

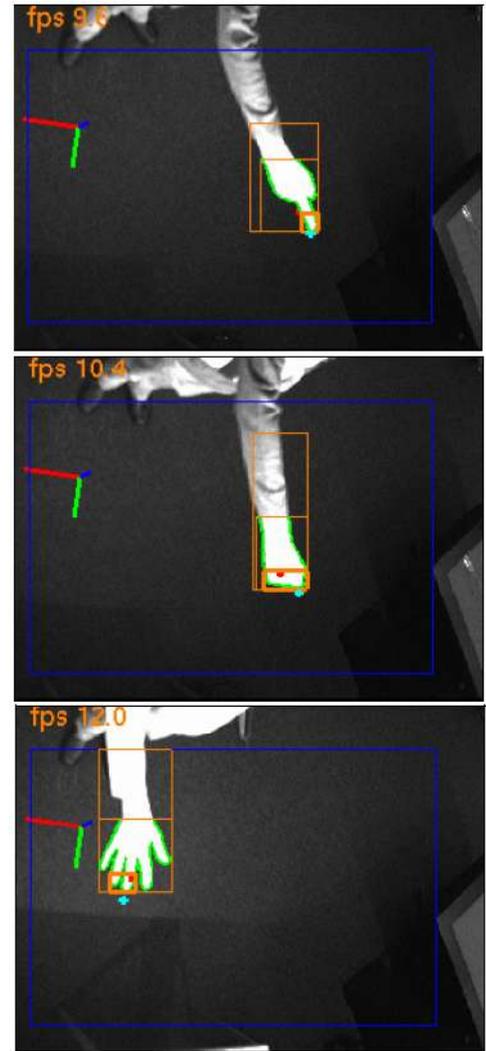
Dynamic Gestural Interaction with Immersive Environments

Cornelius Malerczyk
ZGDV Computer Graphics Center
Darmstadt, Germany



Motivation

- Deviceless interaction using hand gesture recognition with a calibrated stereo system
- Video-based interaction is one of the most intuitive kind of HCI with VR applications
- User is not wired to a computer
- Pointing is one of the most used gestures
- For 3D environments grabbing, moving and releasing objects are elementary tasks
- Methods for fast and robust recognition and tracking with short training phase to compensate individual differences



Technical Setup

- One single standard PC used for both gesture recognition and tracking and rendering of the scenario application
- Standard video beamer display in front of the user
- Feature tracking uses two IEEE1394 Firewire cameras
- Grey-scaled images with VGA res.
- Additional infrared diodes to ensure constant light conditions
- Technical setup is concealed from the user – displaying screen is the only piece of equipment visible to the user



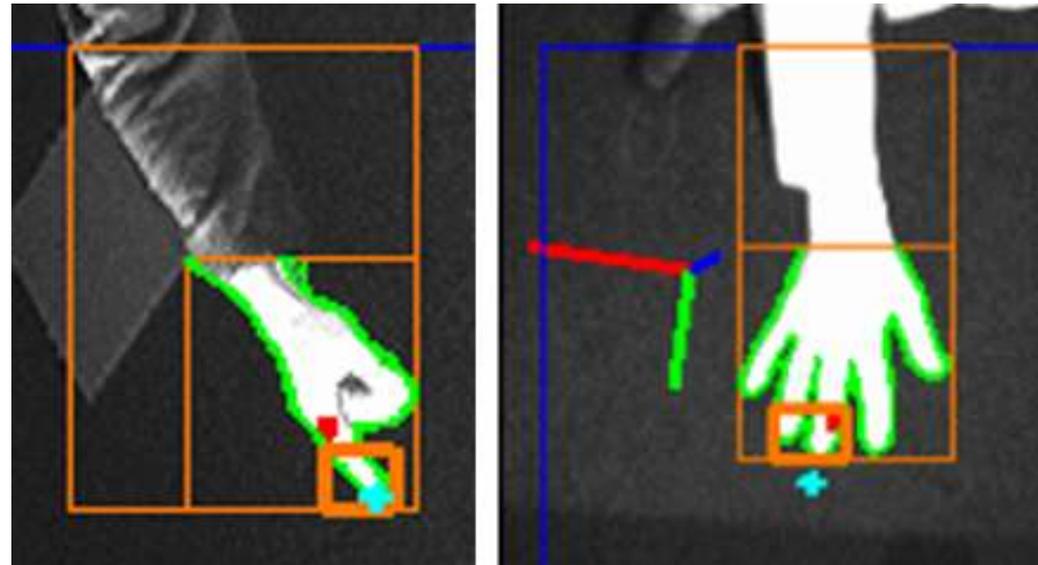
System Calibration

- Recognition and tracking of three different gesture in 3D
- Extrinsic camera calibration necessary after cameras setup
- Swaying a small torch light in the designated interaction volume to determine position and orientation of the cameras
- Definition of the world coordinate system
- Declaration of the corners of the displaying screen to ensure correct interpretation of especially the pointing direction



Gesture Extraction

- Segmentation of user's hand performed on 2D image basis
- Reference images and their edge images take at startup
- Difference images (of edge images) during runtime
- Segmentation at predefined threshold
- Approximate squared rectangle used to separate hand from forearm
- CoG used for 3D position
- Segments used for feature extraction



Feature Extraction

- 2D feature extraction of the human hand for each frame pair such as
 - Boundary length
 - Ratio between boundary length and segment area
 - Eccentricity of the segment
 - Elongatedness of the segment
 - Compactness of the segment
 - Curvature of the segment boundary
- Advantage of 2D: robust and fast determination
- Parameters may differ in both camera images (e.g. rotation)
- Solution: All parameter pairs are sorted by their size and stored as a feature vector for classification of gestures



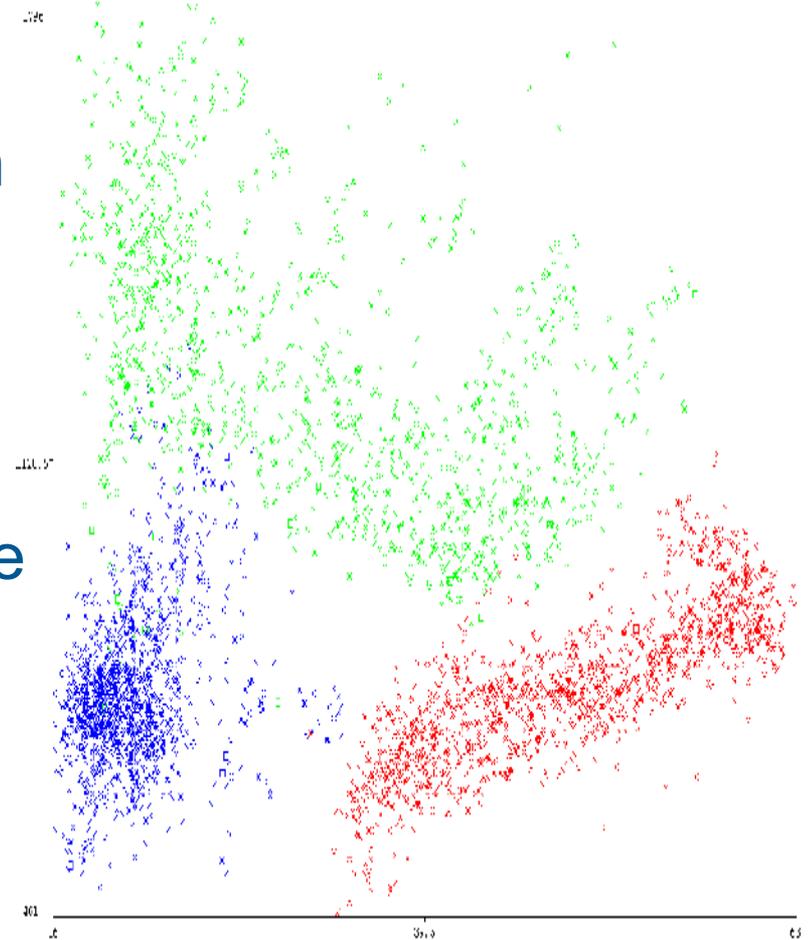
Gesture Classification

- Often used for feature classification: HMM and ANN
- Here: Algorithms for machine learning for data mining tasks:
 - Naive Bayes Classifier
 - Bayesian Network with K2-Hill-Climbing
 - Sequential Minimal Optimization
 - Random Tree Classifier
- Validation of the classification results of up to 800 vectors
- Using standard cross validation methods ($\frac{1}{2}$ for model construction, $\frac{1}{2}$ to examine recognition rate)
- Naives Bayes Classifier leads to sufficient balance for model building <5 sec. and online classification rate $>50\text{Hz}$



Gesture Training

- Short training procedure for each new user necessary to ensure robust and stable recognition with classification results of $>95\%$
- Recognition result decrease by $\sim 10\%$ if used a model build from various users
- User is asked to perform the three different gestures for ~ 10 sec. each at system startup
- Model can be saved for personalized usage



Pointing Gesture

- Gesture recognition systems used for HCI require feedback to reduce user's cognitive load
- VR interaction always needs a visual feedback directly and without delay on the output device
- 3D virtual representation of the human hand and laser pointer metaphor for pointing gesture
- Pointing target calculated as intersection of pointing ray with interaction screen
- Pointing ray starting at a predefined target point behind the user and the finger tip of the segmented hand



Results

- Cameras 30 fps
- Delay of cameras <1/5 sec.
- All image processing and computer vision task <50ms
- Classification results for three different gestures derived from ~250 online training parameter vectors (~10 sec. per gesture)
- Results of 60 sec. of interaction
- Elimination of outliers using post-processing queue

		Classified gesture		
		<i>Pointing</i>	<i>Grab</i>	<i>Release</i>
Gesture performed	<i>Pointing</i>	449	6	33
	<i>Grab</i>	3	426	1
	<i>Release</i>	24	2	379

Table 1. Classification matrix for three different gestures using a Naive Bayes probabilistic model

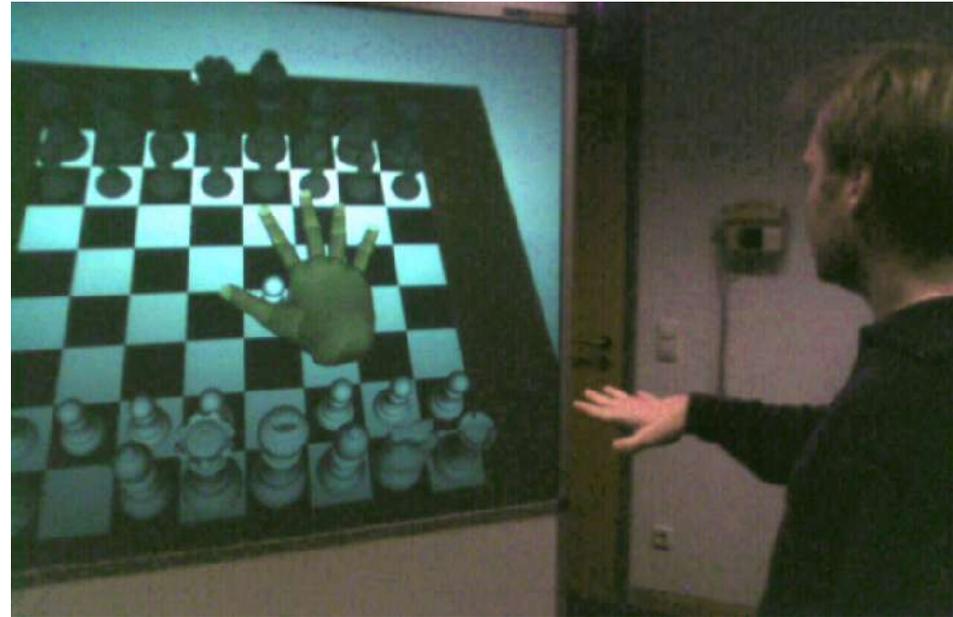
Training data	Test data	Recognition rates
Training data	Cross validation	95.1 %
Testing data	Cross validation	97.6 %
Testing data	Testing data	94.7 %

Table 2. Recognition rates for three different gestures



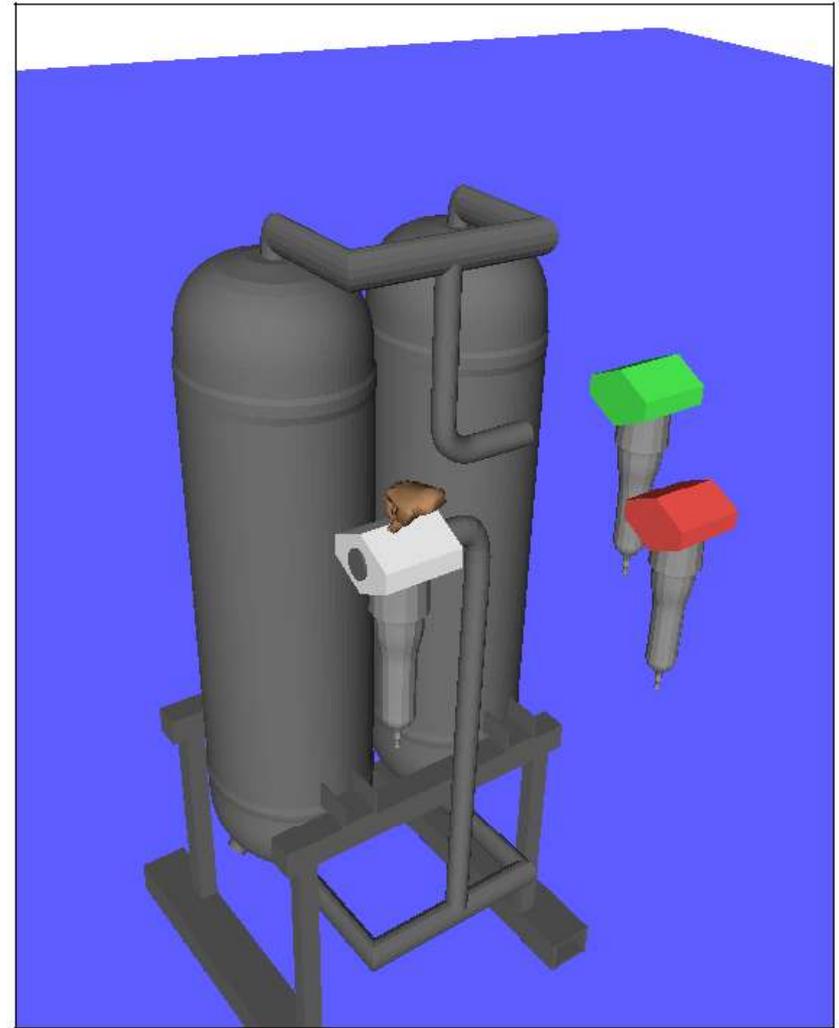
Applications (1)

- 3D virtual chess game using grabbing/releasing chess pieces
- User is standing in front of a large scaled screen
- Permanent visual feedback in real time using representation of hand
- Update of 3D position and recognized gesture
- „Near indicator“ and „snapping“ method



Applications (2)

- Industrial scenario
- User is asked to place color labeled filters to virtual 3D air pump system
- „Snapping“ method ensures a precise assembly of 3D objects, even if the user releases a filter object only roughly at the outlet.



Conclusion

- Video-based gesture recognition system using a calibrated stereo system with two off-the-shelf cameras
- Identification of three different hand gestures (pointing, grabbing and releasing)
- Determination of 3D position and pointing direction to enable an easy to learn and intuitive interaction with 3D environments
- Short training phase necessary to ensure high recognition rates of more than 95%
- Two different scenario applications have been developed

