Visuelle Perzeption für Mensch-Maschine Interaktion

Zusammenfassung / Überblick über die Vorlesung
13.02.2009
<table>
<thead>
<tr>
<th>Day</th>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mo</td>
<td>27.10.2008</td>
<td>Introduction, Applications</td>
</tr>
<tr>
<td>Fr</td>
<td>31.10.2008</td>
<td>Basics: Image Processing</td>
</tr>
<tr>
<td>Mo</td>
<td>03.11.2008</td>
<td>Basics: Image Transformations, Alignment</td>
</tr>
<tr>
<td>Fr</td>
<td>07.11.2008</td>
<td>Basics: Pattern recognition</td>
</tr>
<tr>
<td>Mo</td>
<td>10.11.2008</td>
<td>Computer Vision: Tasks, Challenges, Learning, Performance measures</td>
</tr>
<tr>
<td>Fr</td>
<td>14.11.2008</td>
<td>Face Detection I: Color, Edges (Birchfield)</td>
</tr>
<tr>
<td>Mo</td>
<td>17.11.2008</td>
<td>Project 1: Intro + Programming tips</td>
</tr>
<tr>
<td>Fr</td>
<td>21.11.2008</td>
<td>Face Detection II: ANNs, SVM, Viola &amp; Jones</td>
</tr>
<tr>
<td>Mo</td>
<td>24.11.2008</td>
<td>Project 1: Questions</td>
</tr>
<tr>
<td>Fr</td>
<td>28.11.2008</td>
<td>Face Recognition I: Traditional Approaches, Eigenfaces, Fisherfaces, EBM</td>
</tr>
<tr>
<td>Mo</td>
<td>01.12.2008</td>
<td>Face Recognition II</td>
</tr>
<tr>
<td>Fr</td>
<td>05.12.2008</td>
<td>Head Pose Estimation: Model-based, NN, Texture Mapping, Focus of Attention</td>
</tr>
<tr>
<td>Mo</td>
<td>08.12.2008</td>
<td>Project 1: Student Presentations, Project 2: Intro</td>
</tr>
<tr>
<td>Fr</td>
<td>12.12.2008</td>
<td>People Detection I</td>
</tr>
<tr>
<td>Mo</td>
<td>15.12.2008</td>
<td>People Detection II</td>
</tr>
<tr>
<td>Fr</td>
<td>19.12.2008</td>
<td>People Detection III (Part-Based Models)</td>
</tr>
<tr>
<td>Mo</td>
<td>22.12.2008</td>
<td>Scene Context and Geometry I (Ground-Plane, Hoiem, Leibe)</td>
</tr>
<tr>
<td>Fr</td>
<td>05.01.2009</td>
<td>Weihnachten/Ferien</td>
</tr>
<tr>
<td>Mo</td>
<td>09.01.2009</td>
<td>Facial Feature Detection (deformable templates, ASM, AAM)</td>
</tr>
<tr>
<td>Fr</td>
<td>12.01.2009</td>
<td>Facial Expression Recognition</td>
</tr>
<tr>
<td>Mo</td>
<td>16.01.2009</td>
<td>Gesture Recognition: Taxonomy, Neill, Starner (ASL), Nickel, ...</td>
</tr>
<tr>
<td>Fr</td>
<td>19.01.2009</td>
<td>Project 2: Student Presentations, Project 3: Intro</td>
</tr>
<tr>
<td>Mo</td>
<td>23.01.2009</td>
<td>Tracking I: Kalman, Particle Filter &amp; Applications (AV-Tracking, Body pose)</td>
</tr>
<tr>
<td>Fr</td>
<td>26.01.2009</td>
<td>Tracking II: Articulated Body Tracking</td>
</tr>
<tr>
<td>Mo</td>
<td>30.01.2009</td>
<td>Activity Analysis / Action recognition</td>
</tr>
<tr>
<td>Fr</td>
<td>02.02.2009</td>
<td>keine VL</td>
</tr>
<tr>
<td>Mo</td>
<td>06.02.2009</td>
<td>Project 3: Student Presentations &amp; Lessons Learned</td>
</tr>
<tr>
<td>Fr</td>
<td>09.02.2009</td>
<td>Audio-Visual Speech Recognition</td>
</tr>
<tr>
<td>Mo</td>
<td>13.02.2009</td>
<td>Wiederholung</td>
</tr>
</tbody>
</table>
Basics

- V2: Basics – Image Processing
  - Filtering
  - Linear Systems, properties, notation
  - Gaussian averaging
  - Edges
  - Multi-scale processing
Basics

- Image Transformations & 2D structure
  - Image formation, pinhole model
  - Homogenous coordinates
  - Calibration
  - Alignment
  - Euclidian, affine, perspective transformation
  - Geometric hashing
  - Hough transformation
  - RANSAC
Basics

- V4: Basics – Pattern recognition
  - Building blocks
  - Feature extraction
    - Curse of dimensionality → dimensionality reduction
    - PCA
  - Pattern classification
    - Bayes decision theory
    - Linear discriminant functions, perceptron
    - Instance-based learning
    - K-nearest neighbor
    - KD-trees
Face Detection

- Face detection I
  - Motivation, difficulties, representation
  - Color-based approaches
    - Color-spaces
    - How to model skin-color (parametric, non-parametric)
    - Histogram backprojection vs. Histogram matching
    - Classifiers
    - Postprocessing: Morphological operators
- Ellipsoid head model detection
  - And combination with color-based detection
  - Basically a deformable template
Face Detection II

- Artificial Neural Networks for Face Detection
  - Short repetition of ANNs
  - Preprocessing: histogram equalization
  - Network topology
  - Training

- Viola & Jones approach
  - Haar features
  - Integral image
  - Variant of AdaBoost to select features and to build a strong classifier
  - Classifier cascade for fast processing
Face Recognition I

- Introduction, cognitive issues, history
- Face recognition tasks:
  - open set, closed set, authentication / verification
  - Related metrics
- Feature-based approach
- Eigenfaces
- Fisherfaces
- Applications
Face recognition II

- Local appearance based approaches
  - Modular Eigenspaces
  - Using DCT

- Face Recognition using a 3D morphable model
  - Blanz & Vetter

- Databases & Benchmarks
Head Pose Estimation

- Motivation, relation with focus of attention

- Model-based head pose estimation
  - Needs facial landmark tracking & model

- Head pose estimation with ANNs

- Modeling focus of attention from head pose
  - „who was looking at whom?“
  - Determining whether a robot was addressed or not
People Detection I

- Global vs. Part-Based Approaches
- Levels of supervision
- Contours vs. Colors
- Gradient Histograms
- Interpolation

- Global Approaches
  - HOG
People Detection II

- Silhouette/Chamfer Matching
- Advanced silhouette matching (edge orientation, spatio-temporal templates)
- Earth-Mover’s Distance

- Part-Based Approaches
  - Connectivity structures
  - Geometric models/constraints
  - Wavelet-Based Feature description
  - Efficient body tree matching (Pictorial Structures)
People Detection III

- Local Features
  - Interest points and feature descriptors
  - Hessian-/Harris-Points
  - Automatic scale selection

- Implicit Shape Model
  - Star-Model
  - Codebook Generation, Clustering
  - Model Training, Occurrence distributions
  - Detection Loop
Scene Context & Geometry

- Implicit Shape Model
  - Segmentation, Reasoning about overlapping hypotheses
  - Extensions: Consistent shape voting (4D-ISM), Silhouette/SVM verification

- The Role of Scene Context
  - Dynamics, Perspective, Ground-Plane assumption
  - Estimating horizon position
  - Surface Estimation
  - Structure-from-Motion
Facial Feature Detection

- Repetition of already presented approaches
  - Viola & Jones, PCA-based, Morphable 3D model

- Elastic Bunch Graphs
  - Gabor Wavelets

- Statistical Appearance Models
  - Active Shape Model – ASM
  - Active Appearance Model – AAM

- Deformable Templates
Facial Expression Recognition

- Motivation, Problems, etc.

- Level of description
  - Facial expressions vs. Emotions
  - Six basic emotions
  - Facial Action Coding System (FACS)

- Systems
  - CMU Facial Analysis System (Tian et al)
    - Model-based features, ANNs for classification
  - UCSD system for analysis of spontaneous facial behaviors
    - Appearance-based (Gabor-WL), SVM / HMM
Gesture Recognition

- Definition, applications, types of gestures
- Taxonomies
- Building blocks
- Repetition of HMMs
- Systems
  - Sign Language Recognition (Starner et al)
  - Pointing Gesture Recognition (Nickel et al)
  - Combining Gestures and Speech
- Applications
Tracking I

- Definitions
- Features
  - Templates, color, background models
- Tracking Schemes
  - Mean Shift
  - Kalman Filter
  - Particle Filter
- Examples
  - Audio-Visual Tracking
Tracking II

- Multi-camera systems
  - Calibration, triangulation, stereo processing
- Multi-object tracking
- Tracking of heads and hands
  - Depth from stereo + color, …
- Articulated Body Tracking
  - Taxonomy
  - Models
  - Particle filter for tracking
  - Volume Carving / Voxels
- Metrics for Multi-Object Tracking
Activity Analysis
Audio-Visual Speech Recognition

- Motivation, McGurk Experiment
- Basic Blocks
- Visual Feature Extraction
  - Model-based, appearance-based, hybrid
- Fundamentals of ASR
- Phonemes vs. Visemes
- AV-Fusion
  - Feature fusion
  - Intermediate fusion: multi-stream HMM, product-HMM
  - (Late fusion)
  - Stream weight adaptation