Human Machine Interface for Elderly People

Eva Kollorz[1], Jochen Penne[1], Joachim Hornegger[1], Johannes Kornhuber[2]

[1] Institute of Pattern Recognition, Friedrich-Alexander-University of Erlangen-Nuremberg, Martensstr. 3, D-91058 Erlangen, Germany
{Eva.Kollorz, Jochen.Penne, Joachim.Hornegger}@informatik.uni-erlangen.de
[2] Department of Psychiatry and Psychotherapy, University Hospital Erlangen, Schwabachanlage 6, D-91054 Erlangen, Germany
johannes.kornhuber@uk-erlangen.de

Motivation

Many (elderly) people have problems with:

- menu prompting of electronic devices
- feel of electronical devices like remote-control

$\Rightarrow$ Intuitive handling interface is required
$\Rightarrow$ Gesture vocabulary which is easy to keep in mind

Gesture controlled interface:

- Works touchless $\rightarrow$ No small buttons have to be pressed
- User-independent $\rightarrow$ User does not have to train the system

Time-of-flight (TOF) camera:

- Additional to the gray value image the depth information is provided
- Distance data:
  1. Simple object segmentation possible
  2. Additional information for classification
- Independence of illumination

Aim: Reliable real-time recognition of intuitive gestures for control of electronic equipment [1]

Methods

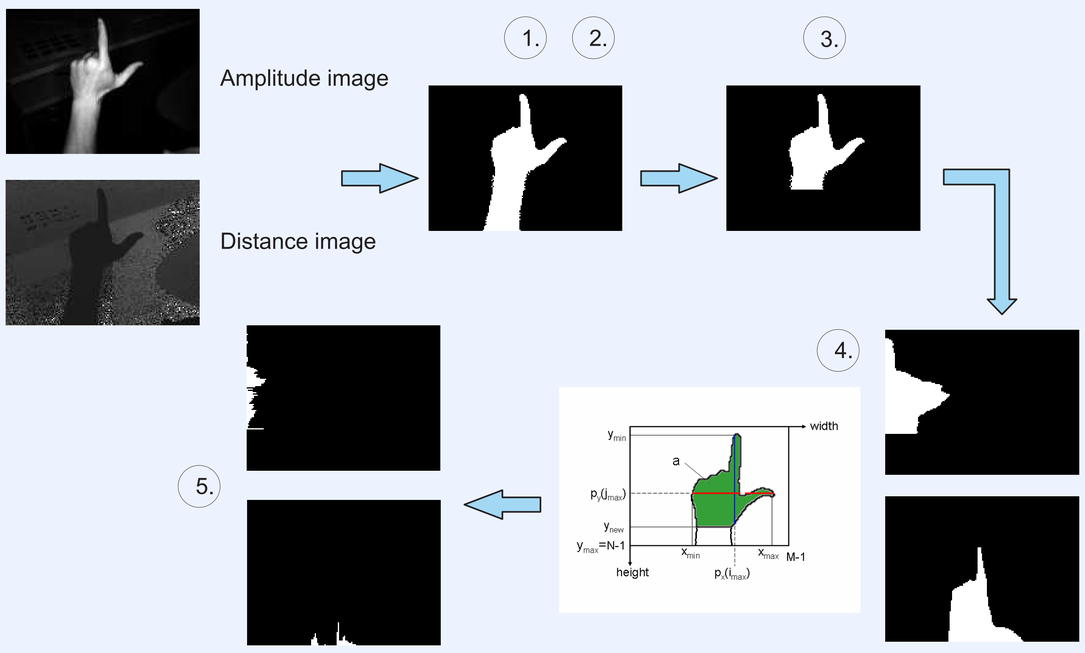

Figure 1: Overview of the introduced algorithm: segmentation of the hand (1.), determination of the bounding box (2.), extraction of the hand (3.), calculation of the depth features of the segmented hand without contour (4.), then projection onto the image axes (4.) and classification (5.).

The proposed algorithm can be separated into five steps:

1. Segmentation of the hand and arm via distance values:
   - Iterative seed fill algorithm
2. Determination of the bounding box:
   - Projection onto the image axes [2]
   - Entry point of the arm
3. Extraction of the hand:
   - Calculation of the initial cut-off position $y_{new} = y_{min} + \frac{12.0 \cdot l}{(cog_z \cdot 0.04)}$,
     l: "virtual" hand length [mm], cog_z: distance at centre of gravity [mm]
   - Adjustment (small bottleneck at the wrist)
4. Projection of the hand region onto x- and y-axis:
   - Projection onto the image axes [2]
   - Calculation of the additional depth features $a_{min}, a_{avg}, a_{max}$
5. Classification with majority decision over k nearest neighbour and m frames:
   - Matching of the actual gesture's bin i, j to the reference (R) gesture's bin c_i, c_j
   - Classification distance $d = \left| \frac{a_{avg} - a_{min}}{a_{max}} - \frac{a_{R,avg} - a_{R,min}}{a_{R,max}} \right| + \frac{x_{\text{diff}}}{x_{max} - x_{min}} + \frac{y_{\text{diff}}}{y_{max} - y_{min}}$,
     $x_{\text{diff}} = \sum_i \left| \frac{p_x(i)}{p_x(i_{max})} - \frac{p_{x,R}(c_i)}{p_{x,R}(i_{max,R})} \right|$, $y_{\text{diff}} = \sum_j \left| \frac{p_y(j)}{p_y(j_{max})} - \frac{p_{y,R}(c_j)}{p_{y,R}(j_{max,R})} \right|$

Results

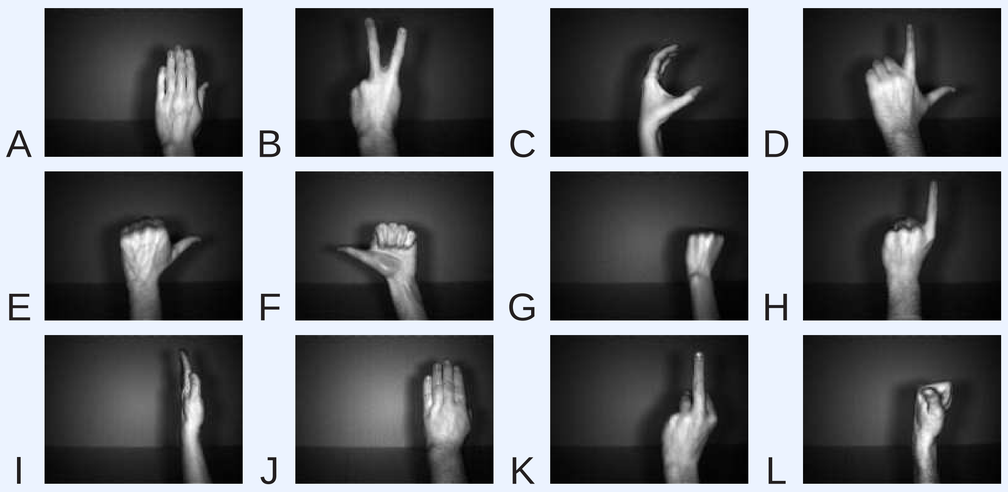

Figure 2: The 12 gestures recognized by the system.

Experimental setup:

- 34 persons à 12 gestures
- Distance range between 70 and 110 cm for the experiments
- PMD[vision] 19k; 160$\times$120 pixels; 40° viewing angle
- 'Leave-One-Out' evaluation

Results:

- Pre-processing: depth information of the camera is used
- Recognition rate of 93.14 % without additional depth features, 94.61 % with additional depth features
- Calculation time of 15 ms for segmentation and 15 ms for classification
- Gesture G and L $\Rightarrow$ Reason for the introduction of the depth features
- Confusion of gestures A-J and H-K

	A	B	C	D	E	F	G	H	I	J	K	L
A	31	0	0	0	0	0	0	0	0	3	0	0
B	0	32	0	0	0	0	0	0	0	0	2	0
C	0	0	33	0	0	0	1	0	0	0	0	0
D	0	0	0	34	0	0	0	0	0	0	0	0
E	0	0	0	0	33	0	1	0	0	0	0	0
F	0	0	0	0	0	34	0	0	0	0	0	0
G	1	0	0	0	0	0	32	0	0	0	0	1
H	0	0	0	0	0	0	0	32	0	0	2	0
I	1	0	0	0	0	0	0	0	32	0	1	0
J	2	0	0	0	0	0	1	0	0	31	0	0
K	0	0	0	1	0	0	0	3	0	0	30	0
L	0	0	0	0	0	2	4	0	0	2	0	26

	A	B	C	D	E	F	G	H	I	J	K	L
A	31	0	0	0	0	0	0	0	0	3	0	0
B	0	32	0	0	0	0	0	0	0	0	2	0
C	0	0	33	0	0	0	1	0	0	0	0	0
D	0	0	0	34	0	0	0	0	0	0	0	0
E	0	0	0	0	33	0	1	0	0	0	0	0
F	0	0	0	0	0	34	0	0	0	0	0	0
G	1	0	0	0	0	0	32	0	0	0	0	1
H	0	0	0	0	0	0	0	33	0	0	1	0
I	0	0	0	0	0	0	0	0	33	0	1	0
J	4	0	0	0	0	0	0	0	0	30	0	0
K	0	0	0	0	0	0	0	4	0	0	30	0
L	0	0	0	0	0	0	1	1	0	0	1	0	31

Table 1: Confusion matrix; Results of the classification. Left: without depth features, right: with additional depth features, vertical: gesture performed by the user, horizontal: recognized gesture class.

Discussion & Conclusions

- Fixed configuration (camera position, user)
- Recognition rate is improved by using additional depth information for pre-processing and classification of the gestures

Outlook

- Different gesture vocabulary because of high confusion probability
- Normalization of the captured gestures
- Improvement of the features
- Other classifiers, e.g., neural networks, self-organizing maps

References

[1] Breuer, P., Eckes, C., Müller, S. 'Hand Gesture Recognition with a novel IR Time-of-Flight Range Camera - A pilot study', *Lecture Notes in Computer Science*, 3rd international Conference MIRAGE 2007, Rocquencourt, France, pp. 247–260, March 2007

[2] Hornegger, J. and Niemann, H. 'Probabilistic Modeling and Recognition of 3-D Objects', *International Journal of Computer Vision*, Vol. 39, No. 3, pp. 229–251, September 2000