Applying Hand Gesture Recognition with Time-of-Flight Camera for 3D Medical Data Analysis

Filip MALAWSKI

AGH University of Science and Technology
Faculty of Computer Science, Electronics and Telecommunication
Department of Computer Science
al. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: fmal@agh.edu.pl

Abstract. This paper describes a human-computer interface based on hand gesture recognition, designed for analysis of 3D medical data. Hand gestures are recognized by employing Shape Context descriptor and SVM classifier on depth maps acquired with a time-of-flight sensor.

Keywords: human-computer interaction, hand gesture recognition, time-of-flight camera

Introduction

Hand gesture recognition is an area of extensive research. Touchless human-computer interfaces have been drawing attention of both scientific and business communities for decades. Hand gestures can be employed as an alternative to the standard mouse interface. In particular, 3D applications, such as games, design tools or data analysis applications, may benefit the most, by utilizing the fact that our hands are able to express a wide range of gestures and can operate in 3D space, contrary to the mouse. Moreover, using hands is the most natural way of interaction with different things.

Multiple solutions for hand gesture recognition have been presented over the years. Initial solutions have been based on RGB cameras, e.g. [1]. However recently, depth sensors, such as Kinect or time-of-flight (ToF) cameras, have gained considerable popularity in this area. Depth cameras are robust to illumination changes and greatly facilitate extraction of relevant data in the image, since foreground/background segmentation can be performed easily with the available depth data. Several solutions based on depth cameras have been proposed, utilizing i.a. HOG descriptors [2], contours[3] and Earth Mover's Distance descriptors [4].

Despite potential advantages and numerous recognition methods, human-computer interfaces based on gestures have not actually gained much popularity. Performing the gestures for a prolonged time has proved to be too tiring and inconvenient. In this paper a hand gesture based interface is presented, which emphasizes the ease of use, rather than wide range of gestures. The user is required to keep the hands in such a position, that they do not overlap. When the hands are seen by the depth sensor as one continuous object this in fact constitutes additional gesture, which indicates that the interaction should be paused. Therefore when the user has a need to rest the hands, he needs only to join them and lay on the table.

Methods

The methods employed in this work are discussed in 3 subsections describing a) design of gestures b) gestures recognition method c) 3D medical data visualization.
**Gesture recognition**

CamBoard Nano ToF sensor has been employed. It provides, along with the depth map, a set of flags for each pixel. Valid pixels have no flags set, while invalid ones have flags describing for what reason is the pixel invalid - e.g. too much noise. For the purpose of gesture recognition a database of the 4 gestures has been acquired. Each gesture was recorded 50 times, resulting in total of 200 images. The gestures have been acquired for one hand only, since gestures of the other hand can be recognized by applying symmetry.

Processing of a single frame is a multi-stage action. In the first stage, the closest pixel is found and all pixels which are more than 30cm further are marked as invalid. This provides a simple, yet effective background/foreground segmentation. Next step is the noise reduction. Although the sensor detects most noise and marks the corresponding pixels as invalid, additional noise reduction is required. For all pixels, the number of invalid pixels in their 3x3 neighborhood is computed and if there are at least 6 invalid pixels in the neighborhood, the pixel is marked as invalid as well.

Subsequently segmentation to left and right hand is performed. The algorithm searches for a vertical line of invalid pixels, starting in the middle of the image and moving towards the edge, further in each iteration, in both directions simultaneously. If no such line is found between 1/4 and 3/4 of image width, the image is classified as the resting position. Otherwise, the image is divided to left and right hand images. Next, valid pixels which have at least one invalid adjoining pixel are marked as contours.

Gesture recognition employs the Shape Context descriptor, which creates a histogram of relative positions of the pixels, based on the contours. For each pixel, length and direction of vectors connecting the pixel with all other pixels are computed. These values are then assigned to bins in log-polar space, therefore creating a 2D histogram. The histogram can be linearized and used as a vector of features describing the gesture. For classification of the gestures several classifiers have been tested - Naive Bayes, Random Forest and Support Vector Machine (SVM) with radial basis function (RBF) kernel, the last one giving the best accuracy (95%). Classifiers implementation was provided by the WEKA package.

Finally, position of each hand is computed as a mean position of all valid pixels in the segment relevant for this hand. However, due to the noise and the fact that the hand is usually not completely steady, this can introduce undesired 'shaking' of the controlled object or property. In order to compensate for this problem, mean hand position from last 5 frames is utilized.

**Visualization**

For the visualization the VTK framework has been employed. It provides various visualization methods commonly used for 3D medical data. In the proposed solution two methods have been utilized - isosurface and cut-planes. The first one creates a 3D surface of points with the same value. Both opacity and the isovalue can be controlled by the user. In the second method 3 cut-planes are displayed, each in different axis. Their position can be controlled as well as level and window parameters of the data in all cut-planes. Both isosurface and cut-planes are presented in fig. 3.

Mapping of the gestures to the controls is as follows. Left hand is used for general image transformations. Open fingers indicate no interaction. Closed fingers with thumb up are used for rotation control, while closed fingers with thumb down indicate translation control. Both modes employ x and y axes for control of the transformations and z axis for the zooming. Right hand is used in similar manner for control of different properties, such as opacity, isovalue, cut-planes position and level and window values. Closed fingers with thumb up indicate changing property value and closed thumb is used for switching between consecutive properties.

**Results and Conclusions**

A human-computer interface based on hand gesture recognition has been devised. Best achieved recognition accuracy is 95%, with the SVM classifier. The recognition method runs in real-time on a typical PC machine. Proposed solution allows for convenient analysis of 3D medical data and addresses the problem of muscle fatigue by utilizing a relatively comfortable hands position.

**BIBLIOGRAPHY**


**Author:** MSc. Eng. Filip Malawski - is a PhD candidate at AGH University of Science and Technology in the Department of Computer Science. He received his MSc degree in computer science at AGH University in 2012. His research interests include computer vision, image processing and human-computer interaction.