

Collection and Classification of Gestures from People with Severe Motor Dysfunction for Developing Modular Gesture Interface

Ikushi Yoda¹(✉), Kazuyuki Itoh², and Tsuyoshi Nakayama²

¹ National Institute of Advanced Industrial Science and Technology (AIST),
Tsukuba, Japan

`i-yoda@aist.go.jp`

² Research Institute, National Rehabilitation Center for Persons with Disabilities
(NRCD), Tokorozawa, Japan

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

Abstract. This study explores gesture controlled user interfaces for people with severe motor function disabilities stemming from cerebral palsy, quadriplegia, and traumatic brain injury. As a result of their disabilities (involuntary movement and spasticity), it is nearly impossible for these individuals to use conventional interface switches and other input devices to access and use a computer. The ultimate objective of this work is to provide these users with user-friendly, cost-effective gesture controlled interfaces that will enable them to comfortably operate a personal computer. We have now succeeded in developing a non-contact, non-restraining interface based on an off-the-shelf image range sensor that recently became available. In addition, we surveyed a large number of disabled subjects and compiled a fairly exhaustive collection of gestures that these users are capable of making, and classified these voluntary movements in terms of the body part involved. Finally, a series of recognition modules have been developed that are optimized to recognize the gestures associated with each body part (hand, head, leg, etc.). This paper provides an overview of the gesture data collection and classification processes, and discusses the development of the recognition modules.

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

1 Introduction

There is a significant number of people who, due to disabilities of one kind or another, are unable to send and receive email or surf the web using a computer. More specifically, we refer to individuals with severe motor function disabilities who find it difficult or impossible to use conventional input devices due to spastic or involuntary motion, or limited range of motion or muscular weakness in the lower limbs. Currently, no convenient interface exists, and those who are stuck using the increasingly limited specially customized switch type interface find it extremely cumbersome to use for operating a computer [8]. Especially for those who are unable to freely venture

outdoors, an environment that allows them to exchange email, surf the web, make online purchases, and other net-based activities really holds the key to a more enjoyable, fulfilling life. Yet for the disabled who are only able to use a simple switch type input device, these more sophisticated operations are all but impossible, not to mention the enormous costs involved in adapting these devices to the user's evolving disability or condition as the user gets older. The information gap will only continue to widen as the information society evolves and the disabled are left further behind.

The goal of this research is to develop a robust gesture controlled user interface that makes it relatively simple to operate a computer (including character input) for the motor function disabled who are uncomfortable or unable to use a keyboard or a mouse. Specifically, we developed a non-contact, non-restraining interface based on a common off-the-shelf range image sensor that provides a cost-effective interface within the budget of almost everyone. Most importantly, the technology must be customizable so it can be readily tailored to the various stages and conditions of the disabled population at low cost. This we achieved by surveying and collecting the widest possible range of movements that might be exploited as gestures, categorizing the movements based on part of the body, and developing modular recognition engines that recognize and identify the movements.

In pursuing work on a gesture controlled interface for IT purposes that involves enormous freedom of movement yet is very difficult to standardize, our objective is to focus first on the most severely disabled where the need is greatest and move toward a standardized gesture interface in the future that is both versatile and essential for categorizing the full range of movements that can serve as gestures.

This paper details our efforts to collect and classify gestures obtained from people with severe motor function disabilities, then develop a basic prototype recognition module capable of recognizing and identifying the gestures.

As part of an earlier project to aid people with severe disabilities, the authors developed a head gesture interface system for individuals with severe cerebral palsy who are unable to operate a wheelchair [1]. The project exploited high-end technology to finally give handicapped users for whom no input devices had been available with an interface they could actually use. This was a groundbreaking development for it brought together intelligence information experts with expertise in cutting-edge intelligence research in collaboration with rehabilitation experts. While this involved state-of-the-art technology, it had to be implemented within the framework of the practical equipment used within the disabled community. So, in terms of actual clinical practice, we gave highest priority to:

- Carefully consulting and listening to the views of patients themselves and their families from the very beginning and every step of the way all through the development.
- Mounting sensors near the joystick, and otherwise conforming to the actual settings of typical electric wheelchairs.
- Ensuring operability indoors or outdoors, in direct sunlight and under tree cover.
- Autonomous self-reliant operation once the caregiver turns the switch.

The work was carried out with these realistic objectives in mind. As a result, the project gave users the ability to move about safely and autonomously within public parks.

Yet two major hurdles remained. First, the unique stereo vision sensor hardware that we developed for generating range images in real time is simply too expensive (cost could be brought down if mass produced, but initial justification for mass production is problematic), and second, it is too costly to tailor the device to accommodate the full range of symptoms and conditions of the disabled population.

At least for indoor use, the first challenge has now been resolved with the availability of several off-the-shelf range image sensors featuring active pattern projection—Xtion PRO [9], Xtion PRO LIVE [10], KINECT for Windows [11], Leap Motion [12], and others—that can be readily obtained by virtually anyone at a modest cost of around 200US\$. We are now starting to see accurate consumer-oriented devices on the market that work just fine over relatively short distances when not exposed to direct sunlight. If we could come up with a solution to the second challenge of tailoring the system to different user conditions, we could provide the viable interface that is so earnestly sought by the disabled community. For indoor environments at least, the only remaining barrier was to figure out how to adapt the technology to various individual conditions and disabilities.

This motivated the authors to develop an image range sensor-based cerebral palsy interface [3] for disabled who are unable to use the common input devices (aside from a caregiver or other familiar attendant or friend who can interpret some spastic movement or a particular bodily movement). Based on the notion of “harmony between man and machine,” we devised an agile scheme over a one-year allotted time frame tailored for a particular subject, a man disabled with typical cerebral palsy who was unable to use conventional input devices. The interface mainly involved finger movements, supported by gyrations of the neck and opening and closing the mouth.

In a similar vein, the authors exploited Microsoft’s Kinect sensor in developing observation and access with Kinect (OAK) [4] as a solution for assisting the activities of the severely disabled. The idea was to enable disabled users to directly or more intuitively operate a computer by combining our scheme with software developed using the Kinect software development kit (SDK) for Windows. Note that this project was primarily intended for the children of disabled parents, and was never really intended as a scheme for organizing and classifying adaptable gestures for the disabled community as a whole. We would also note that this scheme is based on libraries of existing games, which raises a fundamental problem: if you haven’t previously captured the person from the front, then there is no corresponding library in the first place. Finally there is the problem that this device does not work without a particular type of sensor.

For the purposes of this work, we assume that all of the modules for recognizing gestures could be implemented using any of the stereo vision (range image)-based human sensing technologies available including real-time gesture recognition systems [5], shape extraction based on 3D information [6], data extraction based on long-term stereo range images [7], and so on. We also assume that exchanging or swapping out the range sensor should not affect the usability of interface. Our ultimate objective is automatic adaptation of the system to the widest applicable range of parts of the body that might be used to make gestures and long-term shifts in how users make gestures.

2 Collecting and Classifying Subject Data

Collecting Subject Data. Using the range image sensor, we recorded voluntary gestures for the interface from a range of disabled participants affiliated with the National Rehabilitation Center for Persons with Disabilities and a number of other agencies and organizations that deal with disabled individuals in the community. The participants had a range of different disabilities including:

- Children and adults with cerebral palsy (Spastic, athetoid, mixed types).
- Spinocerebellar degeneration, Parkinson's disease, and other neurodegenerative conditions.
- Muscular dystrophy and other muscular disorders.
- Survivors of traumatic brain injury (wounds, injury, stroke).
- Quadriplegics exhibiting spastic or involuntary motion due to genetic factors, syndromes, or unknown causes.
- High quadriplegics.

All of these subjects exhibited spasticity, spastic involuntary movement, or were quadriplegics with severe motor function disabilities. Even though they might have the ability to voluntarily move some part of the body, all subjects had severe motor function disabilities and were extremely limited in the body parts they could move voluntarily; they were significantly hindered by spasms and involuntary movements, and all found it extremely difficult to use existing switch type or other input devices. With this group of severely handicapped quadriplegics and other disabled individuals, we used the range image sensor to collect the full range of gestures they thought they might be able to use.

For these subjects who have great difficulty using an ordinary keyboard or mouse, the following parts of the body showed promise for making gestures that could be used for input:

- Hand and arm (arm, elbow, forearm, hand, finger).
- Shoulder.
- Head (motion of entire head, sticking out/retracting the tongue, eye movement).
- Leg movement (exaggerated movement of the foot or leg).

We collected a wide range of gestures from these four basic regions of the body over an eighteen month period using 33 subjects, while carefully consulting and listening to the views of the disabled users themselves and their caretakers. Counting gestures that could be made using multiple sites or regions of the body, we assembled gestures produced by a total of 104 parts or combinations of body features.

We obtained the consent of the subjects to undertake this work after explaining the nature of this project and had 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.

Classification of Gestures for Each Part of the Body. 3D movements collected from the disabled subjects are systematized as they are classified, assuming that they can be

Table 1. Classifications of gestures

Hand	Move 1 finger		15
	Move wrist		12
	Move hand		13
Shoulder			6
	Up/down	5	
	Forward/backward	1	
Head	Move entire head		17
	Stick out/retract tongue		1
	Eyes		15
	Blinking	10	
	Movement of eyes	3	
Legs	Open eyes wide	2	
	Point foot from ankle		3
	Move one leg from knee down		1
Other	Extend/retract knee		1
	Currently unclassified		12
	Total body sites		104
	Total sites including alternative forms		112

recognized from the range images. By systemization, we mean essentially the same kind of motion, a gesture classification that can be recognized by a recognition module that serves as a base. In other words, we assume that a module can be created that can recognize gestures for each and every region of the body based on the data that was collected. With this approach, since we are focusing on the operation of a computer in a quiet indoor environment with no movement [2], assuming that high-resolution range images are available, the body region of interest can be captured with excellent accuracy without having to use an advanced object model or image features that require significant computational resources. The results are shown in Table 1.

Based on the data collected from the 33 subjects in this project, we classified 3 areas of the body for hand, 3 areas for head, 1 area for shoulder, and 3 areas for the legs. The camera is set up in such a way that it doesn't disturb the subjects and is ideally located to recognize gestures, so the classification is done on the assumption that gestures can be recognized with a single model.

Only 33 subjects were recruited for the study, but we shot the same subject several times on different days to increase the number of regions or parts of the body that were filmed. We found that by reshooting the same subject on different dates, we were able to capture a number of different variations or alternative forms. This proved to be invaluable data for assessing day-to-day variation in the movement of the subjects. Counting these variations, we came up with a total of 112 gesture sites including the alternative forms, as shown in Table 1.

3 Recognition Modules for Different Parts of the Body

In order to recognize or identify the gesture movements that have been assembled so far, a series of prototype recognition modules was developed on the assumption that a single module can accommodate multiple subjects by manually tweaking parameters and other adjustments.

Finger Gesture Recognition Module. For finger gesture recognition, we adopted the following specifications to determine whether a single finger was bent or not and to apply a colored finger cot (single finger of a colored glove).

- Determine if a finger is bent or not.
- Apply finger cot to any 1 of 5 fingers.
- Select red, green, or blue finger cot (choose color that contrasts with clothing).

The prototype implemented for this project is built with recognition parameters set for a particular user, but as one can see from the screenshot shown in Fig. 1, the parameters can be manually tuned for a different user (eventually, this feature will be automated so day-to-day fluctuations are handled automatically).

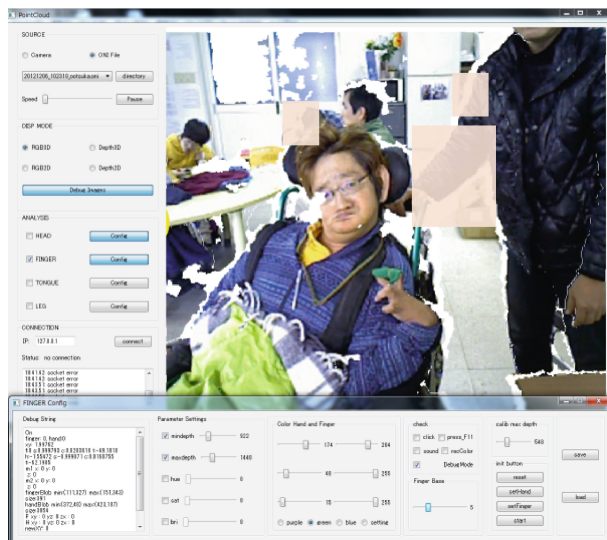


Fig. 1. Screenshot of finger gesture recognition module

The parameters that can be manually adjusted are listed in Table 2.

The recognition algorithm first detects a finger in the specified range space, then extracts a hand based on the position of the finger, and finally calculates the degree the finger is bent from the relationship between finger and hand. Basic steps of the algorithm are as follows:

Detect a finger

- Set 3D space that includes hand
- Extract 3D texture image
- Extract same color as the finger cot from texture image
- Label range of extraction
- Finger is recognized as portion marked by maximum label

Detect a hand

- Extract skin colored region from 3D extracted texture image adjacent to the finger
- Label range of extraction
- Object closest to the finger is recognized as the hand

Determine degree finger is bent

- Calculate moment of finger and hand. Calculate point group for both finger and hand as a 3D moment. Since facing the screen, next, calculate 2D screen
- Calculate 2 axes angle from the moment
- Calculate difference between angles of finger and hand; determine finger is bent when difference exceeds the threshold

Table 2. Parameters that can be manually adjusted

Setting	Description
Parameter settings	Feature to be recognized
• min-depth	Min range image setting
• max-depth	Max range image setting
• hue	Hue
• bright	Brightness
Hue hand and finger	Specify min and max saturation values of hand
	Specify color of finger cot
Finger base	Threshold for determining finger is bent
Check	Specify action to take when finger is bent
	(click; sound is made when pressing the ‘a’ key)

Arm Gesture Recognition Module. For arm gesture recognition, we adopted the following specifications to identify swinging the whole forearm from the elbow (see Fig. 2).

- Determine if forearm of one hand is swinging
- Set up camera so the arm to be recognized fits easily within the angle of view

Basic steps of the algorithm are set forth as follows:

Detect arm

- Set 3D space that includes arm
- Extract arm range image against 3D base



Fig. 2. Screenshot of arm gesture recognition module

Track arm

- Track with particle filter
- Determine likelihood of particle from inter-frame difference
- Now possible to track moving body part (the arm)

Determine swing of arm

- Estimate state of the arm from length of shifts on sway of center of gravity of set of particles
- Classify based on “exaggerated swinging of the arm” and “no movement”

4 Head Gesture Recognition Module

For head gesture recognition, first the normal direction is derived from the range image area centering on the nose, then this orientation is used as the orientation for the face. The user can employ any motion or movement he or she wants to trigger the switch.

- Estimate direction of face in real time
- Operate as switch when setup is oriented toward the face in a particular direction
- Facing to the right generates a click event

The light blue bar near the eyebrow of the subject in Fig. 3 shows the normal direction of the face. By changing the normal direction for turning on the switch, this can be used to describe an action.

The sequence of algorithm steps for estimating head orientation is as follows: face tracking, nose tracking, then calculate the normal area of the face.

Head tracking

- Calculate approximate area of head based on distance information
- Extract just face label using labeling

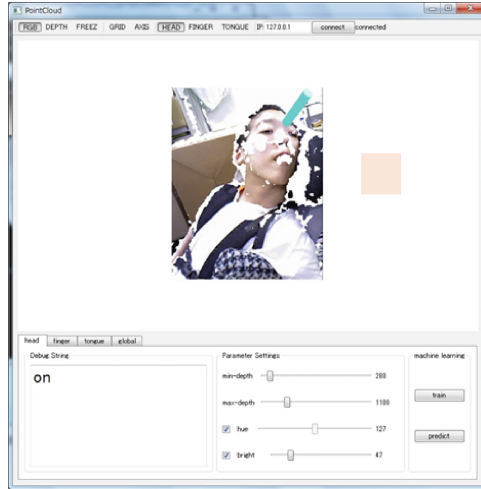


Fig. 3. Screenshot of head pose recognition module

Nose tracking

- Normalize zoomed, rotated, positioned extracted face image
- Nose is closest point to the camera

Calculate normal area of face

- Calculate face normalization (orientation) from the range image area centering on the nose.

Tongue Gesture Recognition Module. For tongue gesture recognition, we simply determine whether the subject is sticking out her tongue deliberately or not; the switch is turned on when the tongue remains out for more than a certain number of seconds. Moreover, the user can assign any motion or movement to trigger the switch. Currently, in setting the color threshold to match the color of individual’s tongue and lighting environment, the individual and the lighting environment are highly dependent.

Tongue Gesture Recognition Algorithm

- As with the head recognition algorithm, this algorithm also starts by tracking the face
- Convert RGB information to HSV information in the face label
- Perform filtering based on the tongue hue threshold setting
- Tongue is recognized when the label exceeds a certain size

Knee Gesture Recognition Module. The finger, head, and tongue gestures can all be recognized simultaneously by camera settings, but the camera is set up differently to recognize knee gestures. For recording knee gestures, an extension arm is used to mount the camera up above the display looking down so the knees are caught at the center of the image (see Fig. 4)

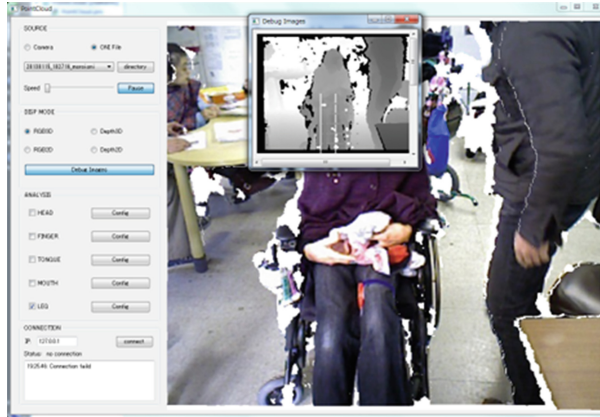


Fig. 4. Screenshot of knee gesture recognition module

- Estimate position of knees in real time
- Switch is triggered by moving the knees together or closing the knees

Basic steps of the knee position estimation algorithm essentially consist of first extracting the knee region, then estimating the knee position on the left and right sides with the hill-climbing method. This particular user defined the act of holding both knees together beyond a certain interval as triggering the switch.

5 Conclusions and Future Work

We began this project with the idea of developing gesture controlled user interfaces to enable people with disabilities to freely access and use information devices using simple gestures. To achieve this goal, the first stage is to compile and classify a collection of 3D actions or gestures that disabled users are capable of making using an economical off-the-shelf image range sensor. In this work, we gathered gesture data from 33 subjects, based on 104 different sites or parts of the body. We systematically categorized this data in terms of 10 total parts of the body that disabled users can employ to make voluntary movements that could be exploited as gestures: 3 areas of the body for hand, 3 areas for head, 1 area for shoulder, and 3 areas for the legs.

In addition, we constructed a series of prototype recognition modules and demonstrated their ability to recognize 5 types of movement among these 10 parts of the body: hands and arms (finger bending and arm waving), head (head swinging and sticking out and retracting the tongue), and legs (opening and closing the knees). Parameters are adjusted manually on the prototype modules, but ultimately we assume such adjustments will be done automatically to easily accommodate a wide range of disabled users.

For this current project we dealt with 33 subjects and 104 parts of the body, but a somewhat larger scale initiative involving around 50 subjects is needed to build a more robust modular gesture recognition platform that we envision. Since we have only

tested the recognition modules developed so far on just a few subjects, we still have not gotten beyond the prototype stage. By increasing the number of subjects and the number of body part sites, we are confident that the approach we advocate here will lead to the development of gesture recognition modules with greater classification accuracy and wider scope.

Acknowledgment. Part of this work was supported by a Health and Labor Sciences Research Grants: Comprehensive Research on Disability Health and Welfare in 2014. The authors gratefully acknowledged the many who have supported and encouraged this work.

References

1. 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, vol. 4, pp. 740–745 (2006)
2. 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)
3. Yoda, I., Nakayama, T., Ito, K.: Development of Interface for Cerebral Palsy Patient by Image Range Sensor. Grant Program Report of Tateishi Science and Technology Foundation, vol. 22, pp. 122–125 (2013) (in Japanese)
4. Iwabuchi, M., Guang, Y., Nakamura, K.: Computer vision for severe and multiple disabilities to interact the world. *ITE Tech. Rep.* **37**(12), 47–50 (2013) (in Japanese)
5. Monekosso, D., Remagnino, D., Kuno, Y.: Intelligent environments: methods, algorithms and applications. In: Yoda, I., Sakaue, K. (eds.) *Ubiquitous Stereo Vision for Human Sensing*, Chapter 6. *Advanced Information and Knowledge Processing*, pp. 91–107. Springer, London (2009)
6. Hosotani, D., Yoda, D., Sakaue, K.: Wheelchair recognition by using stereo vision and histogram of oriented gradients, in real environments. In: *IEEE Workshop on Applications of Computer Vision 2009*, pp. 498–503 (2009)
7. Sato, N., Yoda, I., Inoue, T.: Shoulder gesture interface for operating electric wheelchair. In: *IEEE International Workshop on Human-Computer Interaction in Conjunction with ICCV 2009*, pp. 2048–2055 (2009)
8. 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)
9. http://www.asus.com/Multimedia/Xtion_PRO/
10. http://www.asus.com/Multimedia/Xtion_PRO_LIVE/
11. <http://www.microsoft.com/en-us/kinectforwindows/>
12. <http://www.softkinetic.com/>