Gesture-World Technology

- 3D hand pose estimation system for unspecified users using a compact high-speed camera -

Kiyoshi Hoshino † Motomasa Tomida Takanobu Tanimoto (University of Tsukuba)

1 Introduction

This technology allows people to control devices such as computers, communications devices, household appliances, and robots by means of everyday gestures without using sensors or controllers, which employs the high-speed and high-accuracy computer vision technology capable of estimating the human hand and arm poses captured by a compact high-speed camera.

The technology behind our exhibit will change standard practices with regard to the mechanical operation of equipment such as information devices and robots. Because fingers are articulated structures, they can assume many complex shapes, which often result in the problem of self-occlusion. Despite their small size relative to the rest of the body, fingers are also capable of moving in a wide 3D space, due to movements of the arms or the body. For these reasons, it has not been easy to estimate hand poses by non-contact means, using a monocular camera or a pair of cameras at close-range. In recent years, however, high-speed cameras have become more compact and inexpensive. Therefore, if we can achieve fast and accurate 3D hand pose estimation, using only camera images (in other words, without sensors) and without the need to strictly fix the camera position, we expect that the technology will be applied in a wide range of areas. Examples of such applications could include gesture-based computer operation, virtual games, remote control without a remote controller, digital archiving of artisan skills, and remote robot control. The need to attach sensors or find and use special controllers will disappear.

Our system focuses on achieving highly accurate hand pose estimation for unspecified users by constructing an enormous database including bone thickness and length, joint range of motion, and habitual finger movements, by thoroughly reducing the dimensionality of the image features in the data set used for comparison with the input hand images. If the image features that express each hand pose were of extremely low dimensionality, it would be possible to prepare a database that includes differences among people. We reduce the dimensionality to 64 or less, or $1/25^{th}$ of the original image features. With an image feature reduced to $1/25^{th}$, it is clearly possible to upload data for 25 times as many hand poses to the memory.

In the Conference, we will show you a virtual clay art system, with which you can form 3D models by moving your hand and arm as desired, and/or a remote control system of a robot, with which you can control a manipulator by your hand and arm movement similar to movements in your dairy living.

2 System configuration

Using CG editing software, it is relatively easy to generate an enormous hand pose database [1]. We have a data set of about 30,000 features [2] thus fur, and if we increase the size of the database by 25 times, to 750,000 features, by thoroughly reducing the dimensionality of the image features in the data sets, we will still be able to upload it to a memory. By varying parameters such as bone thickness and length and finger bending habits (in other words, individual differences in joint range of motion and angles of bending within that range), we will generate a hugely varied and diverse collection of hand poses.

Using the above method, we have achieved accurate hand pose estimation even where there are individual differences. The next step is to increase the speed. Our team has previously proposed a classification of features to allow for narrowing a search in terms of wrist rotation, thumb bending, and four fingers bending [3]. We can use this classification of features to perform a first-stage filtering, to be followed by a second stage of precise similarity matching. This 2-stage searching will make it possible to find the most similar pose from a database of roughly one million features at about 150 fps.

References

[1] K.Hoshino and T.Tanimoto: Real time search for similar hand images from database for robotic hand control, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, E88-A, 10, pp.2514-2520, 2005.

[2] K.Hoshino and M.Tomida: Copycat arm, SIGGRAPH 2008, New Tech Demos, Full Conference DVD-ROM, 2164, p.1, 2008.

[3] K.Hoshino and M.Tomida: 3D hand pose estimation using a single camera for unspecified users, Journal of Robotics and Mechatronics, 21, 6, pp.749-757, 2009.

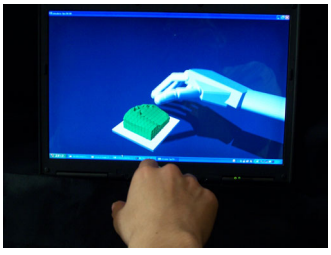

Fig. 1: Virtual clay art system.

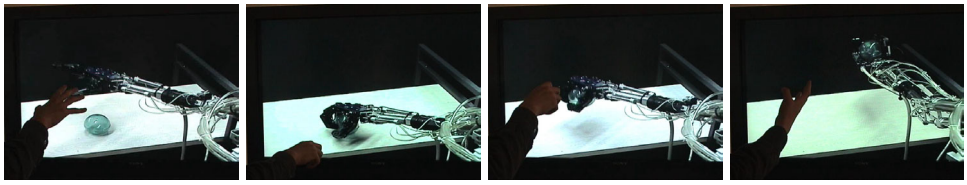

Fig. 2: Remote control system of a robot. The user watches a monitor and decides how to move the robot by gesture.

† hoshino@esys.tsukuba.ac.jp
http://www.kz.tsukuba.ac.jp/~hoshino/

SIGGRAPH 2010, Los Angeles, California, July 25 – 29, 2010.
ISBN 978-1-4503-0210-4/10/0007