we developed a contactless, non-restrictive interface using a commercially-available imaging range sensor with the aim of making it affordable to all users.

The most important factor in developing this interface was that it had to involve technology that could easily be customized to a diverse range of individual users at low cost. To this end, we collected data on the movements of various disabled individuals, classified each of these movements according to body part, and developed a modularized gesture recognition engine [1]. We then utilized the results to start a basic, long-term experiment. In this paper, we describe the framework, methods, and specific conditions of this long-term experiment.

The authors have previously developed an interface based on head gestures for individuals with severe cerebral palsy who are unable to operate a motorized wheelchair, as part of a project to assist severely disabled persons [8]. In this project, we applied high-end technologies to provide an interface required by individuals with severe motor dysfunction who were incapable of operating existing interfaces due to their intense involuntary movements. Our research focused primarily on actual clinical use in order not to deviate from how disabled individuals actually use the devices. As a result, we succeeded in developing an interface that enabled users to independently operate their wheelchairs within a secure park environment.

However, two major hurdles remained, namely the high cost of the proprietary stereo vision sensor hardware that we developed to generate range images in real time, and the high cost of adapting the interface to various disabled individuals.

In response to the first problem, the commercial release of active range image sensors leveraging the pattern projection method has meant that the hardware can now be made for around JPY20,000, making it affordable for most disabled individuals, although with the limitation that it can only be used indoors. Resolving the remaining cost issue would enable the supply of an interface that has been keenly awaited by various disabled persons. Provided that it is only used indoors, the remaining issue of adapting the interface to various disabled individuals would be resolved.

To this end, we developed an interface based on an image range sensor for cerebral palsy patients who had difficulty using existing devices due to involuntary movements outside the target recognition site or spasticity, despite the fact that their gestures could be understood by a caregiver or other experienced individual [3]. Based on the concept of promoting harmony between the human operator and device, we conducted research and development over an expedited 1-year time frame on a single cerebral palsy patient (selected as a user who would typically have difficulty operating a conventional interface) to create a customized interface focusing primarily on this user's finger gestures, as well as head (i.e., nodding) and mouth (i.e., opening and closing) gestures.

A similar research project known as "OAK" ("Observation and Access with Kinect") is being conducted to develop a solution for assisting the activities of severely disabled persons [4]. The research aims to enable disabled users to operate a computer more intuitively through the combined use of software developed with a Windows software development kit. However, this project was primarily intended for the children

of disabled parents, and it is not intended as a system for classifying adaptable gestures for the disabled community as a whole. The research is also based on libraries of existing video games, which raises a fundamental problem in that there is no corresponding library for images of the user that are not taken from a frontal aspect. Another issue is that the device does not work without a particular type of sensor.

In the present study, we assume that all of the gesture recognition modules can be implemented using any available stereo vision (range image)-based human sensing technology, such as a real-time gesture recognition system [5], shape extraction based on 3D data [6], or data extraction based on long-term stereo range images [7]. We also assume that replacing the range sensor will not affect the usability of the interface. Our ultimate objective is to develop an interface that can automatically adapt to a wide range of body gestures, as well as long-term changes in how users perform these gestures.

## 2 Processing Structure

### 2.1 Collecting Data on Body Positions

Using the range image sensor, we recorded different voluntary gestures that the subjects from disability support groups and other organizations would like to see integrated into the interface. All subjects exhibited spasticity, spastic or involuntary movements, or were quadriplegics with severe motor dysfunction. Despite being able to move some parts of their body at will, all subjects had some form of motor dysfunction characterized by a constant impediment in the form of spastic or involuntary movements that made it difficult to use conventional switch-based interfaces and that severely restricted the parts of their body that they could move voluntarily. We used the range image sensor to gather data on the types of gestures that these severely quadriplegic subjects would like to use in an interface.

Focusing on these subjects who have great difficulty using a standard keyboard or mouse, we targeted the following body sites for gesture-based input.

- Hands and arms (arms, elbows, forearms, hands, fingers)
- Head (whole head movement, tongue extension/retraction, eye movement)
- Legs (exaggerated foot or leg movements)
- Shoulders

To date, we have collected gesture data for these body sites from 36 subjects over a period of about 2 years, while also listening to the opinions of the disabled users and their caregivers. In total, we have gathered data on 125 body site movements, including gestures that can be made with multiple body sites.

Before conducting the study, we obtained the informed consent of the subjects based on the approval of the Ergonomic Experimental Committee of the National Institute of Advanced Industrial Science and Technology and the Ethical Review Committee of the National Rehabilitation Center for Persons with Disabilities.

Three-dimensional (3D) movement data collected from the disabled subjects were classified and systematized based on the assumption that the movements could be

recognized from the range images. In this context, the term “systematize” essentially means classifying similar movements as gestures that can be recognized by a single underlying recognition module. In other words, we assume that a module can be created that can recognize gestures for each region of the body based on the collected data. Because this approach focuses on operating a computer in a static indoor environment with no movement [2], providing high-resolution range images should enable high-precision imaging of the body region of interest without the need for a sophisticated object model or image properties requiring significant computational resources. The results are shown in Table 1.

**Table 1.** Classifications of gestures

	Moving one finger	17
Hand	Moving wrist	12
	Moving hand	16
	Moving entire head	19
Head	Sticking out/retracting tongue	15
	Eyes (eye movement, blinking, opening eyes wide)	17
Legs	Moving one leg from the knee down	3
	Extending/retracting knee	5
Shoulder	Up/down, forward/backward	7
Other	Currently unclassified	17
	Total body sites	125

Based on the data collected from the 36 subjects, we classified 3 types of gestures for the hands and arms, 3 types of gestures for the head, 2 types of gestures for the legs, and 1 type of gesture for the shoulders. The camera is positioned so as not to hinder the subject and is ideally located to recognize gestures, so the classification is done on the assumption that gestures can be recognized with a single model. Movements that were classified as other types of gestures were those that clearly differed from the above-mentioned gestures, even when they originated from 1 of the 4 body sites, or those that originated from a distinct body position such as the ear.

## 2.2 Approach Based on the Extent of Voluntary/Involuntary Movements

We have collected and classified gestures obtained from people with severe motor function disabilities, and then developed a basic prototype recognition module capable of recognizing and identifying the gestures. We have classified all subjects into three basic types.

- Type 1:** Little involuntary movement and small voluntary movement
- Type 2:** Large involuntary movement and clear and large voluntary movement
- Type 3:** Large involuntary movement and small voluntary movement (the most difficult type)

If we investigate the gesture interfaces for the three types, the strategies for each type are obviously different. Because the type 1 users cannot move themselves, their recognition areas are not moving. If the system can detect the recognition areas exactly, it monitors only the small movements.

On the other hand, in types 2 and 3, the recognition areas are moving, and the methods of detecting and tracking the areas are the most important. The recognition of voluntary movement is the next step in processing.

We focus on type 1 users in this paper, and we have developed a gesture interface without a body part model. Our method is detecting the recognition area by 3D data exactly and recognizing the movement by the learning method. We consider that the method will be applied to the same type of users at low cost.

### 2.3 Developing Site-Specific Recognition Modules

Based on the data categories shown in Sect. 2.1, we tested a series of prototype recognition modules based on the assumption that a single module could be adapted to suit multiple subjects by manually adjusting the default settings. Below is a list of body parts that have been tested thus far. Based on the above categories, the interface has now been equipped with 2 types of hand modules, 2 types of head modules, and 1 type of leg module (Table 2).

**Table 2.** Recognition modules

Part of body	Recognition module	Model or not
Hand	Finger gesture recognition module	Model
	Arm gesture recognition module	Model
Head	Head recognition module	Model
	Tongue recognition module	Model
Legs	Knee gesture recognition module	Model
Any part	Closest part recognition module to camera	No model
	Simple differential recognition module	No model

Although the research was initially conceived based on body site-specific recognition modules, we subsequently conceived 2 types of recognition modules without site-related models in consideration of the size and nature of the movements. These modules are shown below, specifically the module for tracking the area closest to the camera and the module for extracting subtle movements.

**Recognition Module for the Site Closest to the Camera.** This module tracks the movement of the most proximal part of a recognition site situated closest to the camera. For subjects in whom site-specific classification (modeling) proves difficult, this module captures the target site as accurately as possible using distance data (shape data) and then learns to recognize that portion's movements (i.e., simple properties based on differences between frames).

**Simple Differential Recognition Module.** For subjects in whom site-specific classification (modeling) proves difficult, this module captures the target site as accurately as possible using distance data (shape data) and then learns to recognize that portion's movements (i.e., simple properties based on differences between frames).

In summary, we have developed 5 model-based site recognition modules and 2 recognition modules without models.

### 3 Long-Term Evaluation

We have gathered and classified the site-specific data of numerous disabled individuals and have developed individual recognition models corresponding to the data. While we will continue to gather and classify data, our main emphasis from now on will be to develop the techniques to adapt the recognition modules to individual users. This will involve both initial personalized adaptations and adaptations to long-term subtle changes. The long-term objective is that these adaptations will be implemented semi-automatically with only the caregiver's assistance. We therefore conceived the following experiment with 3 phases in order to develop adaptive techniques for these initial and long-term changes.

#### 3.1 Phase I: Gathering Basic Data

In the initial adaptation to individual subjects, the data requested by the subject will be recorded for a period of about 2 weeks. While the ultimate aim is to automate the personal adaptation process, the aim in this phase is to gather abundant data on subtly different wheelchair and bed positions each day. The data will then be fed into the recognition modules developed thus far so that they can acquire the various parameters.

#### 3.2 Phase 2: Gathering Supervised Data with Recognition Functions

In the phase after completing basic personal adaptation to the recognition modules, a recording system with recognition functions will be used to collect semi-supervised signal data. Specifically, the user will be instructed to play a general music video game (such as "*Taiko Drum Master*") while evaluating the recognition performance and gathering additional data.

This approach is not 100 % accurate, because the expected timing of the user's drum beat is presumed to occur within a certain time frame, and it will likely result in the acquisition of supervised signals exceeding a certain probability. We will spend approximately 2 weeks gathering the data and will then conduct adaptive processing of the various parameters, etc., to further enhance recognition performance.

Figure 1 shows the screen of a web-based application using gesture gymnastics. The application has the same basic architecture as the above-mentioned percussion rhythm game. The parts are designed to respond to gestures from up to 4 different body sites. Each part is matched to a gesture selected by the user, which the user then performs to operate the application.



Fig. 1. Gesture gymnastics

### 3.3 Phase 3: Testing in Long-Term Operation

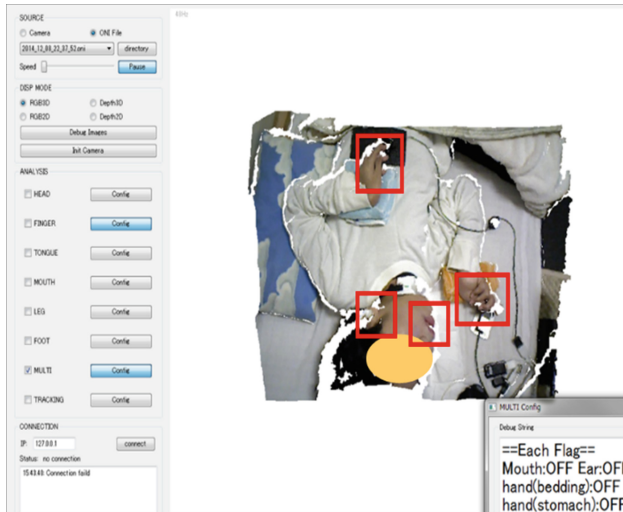
In this phase, we are conducting validation testing of an actual application upon further enhancing the recognition performance based on data such as daily changes in the relative position of the camera and user, and daily changes in the user's movements. The application used the same gesture gymnastics content as in Phase 2, with each user operating the application for 2 to 3 months. Users were asked to use the application every day if possible. The music and operating selections were based on methods best suited to each user. The aim of this testing was to assess the application's ability to handle actual long-term use after adjusting the parameters based on data acquired over a period of about one month.

## 4 Experiment

We have started long-term testing in 3 individuals with severe motor dysfunction. The target body sites, the level of motor function at each body site, and the extent of involuntary movements all differ among these subjects.

### 4.1 Subject with Multiple Target Body Sites with Minimal Involuntary Movement

The main subject in this experiment had little involuntary movement and small voluntary movement. Figure 2 indicates the monitoring image of the subject by the RGB-D camera. He is lying in his bed and basically cannot change his body position by himself.



**Fig. 2.** User and four recognition areas

The four rectangles indicate four recognition parts. The system is able to monitor the specific movement of his mouth, the movement of his right ear, and the movements of his right and left forefingers individually. We are now setting first positions and their parameters manually. On the other hand, the daily parameters are set automatically by the user's particular actions. For applying to the daily position changes of the camera and the user, the user performs the music game for one to two minutes, and the system adjusts all parameters automatically.

The cables in Fig. 2 are for the one-switch device and the trackball. The user could use his right and left forefingers for the interface. Our system uses one RGB-D camera, which monitors his whole upper body, and four switch interfaces can be used simultaneously. We consider that this user is suitable for our interface system to improve it through our experiments.

The long-term evaluation testing targeted a combination of 2 recognizable body sites selected by the user (Fig. 3).

## 4.2 Subject with Typical Head Movements

This test targeted a subject with a high cervical spine injury who was able to operate a wheelchair but required the assistance of a caregiver to operate remote control switches and a personal computer. Head movement was selected as the target recognition site. The subject exhibited almost no involuntary movements and, although he did not have the range of head motion available to a healthy individual, the level of difficulty required for gesture recognition was not high. While simple estimation of head inclination can be performed with existing 2D recognition engines, this experiment was conducted using range images with the aim of not only estimating head inclination, but



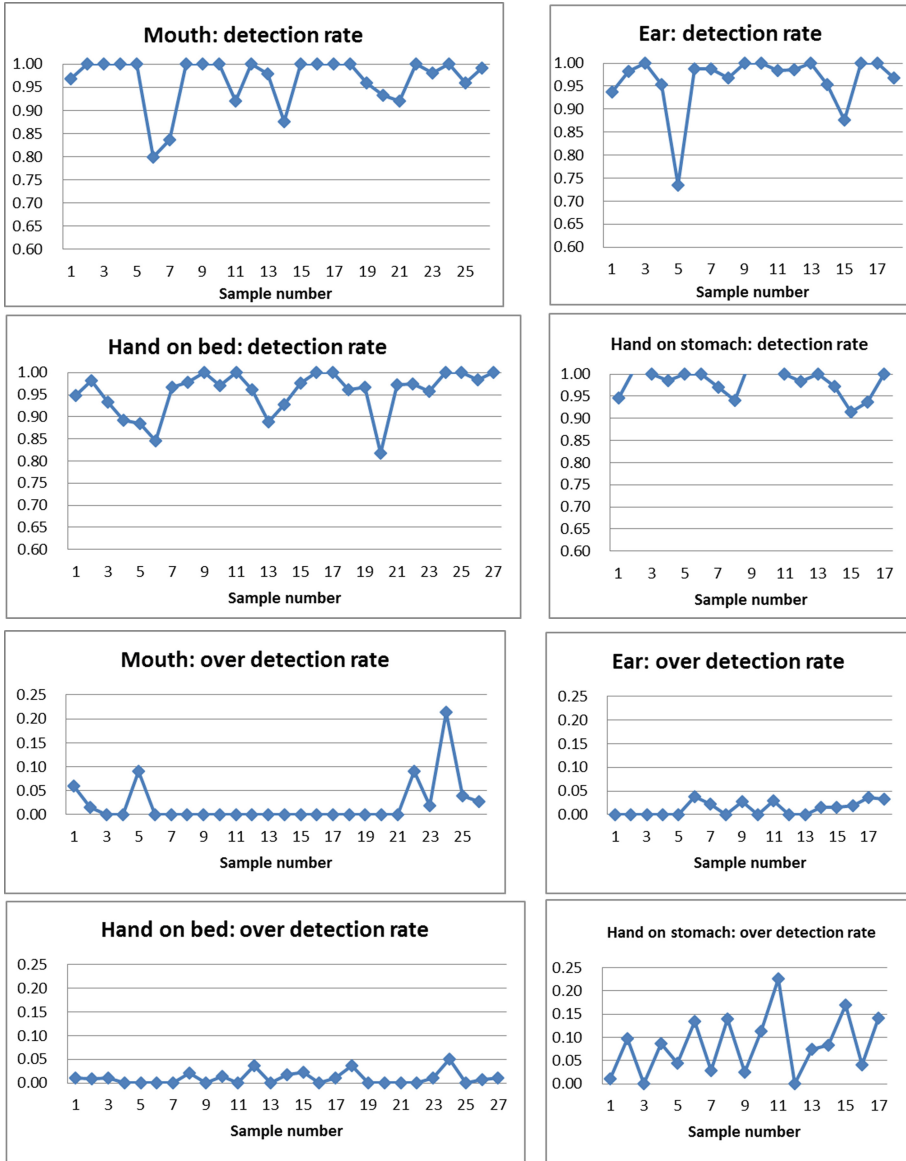


Fig. 3. Parts: simple detection rate and over detection rate

also enabling combined recognition of other sites on the subject’s upper body, and adapting to users whose sideways and vertical movements differed in magnitude.

The test was conducted by assigning 3 channels according to the user’s preferred gestures, namely a frontal → right-facing → frontal gesture, a frontal → left-facing → frontal gesture, and a frontal → downward-facing → frontal gesture. The evaluation testing also obtained results on these 3 operating methods (Figs. 4 and 5).



Fig. 4. Gesture recognition for subject 2

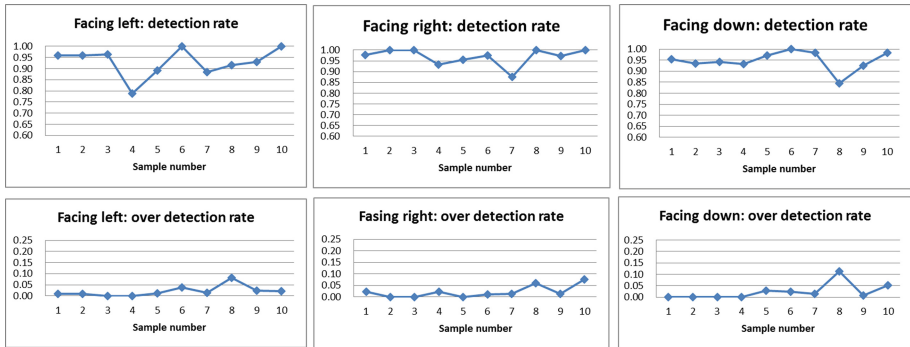


Fig. 5. Facing left, right, and down: simple detection rate and over detection rate

### 4.3 Finger Recognition Subject with Frequent Involuntary Movements

This test targeted a subject with cerebral palsy who had the highest level of difficulty in terms of large involuntary movements and limited recognizable voluntary movements. The subject’s only voluntary movement was the middle finger on one hand, while his arms exhibited significant involuntary movements that prevented his hands from remaining still, thus necessitating constant tracking of hand position. The subject’s constantly-moving hands also made it necessary to perform recognition of the fingers regardless of whether they were facing forward, left, or right (recognition was simply not possible when images of the subject’s fingers could not be captured).

The actual recognition process is depicted in Fig. 6. The red rectangle indicates the target recognition site. A colored sack was attached to enhance the resolution of the range images and to offset the intensity of the subject’s involuntary movements.

There was only 1 type of operation, so the evaluation test was conducted every day, and the test music was changed where appropriate (Table 3).

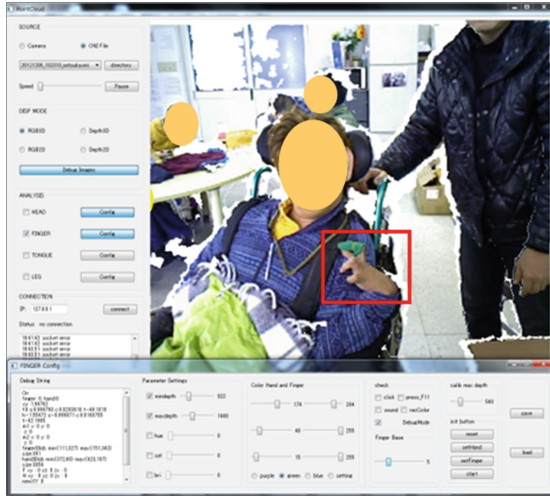
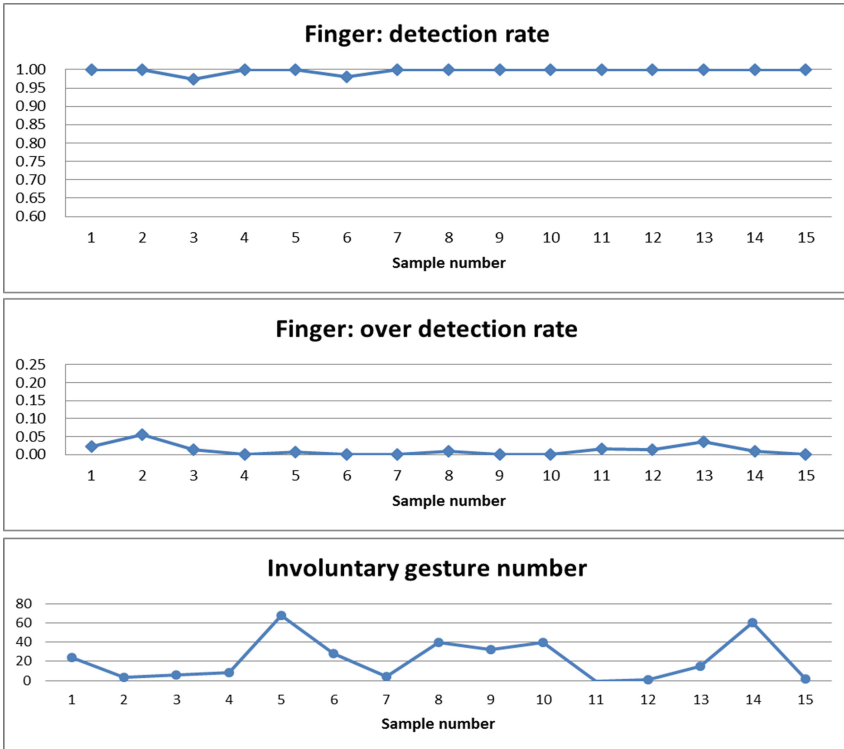


Fig. 6. Gesture recognition for subject 3

Table 3. Finger: simple detection rate, over detection rate, and involuntary gesture numbers



## 5 Conclusions and Future Work

In the present study, we attempted to adapt an inexpensive, contactless, non-restrictive sensor to various disabled individuals by gathering and classifying 3D data on the diverse movements of actual disabled users. Using the data gathered from 36 users, we obtained gesture data from 125 body sites. We then classified these data into a total of 9 body sites capable of voluntary movement, namely the arms and hands (fingers, wrists, forearms), head (head movement, tongue extension/retraction, eye movement), legs (swinging and opening and closing of the knees), and shoulders.

At the same time, we classified disabled users by focusing on the size and nature of target site movements. This classification was based on the extent of involuntary movements, as well as voluntary movements, at target sites. As a result, we were able to classify 3 types of disabled individuals.

We then developed the gesture recognition modules based on the classification of target body site and the type of disabled user.

Specifically, we developed a head recognition module, a finger recognition module, and a simple differential recognition module without a target model, and conducted long-term testing on 3 disabled individuals. The testing was divided into 3 phases, and the recognition results were presented. In terms of basic testing, useful recognition results were obtained in 3 subjects.

In the future, we intend to continue long-term testing by conducting the third test phase involving the use of an actual application and the automatic acquisition of routine data on recognition parameters.

**Acknowledgment.** This work was supported in part by a Health and Labor Sciences Research Grant: Comprehensive Research on Disability Health and Welfare in 2014. The authors gratefully acknowledge the many who have supported and encouraged this work.

## References

1. Yoda, I., Itoh, K., Nakayama, T.: Collection and classification of gestures from people with severe motor dysfunction for developing modular gesture interface. In: Antona, M., Stephanidis, C. (eds.) UAHCI 2015. LNCS, vol. 9176, pp. 58–68. Springer, Heidelberg (2015)
2. Yoda, I., Tanaka, J., Raytchev, B., Sakaue, K., Inoue, T.: Stereo camera based non-contact non-constraining head gesture interface for electric wheelchairs. In: Proceedings of International Conference of Pattern Recognition ICPR 2006, vo. 14, pp. 740–745 (2006)
3. Yoda, I., Nakayama, T., Ito, K.: Development of interface for cerebral palsy patient by image range sensor. Grant Program Report Tateishi Sci. Technol. Found. **22**, 122–125 (2013). (in Japanese)
4. Iwabudhi, M., Guang, Y., Nakamura, K.: Computer vision for severe and multiple disabilities to interact the world **37**(12), 47–50 (2013) (in Japanese)
5. Monekosso, D., Remagnino, P., Kuno, Y.: Intelligent environments: methods, algorithms and applications (Advanced Information and Knowledge Processing). In: Yoda, I., Sakaue, K. (eds.), Chapter 6 Ubiquitous Stereo Vision for Human Sensing, pp. 91–107. Springer-Verlag, Heidelberg (2009)

6. Hosotani, D., Yoda, I., Sakaue, K.: Wheelchair recognition by using stereo vision and histogram of oriented gradients, in real environments. *IEEE Workshop Appl. Comput. Vis.* **2009**, 498–503 (2009)
7. Onishi, M., Yoda, I.: Visualization of customer flow in an office complex over a long period. In: *Proceedings of International Conference on Pattern Recognition (ICPR)*, pp. 1747–1750 (2010)
8. Tanikawa, T., Yoda, I., et al.: Home environment models for comfortable and independent living of people with disabilities. *J. Hum. Life Eng.* **12**(1), 23–27 (2011). (in Japanese)
9. “Survey on persons with physical disability 2006,” Department of Health and Welfare for Persons with Disabilities, Social Welfare and War Victims’ Relief Bureau, MHLW (2006)