Robust Hand Gesture Recognition with Kinect Sensor

Zhou Ren   Jingjing Meng   Junsong Yuan
Nanyang Technological University
50 Nanyang Avenue, Singapore 639798
{renzhou, jingjing.meng, jsyuan}@ntu.edu.sg

Zhengyou Zhang
Microsoft Research Redmond
Washington 98052-6399, USA
zhang@microsoft.com

ABSTRACT

Hand gesture based Human-Computer-Interaction (HCI) is one of the most natural and intuitive ways to communicate between people and machines, since it closely mimics how human interact with each other. In this demo, we present a hand gesture recognition system with Kinect sensor, which operates robustly in uncontrolled environments and is insensitive to hand variations and distortions. Our system consists of two major modules, namely, hand detection and gesture recognition. Different from traditional vision-based hand gesture recognition methods that use color-markers for hand detection, our system uses both the depth and color information from Kinect sensor to detect the hand shape, which ensures the robustness in cluttered environments. Besides, to guarantee its robustness to input variations or the distortions caused by the low resolution of Kinect sensor, we apply a novel shape distance metric called Finger-Earth Mover's Distance (FEMD) for hand gesture recognition. Consequently, our system operates accurately and efficiently. In this demo, we demonstrate the performance of our system in two real-life applications, arithmetic computation and rock-paper-scissors game.

Categories and Subject Descriptors

H.1.2 [**User/Machine Systems**]: Human information processing; I.4.8 [**Scene Analysis**]: Depth cues, Shape

General Terms

Experimentation

Keywords

Human-Computer-Interaction, Hand Gesture Recognition, Kinect Sensor, Finger-Earth Mover's Distance

1. INTRODUCTION

Hand gesture recognition is an important research issue in the field of Human-Computer-Interaction, because of its extensive applications in virtual reality, sign language recognition, and computer games [7]. Despite lots of previous work, building a robust hand gesture recognition system that is applicable for real-life applications remains a challenging problem. Existing vision-based approaches [1, 5, 6] are greatly limited by the quality of the input image from optical cameras. Variations in lighting and background clutters would only worsen the problem. Consequently, these systems have not been able to provide satisfactory results for hand gesture recognition.

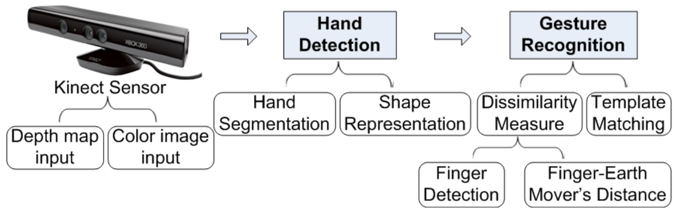

Figure 1: The framework of our robust hand gesture recognition system.

Hand gesture recognition concerns two challenging problems: *hand detection* and *gesture recognition*, namely how to robustly detect the hand and how to efficiently and accurately recognize the gesture of the hand.

The framework of our hand gesture recognition system is shown in Fig.1. As for hand detection, different from traditional methods that use color-markers for hand detection, our system uses both the depth map and color image obtained by Kinect sensor to detect the hand shapes. It ensures robustness to cluttered background. And the segmented hand shapes are represented as time-series curves [2].

As for the gesture recognition, even with the Kinect sensor, it is still a very challenging problem. Because typically, the resolution of a Kinect sensor is only 640×480. Although it works well to track a large object, e.g. the human body, it is difficult to detect and segment precisely a small object from an image at this resolution, e.g., a human hand that occupies a very small portion of the image. Therefore, we use a novel shape distance metric called Finger-Earth Mover's Distance (FEMD) to measure the dissimilarities between different hand shapes, which is proposed in [4]. FEMD metric is specifically designed for hand shape matching, which is robust to hand variations and distortions. FEMD represents a hand shape as a signature which considers each finger as a cluster, and the dissimilarity distance between two hand shapes is defined as the sum of the work needed to move the earth piles and the penalty on the unmatched fingers. In order to accurately detect the fingers, [4] also presents two novel finger detection algorithms, using thresholding decomposition and using a near-convex shape decomposition scheme presented in [3]. Finally, the input hand is recognized by template matching.

MM'11, November 28–December 1, 2011, Scottsdale, Arizona, USA.
ACM 978-1-4503-0616-4/11/11.

2. PERFORMANCE

The goal of this demo is to showcase two real-life applications built on top of our novel and robust hand gesture recognition system. Our technology performs robustly despite variations in hand orientation, scale or articulation. Moreover, it works well in uncontrolled environments with background clutters. As for the efficiency and accuracy, our system operates efficiently with a mean accuracy of 90.6% on the dataset of [4].

3. APPLICATIONS

We demonstrate our hand gesture recognition system using two real-life applications: arithmetic computation and rock-paper-scissors game.

3.1 Arithmetic computation

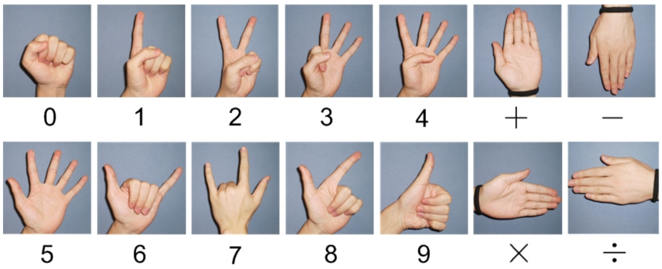

Figure 2: The 14 gesture commands in our arithmetic computation system.

Arithmetic computation is an interesting HCI application. Instead of interacting with the computer using the keyboard or mouse, we input arithmetic commands to the computer via hand gestures. As shown in Fig.2, 14 hand gestures are employed to represent 14 commands, namely number $0-9$ and operator $+$, $-$, $\times$, $\div$, respectively.

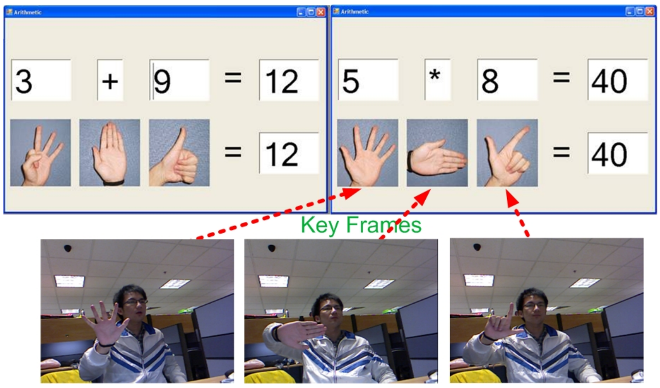

Figure 3: Arithmetic computation.

By recognizing each input gesture as a command, the computer can perform arithmetic computations instructed by the user. Two examples are shown in Fig.3. The key frames are shown as well.

3.2 Rock-paper-scissors game

Rock-paper-scissors is a traditional game. The rule is rock breaks scissors; scissors cut paper; and paper wraps rock. In this demo, we build a Rock-paper-scissors game system played between a human and a computer. Three hand gestures are defined as 3 different weapons in the game, as shown in Fig.4, which can be recognized by our system, and

Figure 4: Rock-paper-scissors game.

the computer just randomly chooses a weapon. Then, according to the game rule, our system can decide the winner between human and computer. Fig.5 shows two examples.

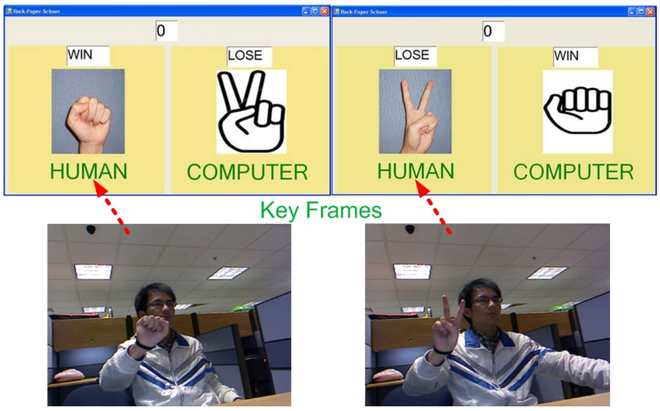

Figure 5: Rock-paper-scissors game.

4. CONCLUSION

In this demo, we present an efficient and accurate hand gesture recognition system using Kinect sensor as the input device. Both the depth and color information obtained from Kinect sensor are used for hand detection, which ensures the robustness of our system to cluttered environments. Besides, the Finger-Earth Mover's Distance metric employed in our gesture recognition module provides an effective mechanism for recognizing hand shapes with input variations and distortions. Such a hand gesture recognition system provides a robust solution in two real-life HCI applications, which can also be applied to many other hand gesture based HCIs.

5. ACKNOWLEDGEMENT

This work was supported by the Nanyang Assistant Professorship (SUG M58040015) to Dr. Junsong Yuan.

6. REFERENCES

[1] C. Chua, H. Guan, and Y. Ho. Model-based 3d hand posture estimation from a single 2d image. *Image and Vision Computing*, 20:191 – 202, 2002.

[2] E. Keogh, L. Wei, X. Xi, S. Lee, and M. Vlachos. Lb_keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures. In *Proc. of 32th International Conf. on VLDB*, 2006.

[3] Z. Ren, J. Yuan, C. Li, and W. Liu. Minimum near-convex decomposition for robust shape representation. In *Proc. of ICCV*, 2011.

[4] Z. Ren, J. Yuan, and Z. Zhang. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *Proc. of ACM MM*, 2011.

[5] N. Shimada, Y. Shirai, Y. Kuno, and J. Miura. Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints. In *Proc. of Third IEEE International Conf. on Face and Gesture Recognition*, 1998.

[6] B. Stenger, A. Thayananthan, P. Torr, and R. Cipolla. Filtering using a tree-based estimator. In *Proc. of IEEE ICCV*, 2003.

[7] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Communications of the ACM*, 54:60–71, 2011.