Robust Real-Time 3D Time-of-Flight Based Gesture Navigation

Jochen Penne Stefan Soutschek Lukas Fedorowicz Joachim Hornegger

Chair of Pattern Recognition
University Erlangen-Nuremberg
{Jochen.Penne,Stefan.Soutschek,Joachim.Hornegger}@informatik.uni-erlangen.de

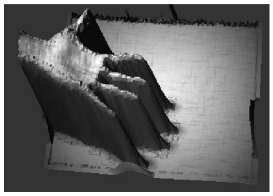
Abstract

Contactless Human-Machine-Interfaces (HMIs) are an important issue in various applications where a haptic interaction with an input device is not possible or not appropriate. Newly developed Time-of-Flight cameras provide 3D information of the observed scene in real-time at constant lateral resolutions of thousands of pixels. Additionally, a gray-value image of the observed scene is available. Our work compromises three major contributions: First, the robust and real-time capable segmentation of the hand by incorporating 3D and gray-value information; Second, the reliable classification of the performed static gesture using robust features; Third, the design of an HMI which uses the classified gesture as well as the 3D position of the hand to enable complex and convenient user interactions. The benefit of using a ToF camera is that the 3D information is just not available from classical 2D camera systems and thus only with ToF cameras the three dimensions of freedom which are given for non-haptic interactions can be fully used. Currently, classification rates of 98.2% are achieved user-dependent and 94.3% user-independent for 6 gestures. Tests with untrained persons yielded a good to very good acceptance of the HMI.

1. Introduction

TOF technology enables the direct acquisition of the distance information about a world point which is projected on a sensor element [1]. Currently, framerates $\geq$15 fps are achieved by TOF cameras.
TOF cameras actively illuminate the scene with an optical reference signal. By Smart Pixels, which are integrated into the TOF camera chip (TOF chip), the reflected optical wave is analyzed and for each pixel the phase shift compared to the reference signal is estimated. Assuming a constant speed for the spread of the signal the phase shift is directly proportional to the distance of a point in the recorded scene.

(a) ToF camera MESA Imaging GmbH [3].



(b) 3D surface reconstruction of a human hand.

Lateral Resolution [px]	144×176
Depth Resolution [mm]	$\geq$1
Framerate [fps]	$\geq$15
3D points	25344
Field of view [degree]	47.5×39.6

(c) Technical specification of ToF camera SwissRanger 3100.

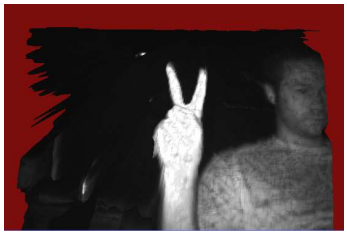
Figure 1. ToF camera SwissRanger 3100 from MESA Imaging (Figure 1(a) and example visualization of available data (Figure 1(b)). Table 1(c) shows the technical specification of the ToF camera used for in our work.

Currently, lateral resolutions of up to 144×176 pixel and z-resolutions of 1 mm are available [3]. Simultaneously to the phase delay, the amplitude of the reflected optical wave is estimated. The amplitude value thus indicates the quality of distance measurement. This information provides a gray-scale image of the scene, with the reflectivity of the material being encoded in the gray-values.

2. Methods and Results

Figure 1 gives an overview of the data and technical details of the ToF camera used in our work. The segmentation of the hand includes an initial segmentation by thresholding the distances and the amplitude information, i.e. the extraction of 3D points within a certain distance to the camera and with a sufficient high quality of distance measurement. As these data may still contain the forearm an additional

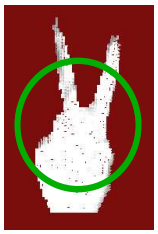
978-1-4244-2154-1/08/$25.00 ©2008 IEEE



(a) Original 3D input data.

(b) Segmented 3D hand.

Figure 2. Original 3D input data (Figure 2(a)) and segmented hand after initial segmentation and removal of forearm (Figure2(b)).



(a) Segmented hand with fitted circle.

(b) Extracted feature vector.

Figure 3. Feature extraction: After segmentation a circle is fitted into the hand (Figure 3(a)) and this circle is sampled in steps of one degree (Figure 3(b)). The feature vector contains the values which have been assigned to each object pixel by a distance transformation [2].

segmentation step including a distance transformation [2], which assigns every object pixel the smallest distance to a non-object pixel, is applied. Original 3D input data and the resulting final segmentation are shown in Figure 2. For the purpose of feature extraction a circle is fitted into the segmented hand. The circle is than sampled in steps of one degree. The resulting feature vector contains the result of the distance transformation of the according point if it is on the hand and zero otherwise. A straight-forward nearest neighbor classifier was chosen for classification. An intuitive gesture set of 6 gestures was designed with the goal to control either the mouse cursor of a computer by gestures or alternatively to rotate/translate a medical volume dataset (e.g. CT,MR) and select points/volumes of interest in the dataset. The gesture functionalities include ”Mouse cursor movement”, ”Point selection/Mouse click”, ”Reset/Back”, ”Rotating a 3D dataset”, ”Translation of a 3D gesture set”. Thereby, the currently classified gesture triggers the functionality and the parameterization of the specific function is derived from the 3D pose information of the hand (e.g. position of the mouse cursor, length and direction of translation of a 3D dataset).

3000 gesture datasets from 15 different persons were acquired. User-dependent, i.e. training and test gestures came from the same person, a classification rate of 98.2% was achieved (ten-fold cross-validation). User-independently, i.e. training and test gestures were not from the same person, a classification rate of 94.3% was achieved. The algorithm operates at 13fps on a 2GHz desktop computer with 2GB RAM. As a rotation of the hand only yields a linear shift in the feature vector, the feature extraction can be accomplished very robust despite rotation of the hand just by realigning the feature vectors. The 15 test persons of the experimental evaluation were asked four questions after having performed the gestures. The questions focused on the intuitiveness, the usability and the response time of the gesture navigation. The overall rating was a good to very good. Thus, we conclude on the practicability of the gesture based HMI.

3. Demonstrator

Our demonstrator consists of a ToF camera which is standing on a table. The user sits in front of the camera (distance approx. 70-100cm) and performs the gestures. An appealing visualization of all processing steps will be available. This will give an outstanding impression of the robustness, speed and stability of segmentation, feature extraction and classification. Then the user can exploit the whole capabilities of the HMI by controlling the computer via gestures or navigating and exploring a 3D medical data set.

4. Acknowledgments

The authors would like to thank Mr. Thierry Oggier from MESA Imaging AG for its valuable support and the fruitful discussion.

References

[1] B. Büttgen, T. Oggier, M. Lehmann, R. Kaufmann, and F. Lustenberger. CCD/CMOS Lock-In Pixel for Range Imaging: Challenges, Limitations and State-of-the-Art. In *1st Range Imaging Day*, Zürich, Switzerland, June 2005.

[2] B. Deimel. Entwicklung eines stabilen videobasierten Verfahrens zur Segmentierung der Hand am Unterarm. Diplomarbeit, Department for Graphical Systems, University of Dortmund/Germany, January 1998.

[3] MESA Imaging AG. www.swissranger.ch, 2007.