Comparing Sensor Based and Vision Based Techniques for Dynamic Gesture Recognition

Bayasgalan Baatar and Jiro Tanaka
Department of Computer Science, University of Tsukuba,
1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8573, Japan
baysa@iplab.cs.tsukuba.ac.jp, jiro@cs.tsukuba.ac.jp

ABSTRACT
Currently, there are mainly two types of techniques for hand gesture recognition: vision based and sensor based. Each technique has its own merits and restrictions and to decide which technique to use in our gesture application is difficult. In this paper, we describe comparison results for these two techniques for dynamic gesture recognition. We implemented a Google Earth Hand Gesture Navigation System for dynamic gesture recognition evaluation purpose. Our results show that vision based Dynamic FP (feature point) is best for dynamic gesture recognition.

Author Keywords
Dynamic gesture recognition

ACM Classification Keywords
H5.2 [Information interfaces and presentation]: User Interfaces. – Interaction Styles

General Terms
Human Factors, Experimentation

INTRODUCTION
Hand gesture is one of the natural interaction methods because hand is the most frequently used manipulation tool for human. There are mainly two types of hand gesture recognition techniques: vision based and sensor based. Vision based systems contain rich visual information which is a strong cue to infer the inner states of an object and this technique can track and recognize the hand even when it is not touching the surface or not wearing a device. At the same time, vision-based systems can be very cost efficient and noninvasive, making vision systems very feasible. However, it has some limitations of the optical sensors, the quality of the captured images is sensitive to lighting conditions and cluttered backgrounds, thus it is usually not able to detect and track the hands robustly, which largely affects the performance. Sensor based technique is another basic alternative to hand gesture recognition which is usually more reliable and are not affected by lighting conditions or cluttered backgrounds. However, it requires the user to wear a device which is inconvenient and may hinder the naturalness of hand gesture. Our research purpose is to compare these two techniques to find out which technique is better for which kind of gesture interaction

APCHI '12, August 28–31, 2012, Matsue-city, Shimane, Japan.

DYNAMIC GESTURE RECOGNITION
We made our own gesture classification which is based on basic task analysis for HCI [3]. We classified gesture into four categories: static, dynamic, dynamic gesture with posture, and object gesture. In this paper, we restricted our work to dynamic gesture recognition. Essentially, dynamic gesture recognition is the recognition of a set of user-centered motions in a single continuous flow. Acceleration based techniques [2] are commonly used as a sensor based techniques for dynamic gesture recognition. Techniques such as optical-flow [5] are used to track motion in case of vision based technique.

GOOGLE EARTH HAND GESTURE NAVIGATION SYSTEM
We developed a Google Earth Hand Gesture Navigation System to compare the sensor based and vision based techniques for dynamic gesture recognition. We selected Google Earth which is a well-known, free and ready to use application where users can navigate to visualize information.

The system recognizes six dynamic gestures; up, down, right, left, clockwise circle and anticlockwise circle. The system enables user to navigate in Google Earth through hand gesture and involves two kinds of interaction techniques; sensor based [4] and vision based. Vision based techniques has two types: dynamic FP [6][7] and static FP[8]. In sensor based technique, user holds Wiimote in his hand to perform gesture. In dynamic FP technique, the system detects user`s hand using skin color segmentation [6] which enables user to move his hand wherever his wants inside the camera view area. In case of the static FP technique, the system does not detect user`s hand. Therefore, user has to move his hand over the decided feature point area to perform gesture.

EVALUATION

We conducted two experiments to compare sensor based technique to vision based technique. The first experiment was made to compare Static FP to Dynamic FP. The second experiment was made to compare Sensor based technique to the technique which has a best result during the experiment 1.

Experiment 1

A total of 10 participants, 3 female and 6 male, joined experiment 1. First, they were asked to perform six kinds of gestures to navigate through Google Earth. During the experiment we measured two parameters; the accuracy and user preference. Table 1 shows the result of experiment 1

	Static FP	Dynamic FP
Accuracy	76%	88%
User preference	20%	80%

Table 1 Result of Experiment 1

Static FP was less accurate because it detects any motion over the feature point area which produces some unwanted gesture commands when user moves his other body parts. For example, when user wants to performs UP gestures twice, first, he has to move his hand up then bring back his hand down and move hand up again. During this process, bring back hand movement produces unwanted gesture command which affected experiment result. However, in dynamic FP technique, simply hiding the fingers stops the gesture.

Also, participants felt that static FP is less natural than dynamic FP because user has to move his hand over decided area which much destructs them from main target.

Experiment 2

Based on the result of experiment 1, we chose Dynamic FP as a candidate for the experiment 2, since 80% of the participants preferred this technique. A total of 10 participants joined the experiment 2, 5 female and 5 male. Table 2 shows the result of experiment 2

	Sensor	Dynamic FP
Accuracy	84%	91%
User preference	40%	60%

Table 2 Result of Experiment 2

Sensor based technique was less accurate because the gesture database was not rich enough. During the experiment, we observed that some people prefer to move his wrist while some people prefer to moves his elbow which produces completely different signals and affected the experiment result. Therefore, we need to learn different people`s hand movements for the same gesture and build up richer gesture database.

CONCLUSION

In this paper, we present comparison results for sensor based and vision based techniques for dynamic gesture recognition. Our results show that Dynamic FP is most accurate and natural technique for dynamic gesture recognition. During the experiment, we observed that what user really wants is to move their hand as usual and normal way as they do. Learning new movement or moving in inconvenient constant speed makes them feel so unnatural.

In each technique, the accuracy was different for each person depending on the person`s speed, hand shape or hand moving behavior etc. System improvement and more user practice are needed to achieve more realistic result.

.

REFERENCES

1. Adam Kendon, “Conducting Interaction: Patterns of behavior in focused encounters”, Cambridge University Press, Cambridge, 1990.
2. Ahmad Akl, A Novel Accelerometer-based Gesture Recognition System, University of Toronto, pp. 2-3, 2010
3. Doug Bowman, Ernst Kruijff, Joseph LaViola, Mark Mime and Ivan Poupyrev, 3D User Interface Design: Fundamental Techniques, Theory, and Practice, SIGGRAPH2000 Course #36, July 2000
4. J. Liu, L. Zhong, J. Wickramasuriya, and V. Vasudevan,“uWave: Accelerometer based personalized gesture recognition and its applications,” Pervasive and Mobile Computing, vol. 5, no. 6, pp. 657 – 675, 2009, perCom 2009.
5. Jud Porter, Mike Thomson, Adam Wahab, "Lucas-Kanade Optical Flow Accelerator", May 2011
6. Mohamed-Ikbel Boulabiar, Thomas Burger et al., “A Low-Cost Natural User Interaction Based on aCamera Hand-Gestures Recognizer”, HCI2011
7. OpenCV: Open Source Computer Vision, http://opencv.willowgarage.com/wiki/
8. Shaowei Chu, Jiro Tanaka, “Hand Gesture for Taking Self Portrait”, HCI2011, pp.4-7, 2011