AUVIS

University of Cologne

Deutsche
Sporthochschule Köln
German Sport University Cologne

Fraunhofer
Heinrich Hertz Institute

Fraunhofer
IAIS

Advances on Video Analysis

Sankt Augustin, Germany, 23rd January 2015

Oliver Schreer
Stefano Masneri

Outline

- Previous status of the recognizer
- Latest developments
  - Improvements to tracker
  - Improvements to annotations
- Current annotation results
  - Tracking results
  - Evaluation of annotation results
- Future work

Previous status of the recognizer

- At the beginning of AUVIS:
  - Limited support for background estimation and camera motion
  - No support for tracking of joined hands
  - Poor tracking results

Previous status of the recognizer

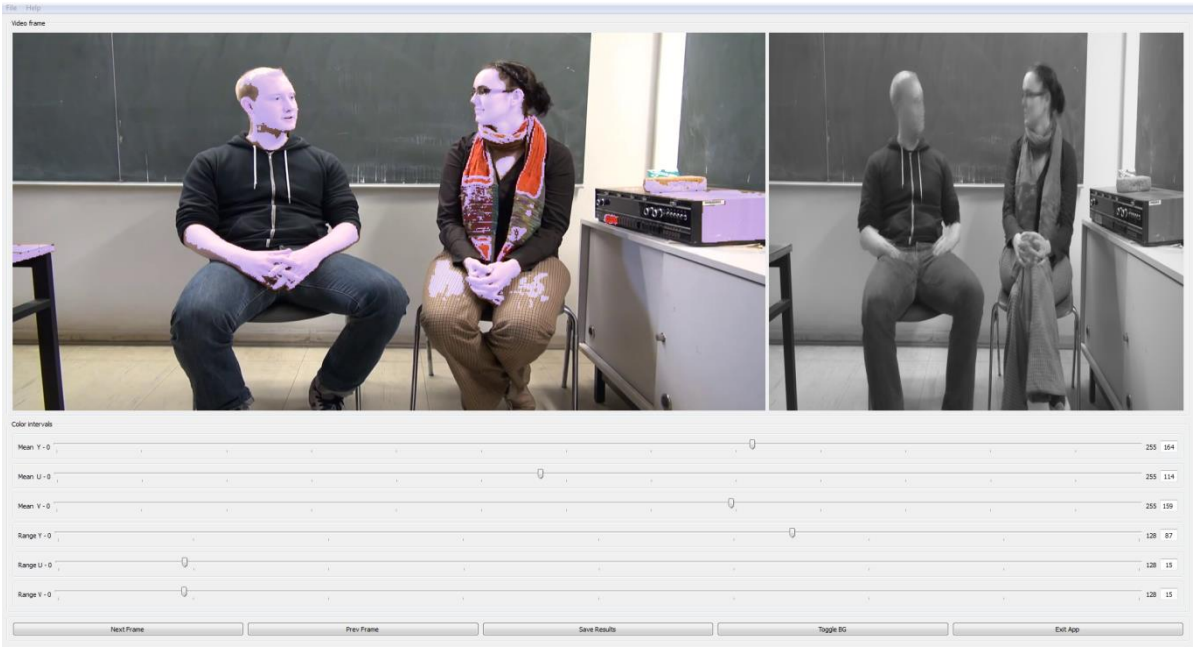
- Last year (AUVIS Workshop 2014):
  - Several algorithmic improvements
  - Partial support for Neuroges annotations
  - Much more reliable detection and tracking of hands

Latest developments

- Hands and head tracking:
  - Adaptive skin colour estimation
  - Probability maps for improved detection and tracking
- Gesture analysis:
  - Outlier removal, filling of missing data
  - Structure annotation from full trajectory analysis
  - Improved analysis of intrinsic motion

Skin colour estimation

- Automatic skin colour estimation usually fails on AUVIS videos
- Manual estimation possible via GUI interface, but:
  - Results may still be unsatisfactory
  - It takes time to learn to use the tool

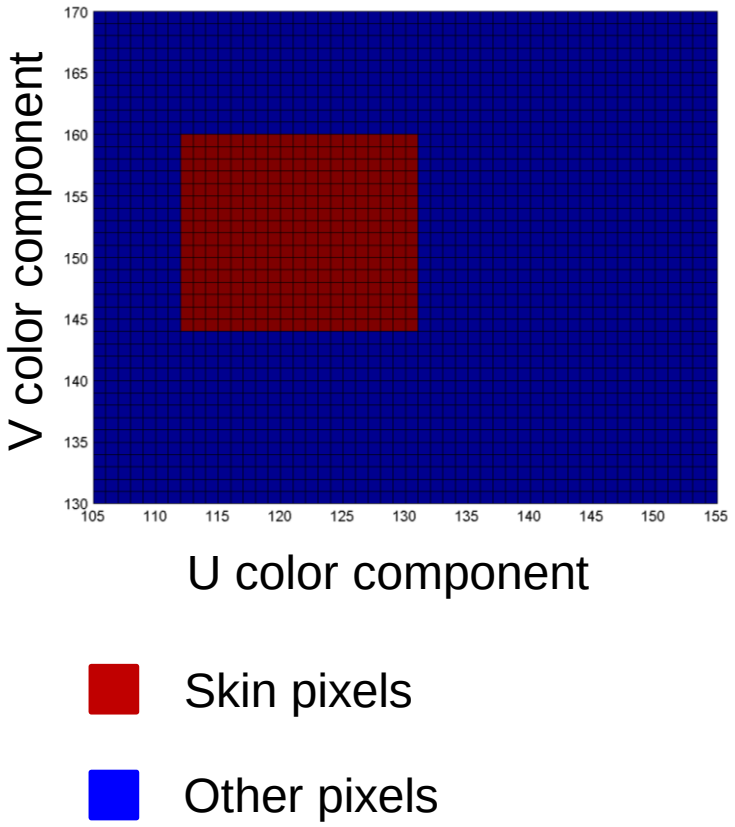


Initial approach

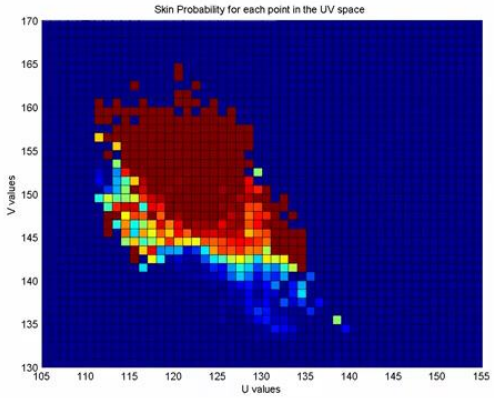
Limitations of the approach:

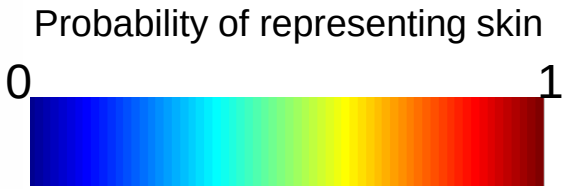
- Fixed segmentation parameters
- Binary approach
- Only rectangular distributions allowed
- Temporal changes of skin colour parameters cannot be considered
- Difficult videos require fine-tuning by the user

Pixels distribution in UV colorspace



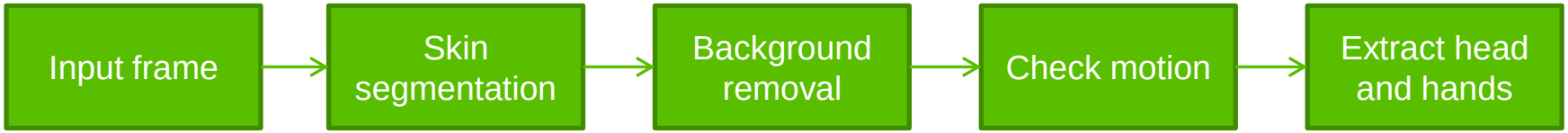
Adaptive skin colour estimation





- The estimation process is faster and more accurate
- The first 20 frames are used to learn skin colour parameter
- Fuzzy approach
- Skin colour updated over time ➔ robust to changes of colours
- Available in two modes:
  - Fully automatic: training done using face pixels
  - Semi-automatic: using user-defined initial parameters

Initial tracking framework



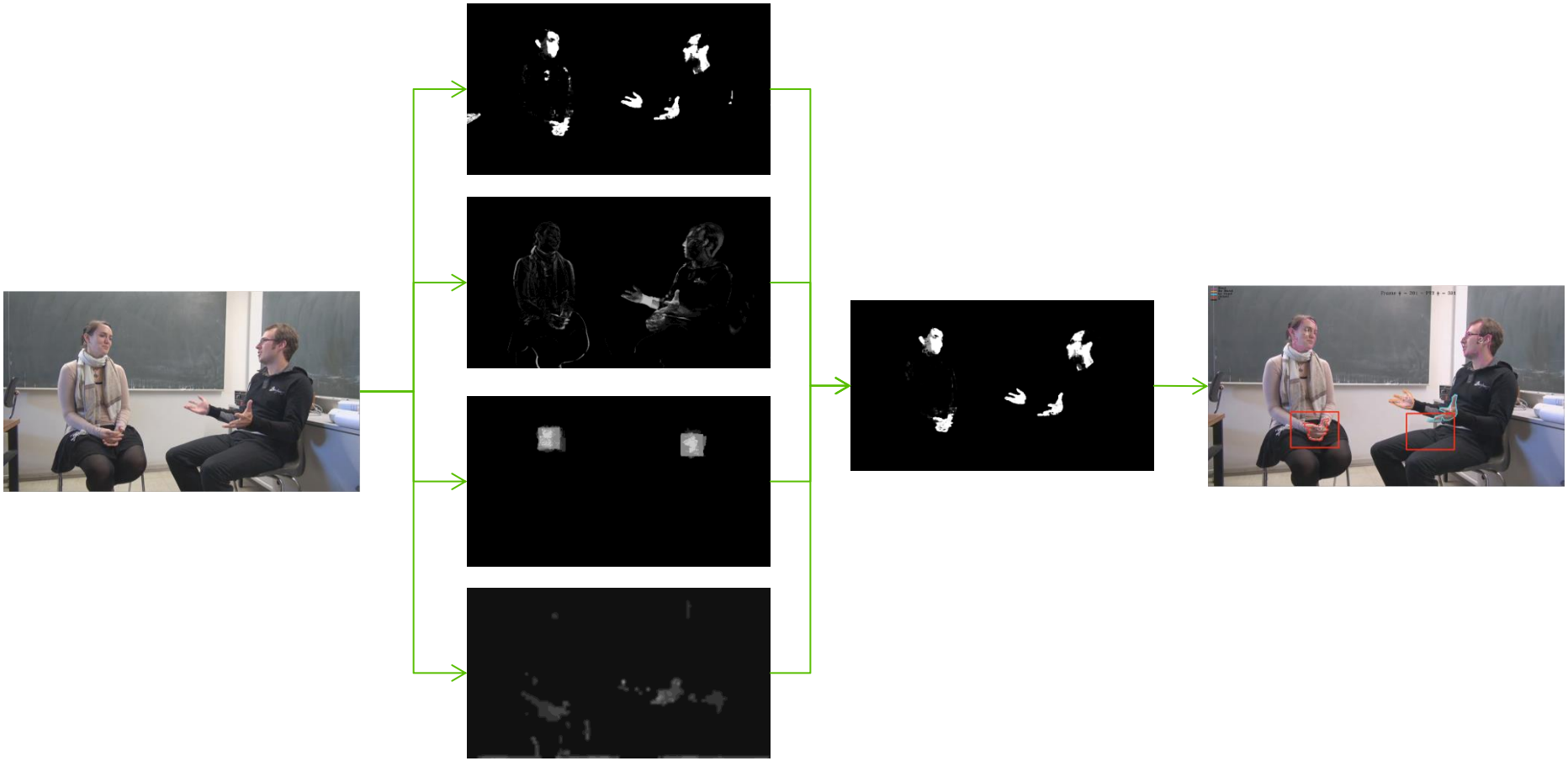
- 3 different modalities (colour, background and motion) are used as binary input
- Successive processing of different modalities

Identified Problems:

- Information loss due to binarization
- Information loss due to sequential processing

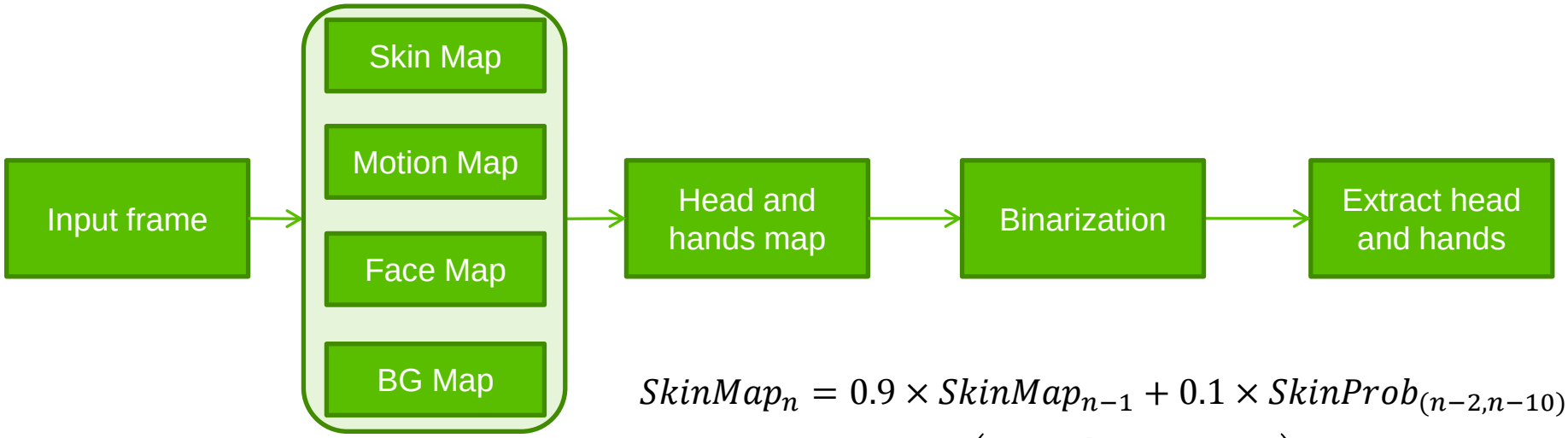
Novel tracking framework

- From binary images to *probability maps*



Novel tracking framework

- From binary images to *probability maps*



$$Map_n = \sum w_i M_{i,n} \,;\, \sum w_i = 1$$

$$SkinMap_n = 0.9 \times SkinMap_{n-1} + 0.1 \times SkinProb_{(n-2,n-10)}$$

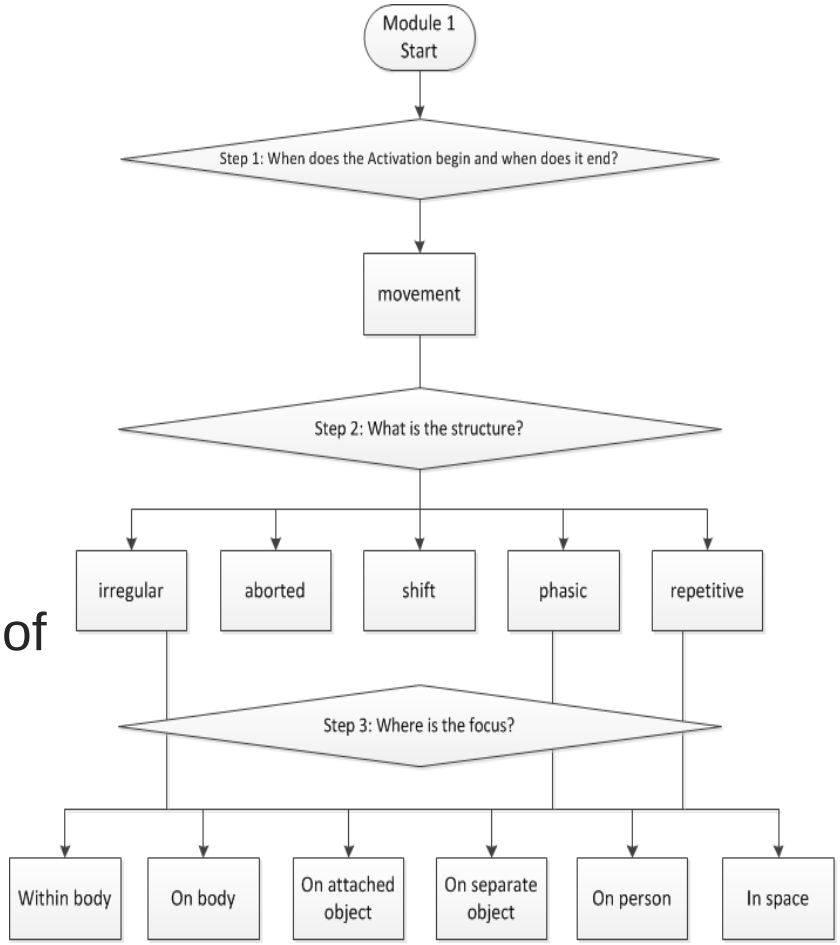
$$MotionMap_n = \max\left(255, \sqrt[2]{MV_x^2 + MV_y^2}\right)$$

$$FaceMap_n = \begin{cases} \max(128, 1 + FaceMap_{n-1}), & if\ face \\ 0.8 \times FaceMap_{n-1}, & otherwise \end{cases}$$

$$BGmap_n = 255 - \left(|I_y - BG_y| + |I_u - BG_u| + |I_u - BG_u|\right)$$

Gesture analysis

- Once hands position is available, movements can be analysed:
  - Assignment of left / right hand for each participant
  - Detection of start and end of each movement
  - Assign the *structure* and the *focus* of movement



Initial gesture analysis framework

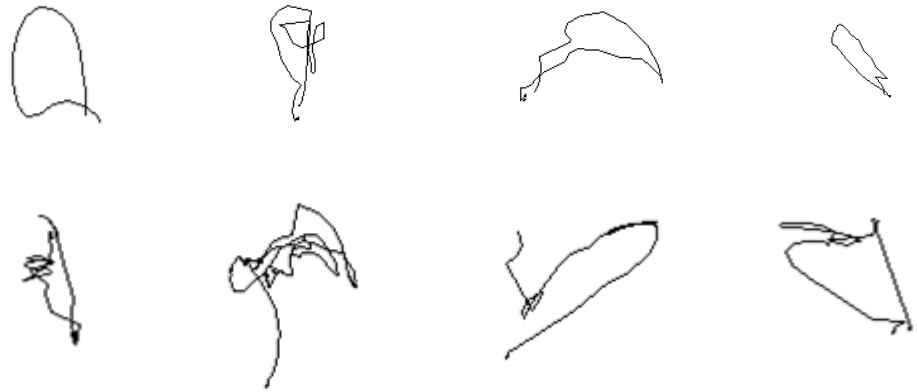
- No post-processing of the hands trajectory:
  - Single-frame errors could possibly lead to wrong annotations
- Structure assignment based on directions rather than trajectories
  - Simple approach
  - Not accurate enough for most of the movements

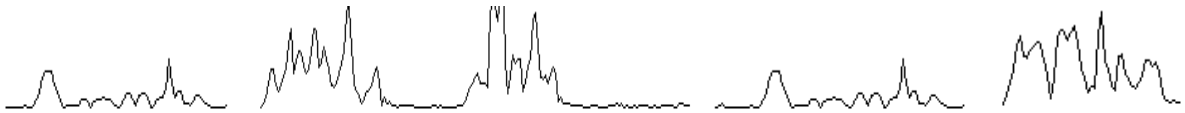
Novel gesture analysis framework

- Post-processing of the hands trajectory:
  - Outlier removals
  - Linear fitting of missing data
  - Smoother trajectories
- Using trajectory and speed information rather than just direction of motion
- Several constraints:
  - video framerate
  - movement duration and characteristics

Novel gesture analysis framework

- Hand **trajectories**:
  - Phasic
  - Repetitive
- Hand **speed**:
  - Phasic
  - Repetitive





Novel Tracking Framework Results

Novel Tracking Framework Results

Annotation Results

- Standard evaluation measure is the interrater agreement
- In our case, agreement between human and automatic annotation
- For Activation units, agreement is defined as the fraction of overlapping movements over the total of annotated movements
  - $A_{act} = (Act_1 \cap Act_2)/(Act_1 \cup Act_2)$
- For Structure units, agreement is defined from a modified version of Cohen's Kappa measure
  - $A_{str} = \kappa = \frac{p_a - p_e}{1 - p_e}$
  - Takes into account partial overlaps between movements

Annotation Results

- Good interrater agreement values for activation units on most of the test videos

Video	Agreement Left	Agreement Right
Alex	0.71	0.75
Caro	0.74	0.76
Feli	0.70	0.70
Dani	0.43	0.40
Karo	0.48	0.45
Nele	0.51	0.54
AVERAGE	0.59	0.60

Annotation Results

- Automatic-human rater agreement for structure units is still too low
  - Interrater agreement is considered good when k > 0.60
- Main sources of error:
  - Single movements detected as multiple movements
  - Shift and Aborted untis not considered by automatic annotations
- Frame-based results slightly better than annotation-based ones:

	Precision	Recall	F-Measure
Phasic	48.5%	19.9%	28.3%
Repetitive	47.4%	40.2%	43.6%
Irregular	54.0%	64.0%	58.6%

Future work

- Further improvements to the tracking framework
  - Hand / Face overlaps
  - Arm / Hand separation
- Reach automatic-human agreement comparable to human-human interrater agreement based on EasyDiag (standard evaluation tool)
- Implement shift and aborted annotations for structure units
- Evaluate results for Focus annotations

Summary

- The tracking framework has been completely revisited
- The gesture analysis framework has been refined and made more robust
- Evaluation showed good annotation results for *activation* units, less so for *structure* units
- Future work will focus on improving annotation results as well as expanding the types of annotated units

Thank you !!!

Contact: Oliver.Schreer@hhi.fraunhofer.de
Stefano.Masneri@hhi.fraunhofer.de