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

08.09.2008

Jochen Penne

Chair of Pattern Recognition (Informatik 5)

Friedrich-Alexander-University Erlangen-Nuremberg

Content

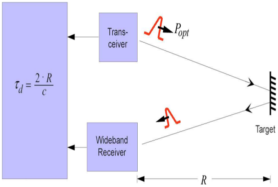
Introduction

ToF (Time-of-Flight) technology

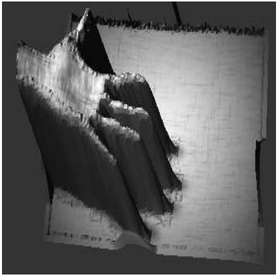
Technical specifications

- Framerate: 12...50 fps
- Resolution: 16×64 pixel ... 144×176 pixel
- Depth resolution per pixel: 1-6 mm
- Output: **3-D coordinates**, **gray value image** (encoding the amplitude of the reflected signal)

(a) TOF camera MESA Imaging GmbH



(b) TOF principle (distance R, speed of light c, travel time of impulse P_{opt} is τ)



(c) Reconstruction of human hand

Introduction

Time-of-Flight and Human-Machine-Interaction

Beneficial	Limiting
+ Markerless measurement principle	- Costs of ToF camera system[1]
+ Active illumination	- Small field-of-view ($\approx 40^\circ$)[2]
+ Sufficient frame rate	- Pixel resolution[3]
+ Constant resolution	

Table: Beneficial and limiting aspects of ToF data acquisition for touch-less Human-Machine- Interaction. [1]2006: >5000 Euro; 2008: 1400 Euro; January 2009: 300 Euro. [2]Increasing the field-of-view also implies increasing the actively illuminated part of the scene. [3]Not in the range of standard CCD cameras (640×480px. or higher).

Appropriate application fields:

- Intra-operative medical applications
- Industrial robot control
- Automotive applications
- Gaming industry (?)

Human-Machine-Interaction(?)

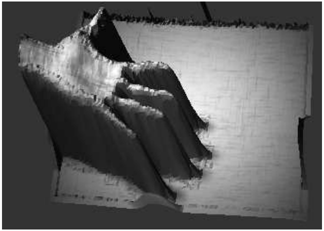
Introduction

Scope of the presented work

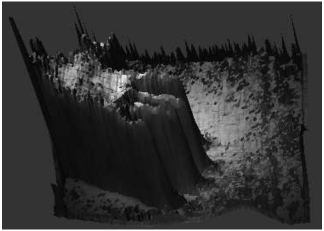
The presented work aimed at giving a proof of concept by successfully addressing the issues:

- How can the hand (and only the hand) reliably be identified?
- How can meaningful features be extracted to enable a gesture classification?
- What gestures shall be used?
- How do users response to the system?

Example: Illustrating the problem of segmentation and feature extraction



A hand can look like this...

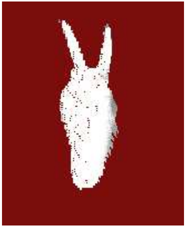
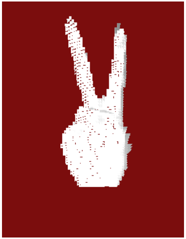


... or like this.

Gesture Recognition and Classification

Segmentation

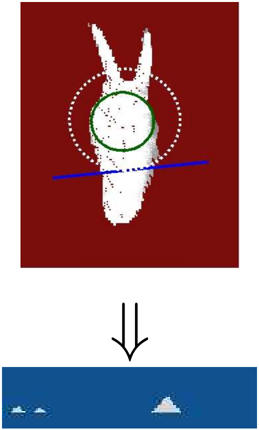
Steps from original 3D and amplitude data to fine segmentation of the hand:

 ⟹  ⟹ 

Original 3D+amplitude data ⟹ Coarse segmentation (distance threshold) ⟹ Fine segmentation (distance transformation)

Gesture Recognition and Classification

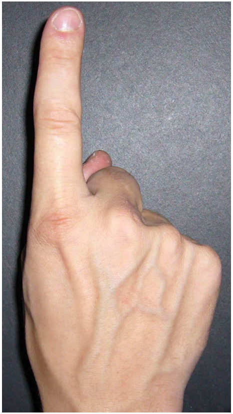
Feature Extraction



- Distance transformation yields center pixel of hand and radius of circle completely located in the palm (green circle).
- Increasing the radius by a factor of 1.68 yields the *sampling circle* (white dotted circle).
- Sampling the *sampling circle* in steps of one degree yields the feature vector which contains the distance transformation values of pixels lying on the sampling circle.
- These feature vectors can easily be aligned via the maximum value (always on the intersection of hand and forearm) by cyclic shifts $\Longrightarrow$ rotation-invariant feature vector.

Human-Machine Interface

Gesture Set

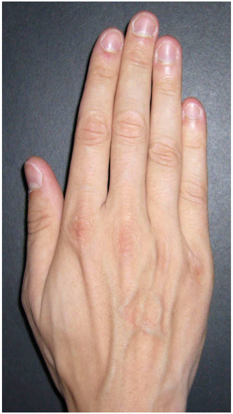


Mouse Cursor (Movement)

Mouse Cursor (Click)

Human-Machine Interface

Gesture Set (cont.)

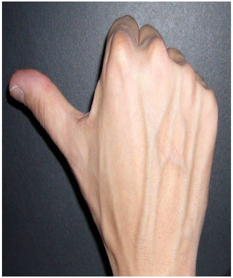


3D Rotation

3D Translation

Human-Machine Interface

Gesture Set (cont.)



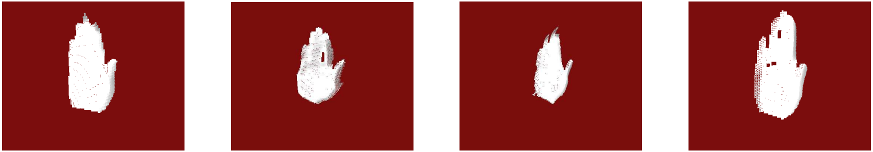
Reset

Evaluation

Data used for evaluation

- 15 persons; 40 reference datasets of each of the 5 gestures $\Longrightarrow$ 3000 datasets
- Brief explanation of the gestures
- No restrictions regarding position of sleeves or wearing a ring
- Only requirement: Hand is in the field-of-view of the camera and within the working range (<1.2 meter)

Example data: 4 different persons perform the *3D rotation* gesture



Evaluation

User-dependent classification rates

Nearest Neighbor	**Number of equidistant samples extracted from sampling circle**			
Person	**90**	**120**	**180**	**360**
P1	94.0 %	95.0 %	96.0 %	95.0 %
P2	99.0 %	98.5 %	99.0 %	98.5 %
P3	99.0 %	99.0 %	99.0 %	99.0 %
P4	99.0 %	99.0 %	99.5 %	99.5 %
P5	97.0 %	97.0 %	98.0 %	97.5 %
P6	99.5 %	99.5 %	99.5 %	99.5 %
P7	93.5 %	93.0 %	93.0 %	92.5 %
P8	98.5 %	98.5 %	99.0 %	98.5 %
P9	99.0 %	99.5 %	99.5 %	99.0 %
P10	98.5 %	99.0 %	99.0 %	98.5 %
P11	98.5 %	98.5 %	98.0 %	98.0 %
P12	99.0 %	99.5 %	99.5 %	99.5 %
P13	99.5 %	99.5 %	99.5 %	99.5 %
P14	99.0 %	99.0 %	98.5 %	98.5 %
P15	99.5 %	99.5 %	99.5 %	99.5 %
Mean	**97.6 %**	**98.3 %**	**98.4 %**	**98.2 %**

Table: User-dependent classification rates.

Evaluation

User-dependent classification rates

- The user-dependent evaluation used reference and test gestures of the same person.
- For the user independent evaluation 45 data sets for each gesture (3 data sets from each of the 15 persons) were used as reference data sets.

	Number of equidistant samples extracted from sampling circle			
	90	**120**	**180**	**360**
Nearest Neighbor	80.9 %	81.3 %	81.8 %	83.6 %

Table: User-independent classification rates.

Evaluation

Intraoperative usability - Preliminary results

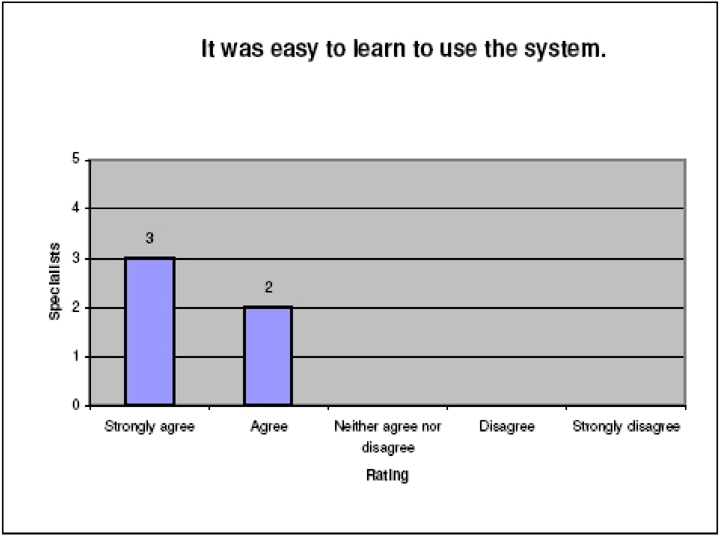
- 5 surgeons (2 female, 3 male) utilized the presented Human-Machine-Interface to manipulate 3D CT volume data sets (using InSpace volume renderer)
- Step 1: Brief explanation of the gesture data set
- Step 2: Manipulation of volume data set
- Step 3: Questionnaire

Setup for evaluation:

- Right laptop: Gesture classification
- Middle: ToF camera
- Left laptop: Volume renderer where actions triggered by gestures are performed

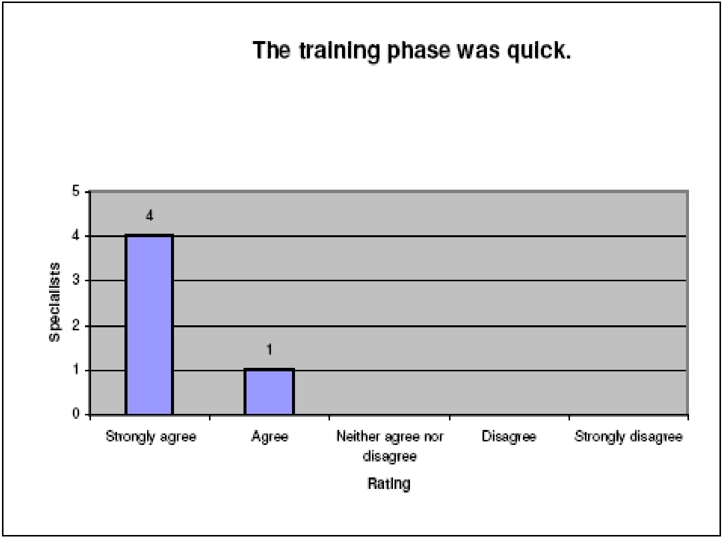
Evaluation

Intraoperative Usability - Preliminary Results (cont.)



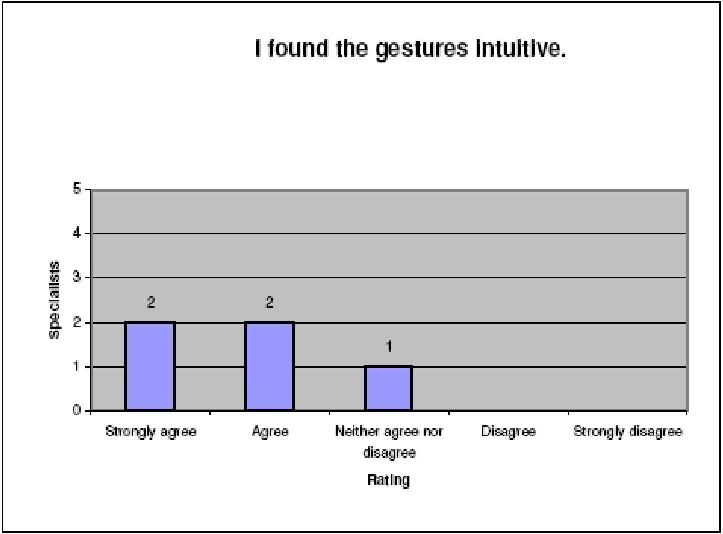
Evaluation

Intraoperative Usability - Preliminary Results (cont.)



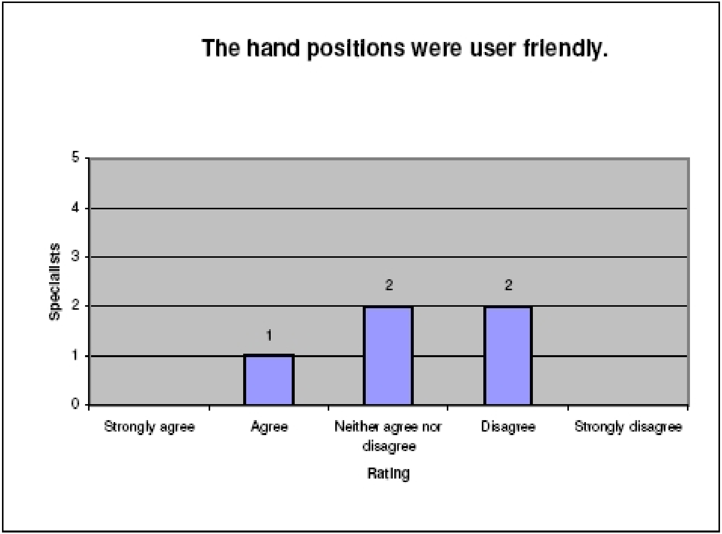
Evaluation

Intraoperative Usability - Preliminary Results (cont.)



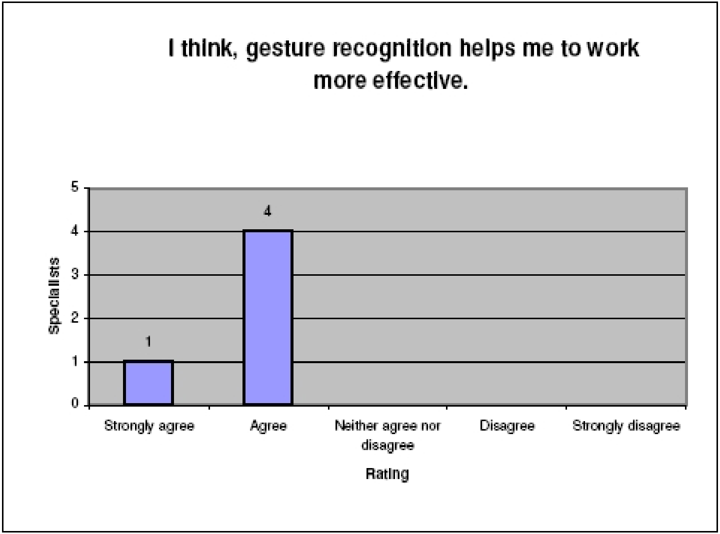
Evaluation

Intraoperative Usability - Preliminary Results (cont.)



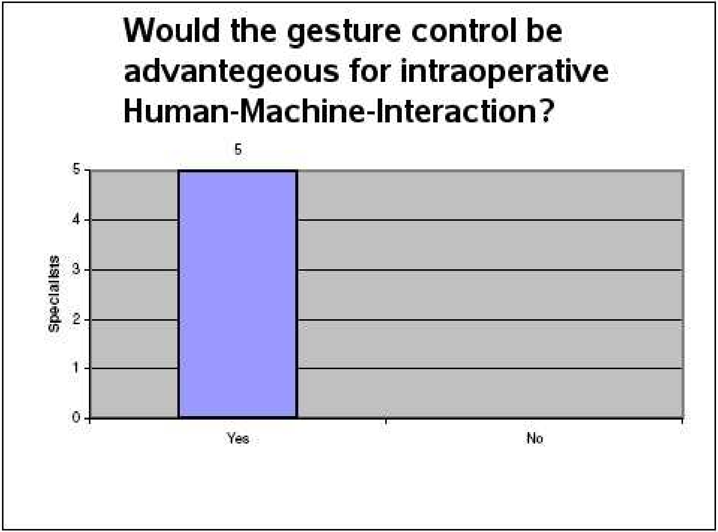
Evaluation

Intraoperative Usability - Preliminary Results (cont.)



Evaluation

Intraoperative Usability - Preliminary Results (cont.)



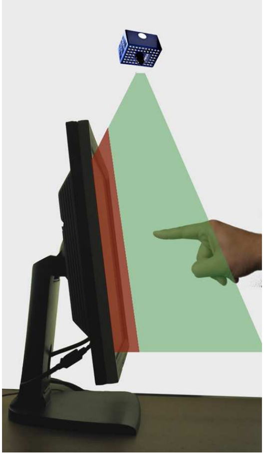
Evaluation

Conclusion

- In terms of algorithms the ToF-based 3D gesture recognition provides sufficiently high recognition rates at least when user-specific training data sets are used.
- More advanced classification approaches (Principal Component Analysis) will be suitable and lead to better classification rates for user-independent training data sets, too.
- The evaluation by the 5 surgeons indicates that ToF-based 3D gesture interaction will be beneficial for intraoperative applications, but: The hardware setup definitely has to be changed to enable a more suitable and easy interaction (hand position was rated very inconvenient).

Outlook

Alternative hardware setup



- Mount the camera above the display.
- The intersection of pointing finger and display defines the mouse position.
- Other gestures trigger certain actions like manipulation of volume data sets or marking regions in an image.

The End

- Thanks to Lukas Fedorowicz: Parts of the presented work have been accomplished by him during his diploma thesis at the Chair of Pattern Recognition.
- Thanks to the MITI group (workgroup for minimally invasive therapy and intervention) of Prof. Hubertus Feußner at the Klinikum rechts der Isar: Members of the MITI group participated in the evaluation of the 3D gesture control system.

Thank you for your attention.