Face Detection, Facial Gesture Recognition

Attila Fazekas
Debrecen

- **Debrecen** is Hungary's second largest city after Budapest.
- It was the largest city in the 18th century and was also the capital during the revolution in 1848–1849 and by the end of the World War II in 1944–1945.
University of Debrecen

- 33000 students
- 20% in English Program
- 14 faculties
- 24 doctoral schools
- Top Ranked
- Long history started from Reformed College of Debrecen in 1538
Faculty of Informatics

**Students:** 1995  
- Full time: 1722  
- Part time: 233  
- PhD students: 33  
- Students in English program: 429

**Academic staff**  
- 7 full professors  
- 20 associate professors  
- 34 senior lecturers  
- 12 assistant lecturers
Road Map

• Multi-Modal Human-Computer Interaction (HCI)
• Face Analysis
• Face Detection
• Facial Gesture Recognition
• Turk-2 and Other Examples
• Support Vector Machines
What is the Multi-Modal Interaction?

• There are two views on multi-modal interaction:
  
  • The first focuses on the **human** side: perception and control. There the word modality refers to human input and output channels.
  • The second view focuses on synergistic using two or more **computer** input or output modalities.
Modalities

• We can divide the modalities in seven groups
  • Internal chemical (blood oxygen, etc.)
  • External chemical (taste, etc.)
  • Somatic senses (touch, etc.)
  • Muscle sense (stretch, etc.)
  • Sense of balance
  • Hearing
  • Vision
Definition of the Multi-Modality

• "Multi-modality is the capacity of the system to communicate with a user along different types of communication channels."

• Both multimedia and multi-modal systems use multiple communication channels. But a multi-modal system strives for meaning.
Two Types of Multi-Modal Systems

- The goal is to use the computer as a tool.
- The computer as a dialogue partner.
Benefits

• **Efficiency** follows from using each modality for the task that it is best suited for.

• **Redundancy** increases the likelihood that communication proceeds smoothly because there are many simultaneous references to the same issue.

• **Perceptability** increases when the tasks are facilitated in spatial context.
Benefits

• **Naturalness** follows from the free choice of modalities and may result in a human-computer communication that is close to human-human communication.

• **Accuracy** increases when another modality can indicate an object more accurately than the main modality.
Applications

• Mobile telecommunication
• Hands-free devices to computers
• Using in a car
• Interactive information panel
Face Analysis

• Face carries a lot of important information in communication
• Monitoring the face is fundamental in HCI
• First step: face detection (localization)
• Using the localized face can be performed:
  • Tracking face and facial features
    • 2D face tracking, gaze estimation, head-shake detection
  • Face classification
    • gender, age, facial expressions, race
  • Feature extraction
    • skin/eye/hair color
    • mustache/beard detection
Related Topics

• Face detection (one face/image)
• Face localization (more faces/image)
• Facial feature detection (eyes, mouth, etc.)
• Facial expression recognition
• Face recognition, face identification
• Face tracking
Problems of the Face Detection

- **Pose**: The images of a face vary due to the relative camera-face pose
- **Presence or absence** of structural components (beards, mustaches, glasses etc.)
- **Facial expression**: The appearance of faces are directly affected by the facial expression
- **Occlusion**: Faces may be partially occluded by other objects
- **Image orientation**: Face images vary for different rotations about the optical axis of the camera
- **Imaging conditions** (lighting, background, camera characteristics)
Face Detection

- **Scanning** of the picture by a running window in a multi-resolution pyramid
- **Normalize** of the window
- Hide some parts of the face (**masking**)
- **Normalize** of the local variance of the brightness on the picture
- Localization of the face (**decision**)
Face Detection

- Eigenfaces – PCA
- Neural Networks
- SVM
- AdaBOOST (by Viola and Jones, 2001)
  - Stable
  - Fast method (10-15x faster)
  - Open Source implementation can be found in OpenCv (by Lienhart, 2002)
  - A Google Portrait uses this method.
Boosting

• Make a strong classifier from a lot of weak ones.
• It based on selection of the good candidates.
Face detection/Face Tracking

- Viola and Jones detector is used
  - Face detection is reduced to image classification problem
- Given a set of feature types:
- Training:
  - positive (faces) and negative (random images) examples
  - those features are selected which fits on the positive set (finding position and their extent)
  - the selected features are collected into a cascade file
- Face detection
  - the different scale of the input image is scanned through
  - fitting the set of features:
    - if all are fitting => there is a face
Face Classification

- Can be used for: gender, age, facial expression, race detection
- Preprocessing – for feature extraction
  - Cutting the face, **LBP transform**, Gabor transform
- Classification (using the model file)
  - **SVM**, AdaBoost
- Databases: Cohn-Kanade, FERET
- Typical classifiers trained (LBP+SVM)
  - Gender (male vs. female), Facial expressions (happy, sad, surprised, angry), Age estimation (10-29, 30-49, 50-), Race (asian, hispanic, black, white)
Face Gesture Recognition

• Let us consider a set of the facial pictures.
• Let us set up a finite system of some features related the pictures.
• It is known any pictures is related to only one class: face with the given gesture, face without the given gesture.
• The problem to find a method to determine the class of the examined picture.
• One possible way to solve this problem: Support Vector Machine.
Facial Feature Color Extraction

- Determining the color of various facial features: skin, hair and eyes.
  - The full color range of the segmented face image will be reduced to color categories based on human cognition principles
  - The segmentation steps of the HI plane in case of hair colors: (a) color marker points, (b) convex hull, (c) distance transform and (d) the segmented plane:
Facial Feature Extraction

• The colors of the facial features are determined in two steps:
  • First, the skin, eyes and hair are segmented in the image using only structural information
  • Then, within the segmented regions the huge number of colors in real color images is substituted by a smaller color set, which is used to determine the color of a given feature.
Head-pose Tracking, Gaze Estimation

• Head pose estimation
  • Detecting the location of the facial features using individual feature detectors (Viola and Jones detectors): eyes, mouth, nose tip
  • Based on the position of the facial features, the POSIT algorithm is used to estimate the three rotation and the translation vector of the head pose
Turk 2 Chess Player

• Parts of the system:
  • Robot arm
  • Visual perception module
  • Chess engine
  • Talking head
Structure of the System

• Input channels:
  • Chess board
  • Gestures of the human player
  • Speech of the human player

• Output channels:
  • Movements of the robot arm
  • Gestures of the virtual player
  • Speech of the virtual player
Face Detector

• Detection of the player’s existence
• Detection of the player’s face
• Extracted features:
  • Emotions
  • Sex
  • Age
• Using:
  • SVM
  • Cohn-Kanade database
Speech Recognition

• Recognizing only keywords to control the game
• Using: Hidden Markov Model Tool Kit (HTK)
Talking Head

• The first Hungarian talking head able to express emotions
• Emotions: sad, natural, happy, bored
• Gestures to make the behaviour of the head more life-like (Perlin – noise)
• MPEG4 FAP parameters
• Using: charToon, ProfiVox TTS
Lip sync

- MPEG-4 FAP parameters
- Candide mask
Chess State Recognizer

• Detecting the changes on the chess board:
  • Determining the fields containing pieces
  • Determining the color of the piece on the field
Robot Arm

• The robot arm is a mechanical arm with 4 degrees of freedom developed especially for this project

• There are similar products that can be used like this (for example LEGO® robots)
The Controller

• Synchronizing the other components
• Following the abstract state of the game:
  • Human player is thinking
  • Human player makes the move
  • Virtual player is thinking
  • Virtual player makes the move
• Each state is characterized by a set of properties (to control the behaviour of the output interfaces):
  • Gaze direction
  • Probability of the gaze
  • Utterances
  • Time of emotion holding
Chess Player Turk - 2
HeadPose
BillBoard
Support Vector Machine

• Statistical learning from examples aims at selecting from a given set of functions $f_\alpha$, the one which predicts best the correct response.
• This selection is based on the observation of $l$ pairs that build the training set:

$$(x_1, y_1), \ldots, (x_l, y_l), \text{ } x_i \text{ real number and } y_i = +1 \text{ vagy } -1$$

which contains input vectors $x_i$ and the associated ground ”truth” given by an external supervisor.
• We would like to find the function $f_{\alpha_0}$ which minimizes the risk function.
Support Vector Machine

• The basic idea of SVM to construct the optimal separating hyperplane.

• For the linearly separable case, SVM simply seeks for the separating hyperplane with the largest margin.

• For linearly nonseparable data, by mapping the input vectors, which are the elements of the training set, into a high-dimensional feature space through so-called kernel function.

• We construct the optimal separating hyperplane in the feature space to get a binary decision
Thank you for your attention