GESTURE RECOGNITION RESEARCH FOR HUMAN-MACHINE SYMBIOTIC ENVIRONMENT

T.Kirishima[1], K.Sato[2], and K.Chihara[3]
[1] Department of Electrical Engineering, Nara National College of Technology,
22 Yata-cho, Yamatokoriyama-shi, Nara, 639-1080 Japan
[2] Graduate School of Engineering Science, Osaka University,
1-3 Machikaneyama-cho, Toyonaka-shi, Osaka, 560-8531 Japan
[3] Graduate School of Information Science, Nara Institute of Science and Technology,
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101 Japan

ABSTRACT: The drastic increase in computing performance of mobile sensing devices will soon trigger emergence and permeation of ubiquitous computing technologies. Breakthroughs in hardware technology are already demanding novel applications for intelligent robots, security, telemedicine, and virtual reality. Gesture recognition technology that enables the recognition of human natural behaviors in both real and virtual worlds can play important roles and will become indispensable in our future lives. Gesture recognition research is a field that attempts to make computers understand human intentions and body configurations through visual observations of their behaviors and activities. It can contribute as a human interface in both real and virtual worlds.

1. INTRODUCTION

Everyone knows that human body movements can bear rich meanings in human-to-human communication. We call it a "gesture." Today, it began to emerge as a means for interaction with digital computers, or intelligent machines. Using a gesture as a means for human-machine communication is a new paradigm that leads to human-machine symbiotic environments in future ubiquitous society. For the symbiosis of humans and machines, machines are required to understand human gestures[1]. Thanks to the innovations in computer hardware technology, even a tiny mobile computing devices can now process streams of visual information in real-time. The rest is how to implement and organize the machine intelligence that will widely spread in the real world. In this paper, multilateral aspects of human gestures are briefly noted and some of the bottlenecks are explained that researchers frequently encounter when constructing a gesture recognition system. Then, some ideas are presented on how to design and how to evaluate a gesture recognition system.

2. IS GESTURE RECOGNITION AN “EASY-TO-SOLVE PROBLEM?”

Today, gesture recognition systems are being developed worldwide and one can easily find highly successful systems. Some may think that most of gesture recognition problems in terms of engineering have already been resolved and they belong to “easy-to-solve problems.” In my view, only a part of problems that are easy to solve has been solved since gesture-related research belongs to an interdisciplinary domain that is so wide and so deep. Before dealing with gestures in terms of engineering, multilateral aspects of gestures need to be addressed.

2.1 Social / Psychological / Cognitive Aspects

Gestures are used by people of any age, any sexuality, and any race. The use of body movements is universal and indispensable for acquiring man's such basic skills as following:

(1) The ability to behave and interact with objects / people in the real world.
(E.g., eating, touching, grasping, walking, collision detection and avoidance)

(2) The ability to acquire one’s own body image.
(E.g., interaction with mirrors in order to know “I am real and a social existence”)

Moreover, the use of gestures is indispensable for developing man's such social communication skills as following:

(1) Understanding basic emotions and intentions of communication partner from his/her behavior[2]. (E.g., neutral, anger, sadness, happiness, fear, surprise, disgust, and teasing.)

(2) Understanding Body Language.
Each typical body motion and its meaning are loosely defined in an arbitrary manner. This requires only a local agreement.

(3) Understanding Sign Language.
Special body configuration, motion, and its meanings are strictly defined and shared in deaf community. This requires a long-term education and training.

In a social situation, gestures that are based on communication protocols are intuitive, flexible, and easy to understand and share. Without communication protocols, gestures are often ambiguous, misleading, and difficult to understand and share. How could we make use of such gestures in the context of engineering?

2.2 Engineering Aspects

Gesture recognition techniques are particularly useful in our daily domains or contexts:

* Smart / Intelligent Homes
  - Security use (intruder detection / behavior monitoring / person identification by gait analysis)
  - Appliance use (room-guidance / automatic control of lighting, air-conditioning and ventilation)
  - Welfare use (nursing-care at home by partner robots that can understand human intention)
* Human-Computer Interaction (HCI)
  - Entertainment (intuitive control of interactive games and navigation in virtual worlds)
  - Repetitive stress injuries (RSI) syndromes prevention (especially, avoiding tenosynovitis)
  - Physical exercises assistance (bedridden-state prevention, caregiver's distraction support)

Especially, in HCI, gestures can be an excellent means for:

- Operating a computer (substitution for keyboards and mice)
- Navigation in VR / AR / MR environments (walkthroughs, issuing gesture commands)
- Interaction with a virtual agent / object (CG) (direct communication / manipulation)
- Interaction with a real agent / object (robots) (direct communication / manipulation)

Apparently, the usability of gestures in above domains is promising and we have witnessed many excellent demonstrations. But only small portions of these are put into practical use. Why?

3. ARE THERE ANY TECHNICAL BOTTLENECKS?

3.1 Hardware aspect

Although hardware resources are limited (e.g., limited number of processors, and limited amount of memory), hardware-related limitations are not an obstacle to gesture recognition research since hardware performance and architectures have been dramatically improved.

3.2 Sensor aspect

Taking a gesture is a spatio-temporal event. If one wishes to deal with all kinds of gesture, the highest spatio-temporal resolution will be required. In reality, most of vision sensors that are available today and tailored to fit human vision properties strictly operate at fixed sampling rate and at fixed resolution (NTSC, PAL, SECAM, etc). Unfortunately, facial or hand gestures could require much higher spatial or temporal resolution. Validity of particular spatial or temporal resolution cannot be judged without considering the nature of target gestures. Potentially, there is a great need for a vision sensor that can change its sampling rate and spatial resolution dynamically responding to the demands from computer vision algorithms.

3.3 Software/Algorithm aspect

3.3.1 Sampling problem

Sampling theorem is basically applicable to the signals of one-dimensionality. It provides criteria for the minimum sampling frequency to recover or approximate original signal. But, is it also applicable to the gesture image sampling problem? Lack of sampling theory that takes recognition rate into account suggests that the ideal temporal resolution cannot be determined theoretically. Then, how can ideal sampling rate be determined that could appropriately classify the target actions? It is clear that a set of similar gestures could require much higher sampling rate than a set of distinct gestures. The ideal sampling rate will depend on the combination and the complexity of the gestures to be recognized.

3.3.2 Focus of attention / spatial-segmentation problem

To figure out the meaning of given gestures, one needs to know which part of body to see and what kind of motion features to see. This knowledge or protocol is formed through the face-to-face communication in our daily lives. This protocol-forming activity requires the so-called “focus of attention” capability. Unfortunately, it is not always evident which part of the body to extract from an image. From image recognition perspective, the body parts extraction is not a prerequisite process. When we try to take or interpret gestural actions, we usually pay greater attention to the moving parts of body. Actually, we unconsciously observe both regions of change and regions of little or no change since the combination of these regions could be very useful segmentation cues that reflect the signer's intention. A gesture is taken and is understood according to this visual communication protocols. Forming and sharing these protocols on gestures should greatly facilitate the communication between human and machines[3].

3.3.3 Temporal-segmentation problem

Communication protocols stipulate when a gesture begins, continues, and ends. Signed gestures are usually complex and concurrently use facial gestures, stipulated hand postures and movements. They usually accompany organized, systematic, and strong rules, and require language level understanding. Only well-trained people can understand sign language communication since it requires temporal-segmentation in light of contexts. Sign language protocols that are consistent and precisely defined are necessary. These are acquired through intensive training with a teacher.

On the other hand, understanding of local gestures requires only protocols that are arbitrary and temporary. Usually, local gestures are simple and repetitive movements. This suggests that only agreements on spatial segmentation cues are needed. In reality, computation of image difference is usually enough for the temporal segmentation of local gestures.

From above examples, we can see that there can be a wide variety of human communication protocols. When we consider the nature of Human Computer / Robot Interaction, gesture protocol learning[3] is very useful since it does not require intensive training on the part of user. At the same time, protocol-based gesture recognition provides the basis for sign language understanding. It can generate a sequence of symbols that will help symbolic representation and interpretation of signed gestures. Temporal-segmentation problems should be treated separately at different levels.

3.3.4 Trade-off problem between processing speed and recognition rate

In gesture recognition systems, we need to pay attention to both processing speed and recognition rate. Parallel processing hardware can satisfy both requirements at the same time. But, once the hardware architecture is determined, the maximum hardware performance is also fixed. This means that we cannot expect much faster processing speed without a means to control it.

Actually, the increase in the number of recognition modules and layers inevitably causes the increase in computation time. But, severe degradation in recognition performance is not allowed for real-time applications. Unfortunately, it is inherently difficult to detect the recognition rate for which there are no physical sensors. A method that can automatically detect recognition rate on-line and in real-time is strongly required in order to maintain robustness and accuracy of the recognition system.

4. HOW DO WE DESIGN A GESTURE RECOGNITION SYSTEM?

4.1 Too many different goals

Unlike face recognition research, there are many different goals in gesture recognition research. As shown in Figure 1, target body part itself could be head, arms, hands, torso, knee, legs, foot, etc., or combination of these parts. The target problem could be static posture analysis or dynamic motion analysis. The required functionality could be body parts or joints detection, tracking, or recognition. Moreover, some researchers may try to deal with the problems on multiple person, specific person, or unspecified person. For the reasons described above, goals and fundamental design of a gesture recognition system can vary greatly. Of course, no single research project could undertake all these problems. For the time being, researchers will have to deal with individual or different problems under different approaches until they find and share fundamental problems in this field.

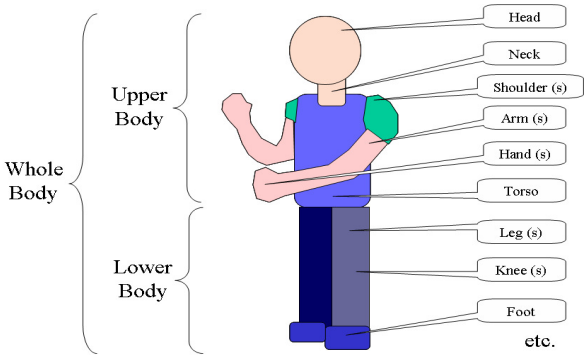

Figure 1: A human body is made up of so many parts!

4.2 Mathematical tools and recognition frameworks

There are popular mathematical tools such as HMM (Hidden Markov Models), NN (Neural Networks), DP (Dynamic Programming), and SVM (Support Vector Machines) that have often been applied to gesture recognition problems. Surely, gesture recognition is one of the application fields that mathematical tools can handle. But, is it possible to adequately choose mathematical tools without knowing each tool's advantages and disadvantages? Also, is it possible to implement state-of-the-art of each mathematical tool and to evaluate with perfect impartiality? For this, one will have to apply as many mathematical tools as possible and compare their results. Without any selection criteria for mathematical tools, it should be difficult to say which one is the best.

On the other hand, a recognition framework also can be temporary / problem-specific / comprehensive. But any recognition framework needs to reflect the nature of target problems. Interestingly, researchers tend to develop similar recognition framework when the goals or problem definitions are similar. Here, the same questions arise. Is it possible to implement state-of-the-art of recognition framework and to evaluate with perfect impartiality? For this, one will have to implement as many frameworks as possible and compare their results. Without any selection criteria for recognition frameworks, it should be difficult to say which one is the best.

Generally, researchers who are more interested in the usefulness of mathematical tools, they usually develop a sequential processing flow as shown in Figure 2, focusing on a particular mathematical tool. On the other hand, researchers who are more interested in the framework for gesture recognition, they usually develop a parallel processing flow as shown in Figure 3, focusing on a particular recognition framework. Interestingly, in either case, researchers finally notice that both the mathematical tools and the recognition framework are important and necessary.

In summary, in the design phase, it is recommendable to consider following topics before implementing the system.

(1) Problem formulation and definition (this should not be affected by the trend of the times)
(2) Design of recognition framework (this could be affected by the hardware/OS environment)
(3) Selection of mathematical tools (this will be affected by the trend of the times)

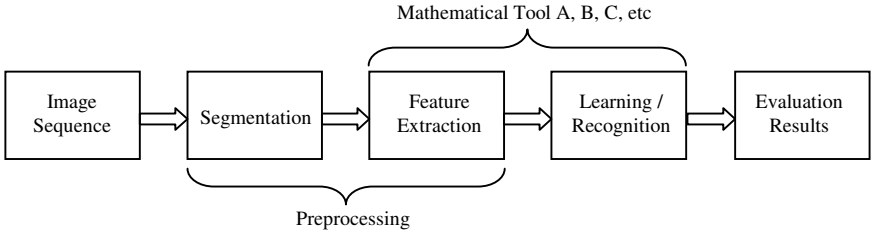

Figure 2: A sequential processing flow for gesture recognition

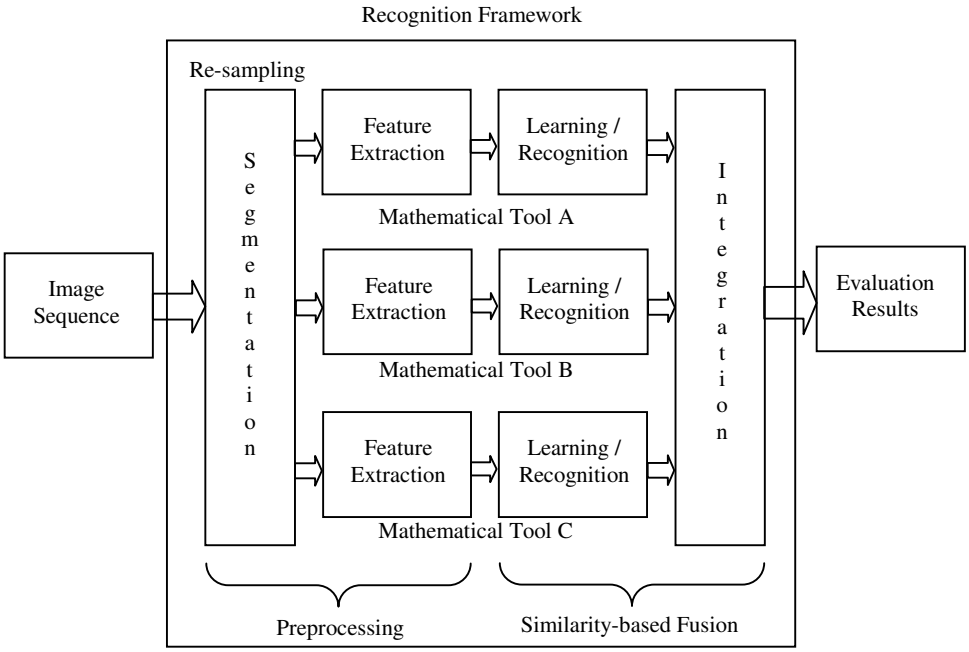

Figure 3: A parallel processing flow for gesture recognition

5. HOW DO WE EVALUATE A GESTURE RECOGNITION SYSTEM?

Although an evaluation method for gesture recognition systems deserves a research in itself, a more sophisticated and reliable evaluation framework is needed to succinctly compare the results of gesture recognition researches. Currently, following two issues are the major concerns.

5.1 Lack of standardized gesture image database

The problem of evaluation is closely related to the problem of what should be recognized. As has been mentioned in section 4.1, in gesture recognition research, there are too many variations in recognition targets. This suggests that each target may require domain-specific image databases. For this reason, the use of standardized gesture image database is not regarded as mandatory among most of the researchers in this field. But, the need for standardized gesture image database arises once common goals on certain topics are shared among researchers. For example, we can obtain standardized image databases for person identification by gait. By sharing standardized image databases, the comparison of results among different approaches can be done very easily.

But, standard gesture image databases will have to accommodate locality-specific movements and their meanings for sign language recognition problems. We know some cases in which different meanings are allocated to the same movements in different parts of the world. For locality-specific problems, locality-specific databases will have to be developed. For the reasons described above, it is difficult to expect any standardized gesture image databases that can satisfy all problem domains in this field. Probably, the old-fashioned approach of accumulating database materials at one place, at one time, and at one organization cannot provide better standard gesture image databases. Distributed databases that any researchers can easily access and contribute anytime over the Internet will be one of the solutions to this problem.

5.2 Lack of standardized evaluation procedure

By sharing a standardized evaluation procedure, it will be easier to compare improvements among different approaches. But, is it really possible to standardize an evaluation procedure? Currently, recognition rates are calculated after human judgment on the given results. The formulas and the methods to obtain recognition rates are not always the same among researchers. Moreover, we have to admit the fact that researchers cannot always follow the same evaluation procedure.

But we can be optimistic by taking another approach. Essentially, what is required is the comparison of different approaches under the same experimental framework and conditions. This will be possible by exploiting a recognition framework as shown in Figure 3. A recognition framework that can accommodate different algorithms and integrate their results will provide the basis for the comparison. Especially, full-automatic or semi-automatic evaluation procedure is important. It will help guarantee the reproducibility of the evaluation experiments. Moreover, by automating the evaluation procedure, one can establish a well-known P (Plan) - D (Do) - C (Check) - A (Action) cycle that enables the parameter-tuning of the recognition algorithm, hence, leading to a framework that can optimize the recognition performance for particular needs (see Figure 4).

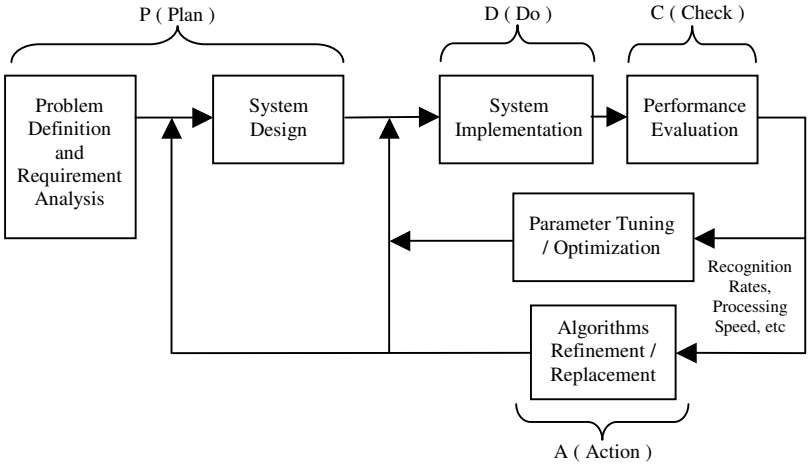

Figure 4: P-D-C-A framework for the research on gesture recognition.

6. SUMMARY

Considering the interdisciplinary nature of gesture recognition research, it is still at the beginning stage. There are too many topics to be studied. Fortunately, various kinds of mathematical tools have been actively examined and some have proven to be effective. But, apart from these mathematical tools, discovering an intrinsic problem on gesture recognition is particularly important since it requires novel approaches that will lead to the true advancement in this field. On the other hand, innovations in hardware technology have made it possible to implement gesture recognition algorithms on various kinds of hardware platforms. To obtain better recognition framework and algorithms, researchers should share new findings and problems while taking advantage of the state-of-the-art hardware and environment. In future man-machine symbiotic ubiquitous society, gesture recognition systems will be playing a prominent role since human body-mediated communication with intelligent machines is essential.

REFERENCES

[1] T. Matsuyama: “Preface”, Proc. First Int’l Workshop Man-Machine Symbiotic Systems, pp. iii-vii, Nov. 2002.

[2] R. Nakatsu: “Nonverbal Information Recognition and Its Application to Communications”, Proc. Third IEEE Int’l Conf. Automatic Face and Gesture Recognition (FG'98), pp.2-7, Apr. 1998.

[3] T. Kirishima, K. Sato, K. Chihara: “Real-Time Gesture Recognition by Learning and Selective Control of Visual Interest Points”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 27, No.3, pp. 351-364, Mar. 2005.