IM2.VP
Visual Information Processing

Phase II - Achievements

Prof. Jean-Philippe Thiran

IM2.VP - Tasks

IM2.VP.1- Face detection and recognition
- IDIAP (Dr S. Marcel) & EPFL (Prof. J.-Ph. Thiran)

IM2.VP.2 - Multi-source image representation for joint processing and coding
- EPFL (Prof. P. Frossard)
- task finished in December 2008

IM2.VP.3 - Multi-person activity discovery
- IDIAP (Dr D. Gatica-Perez & Dr J.-M. Odobez)
- task finished in December 2007

IM2.VP.4 - Head activity analysis and visual Focus of attention (VFOA)
- IDIAP (Dr J.-M. Odobez)

IM2.VP.5 - Recognition and interpretation of body gestures
- ETHZ (Prof. L. van Gool, Dr B. Fasel) & EPFL (Prof. A. Billard)
- task finished in December 2007

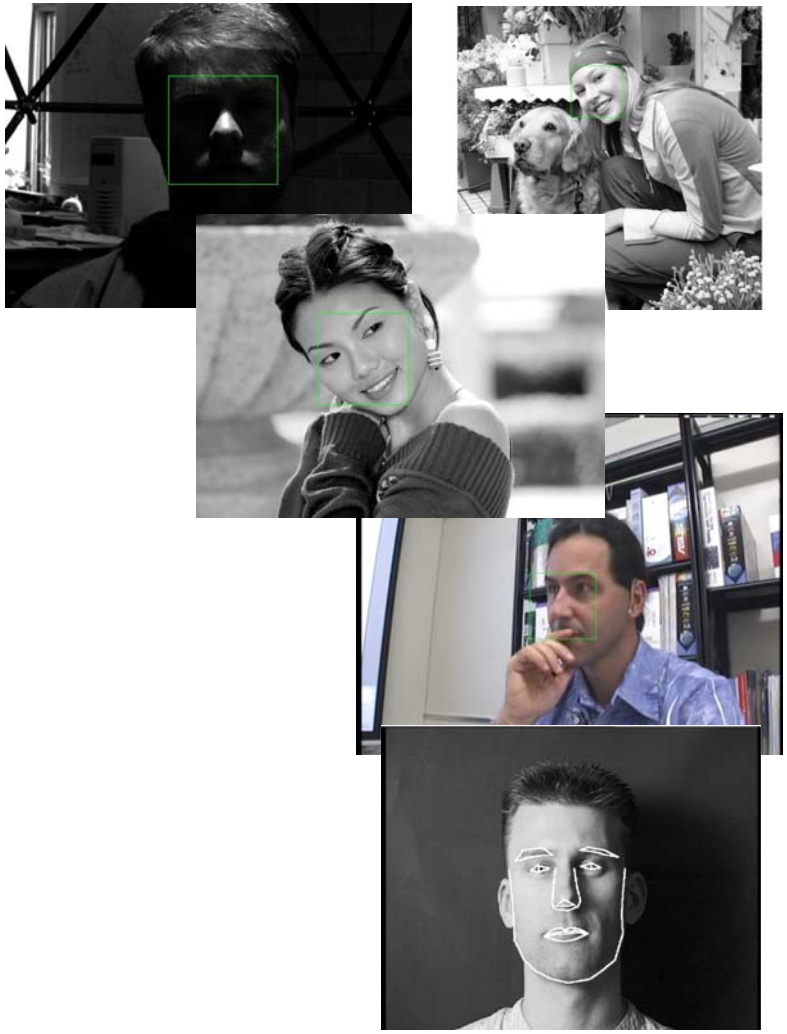
IM2.VP.6 - Handwriting recognition and whiteboard data analysis
- Uni. Bern (Prof. H. Bunke)

IM2.VP.7 - Omnidirectional visual attention
- Uni NE (Prof. H. Hügli)

IM2.VP.8 - Vision-supported speech understanding
- ETHZ (Prof. L. van Gool & Dr B. Pfister)

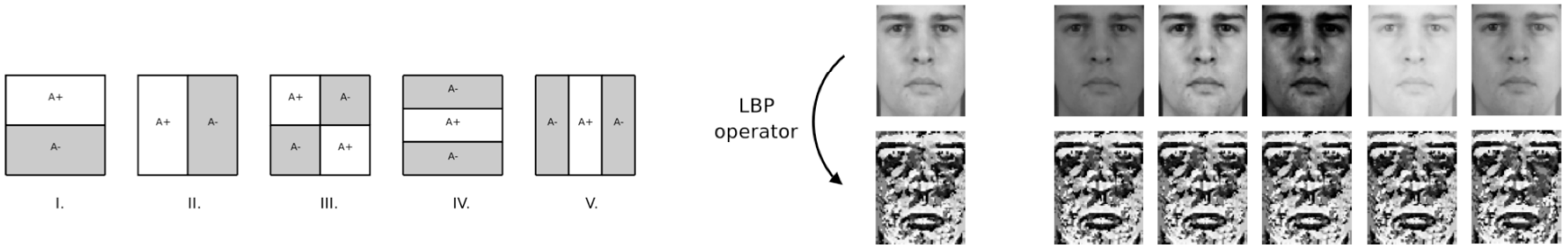
Face detection, alignment & recognition

- Frontal and MultiView Face Detection System
  - New robust features
  - Pyramid of classifiers for simultaneous non-frontal face detection and pose estimation
  - Information theoretic optimal classifier combination
- Robust-to-illumination Face Alignment System
- Generative Models for Face Recognition



Haar-LBP Features

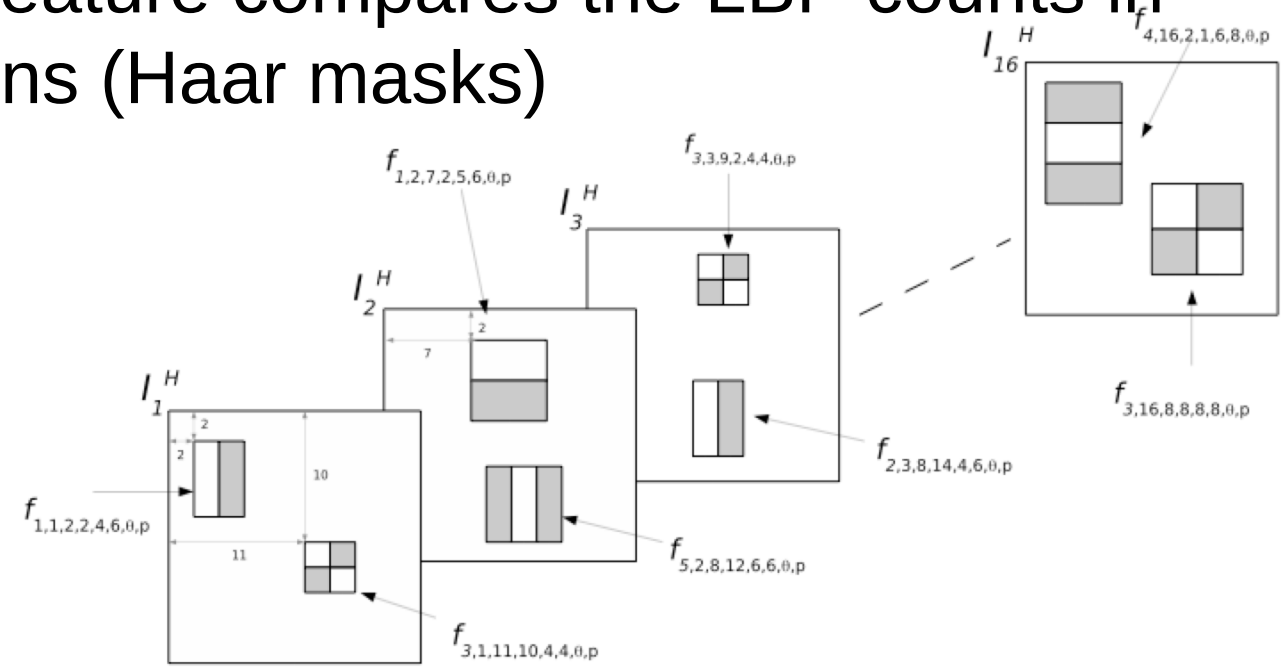
- We investigated a novel feature set for fast illumination invariant face detection
  - Combines the advantages of both Haar and LBP



  - Similar boosting framework with an augmented Haar-like feature

Haar-LBP Features

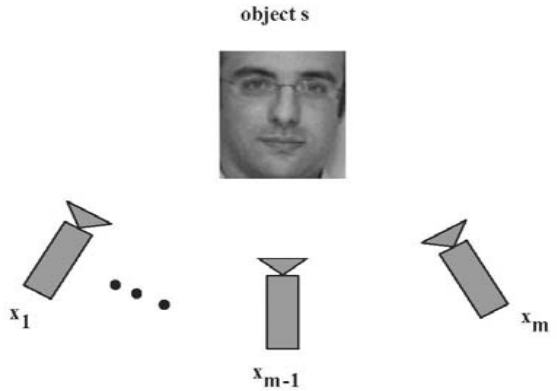
- An HLBP feature compares the LBP counts in 2 subregions (Haar masks)



- Results: more robust to illumination variations compared to Haar and LBP

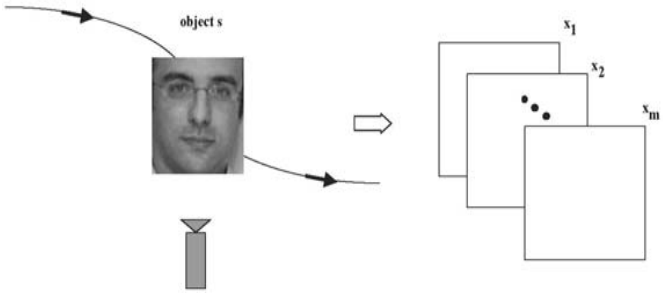
Multi-camera object recognition (P. Frossard)

- Recognition from multiple observation sets
  - Diversity increases performance
  - Multi-camera object recognition
  - Video face recognition



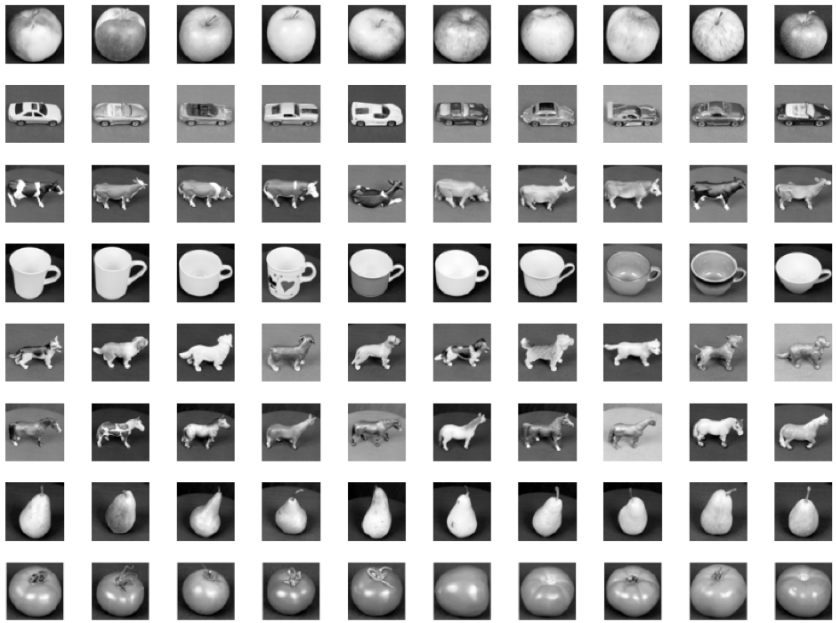
Vision sensor networks

- Graph-based label propagation solution
  - Preserved data and label smoothness
  - Low-complexity solution
- Distributed solution by consensus



Video sequence

Sample achievements



MASC	MSM	KMSM	KLD
88.88 (1.71)	74.88 (5.02)	83.2500 (3.4)	52.5 (3.95)

TABLE I
OBJECT RECOGNITION RATE IN THE MEAN(STD) FORMAT, MEASURED ON THE ETH-80 DATABASE.

Classification of multiple observation sets by semi-supervised learning
Effrosyni Kokiopoulou and Pascal Frossard
submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence, October 2008.

3D Face Recognition with Sparse Spherical Representations
Roser Sala Llonch, Effrosyni Kokiopoulou, Ivana Tosic and Pascal Frossard
Accepted for publication, Pattern Recognition, April 2009.

Video face recognition with graph-based semi-supervised learning
Effrosyni Kokiopoulou and Pascal Frossard
Invited paper, IEEE ICME, Cancun, Mexico, July 2009.

Graph-based Classification for Multiple Observations of Transformed Patterns
Effrosyni Kokiopoulou, Stefanos Pirillos and Pascal Frossard
Proceedings of ICPR, Tampa, Florida, USA, December 2008.

3D Face Recognition using Sparse Spherical Representations
Roser Sala Llonch, Effrosyni Kokiopoulou, Ivana Tosic and Pascal Frossard
Proceedings of ICPR, Tampa, Florida, USA, December 2008.

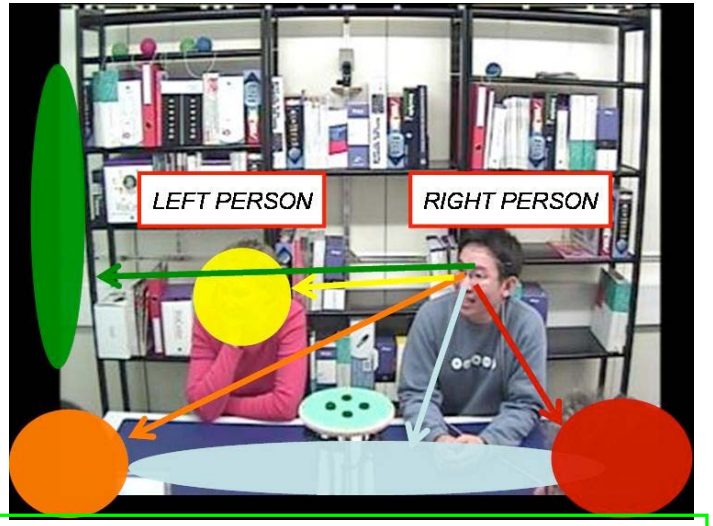
Multimodal multiparty joint VFOA recognition

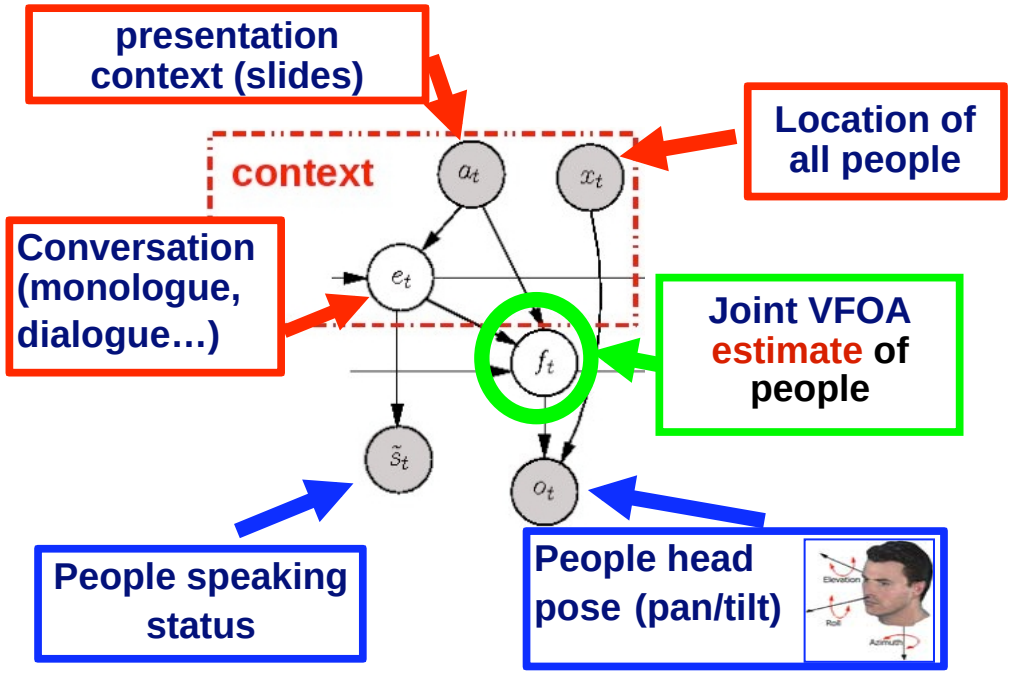
Visual Focus of Attention (VFOA) Recognition

who is looking at whom or what ?

Relevant for

human interaction modeling, addressee detection





Set of Dynamic Bayesian Networks for conversation modeling

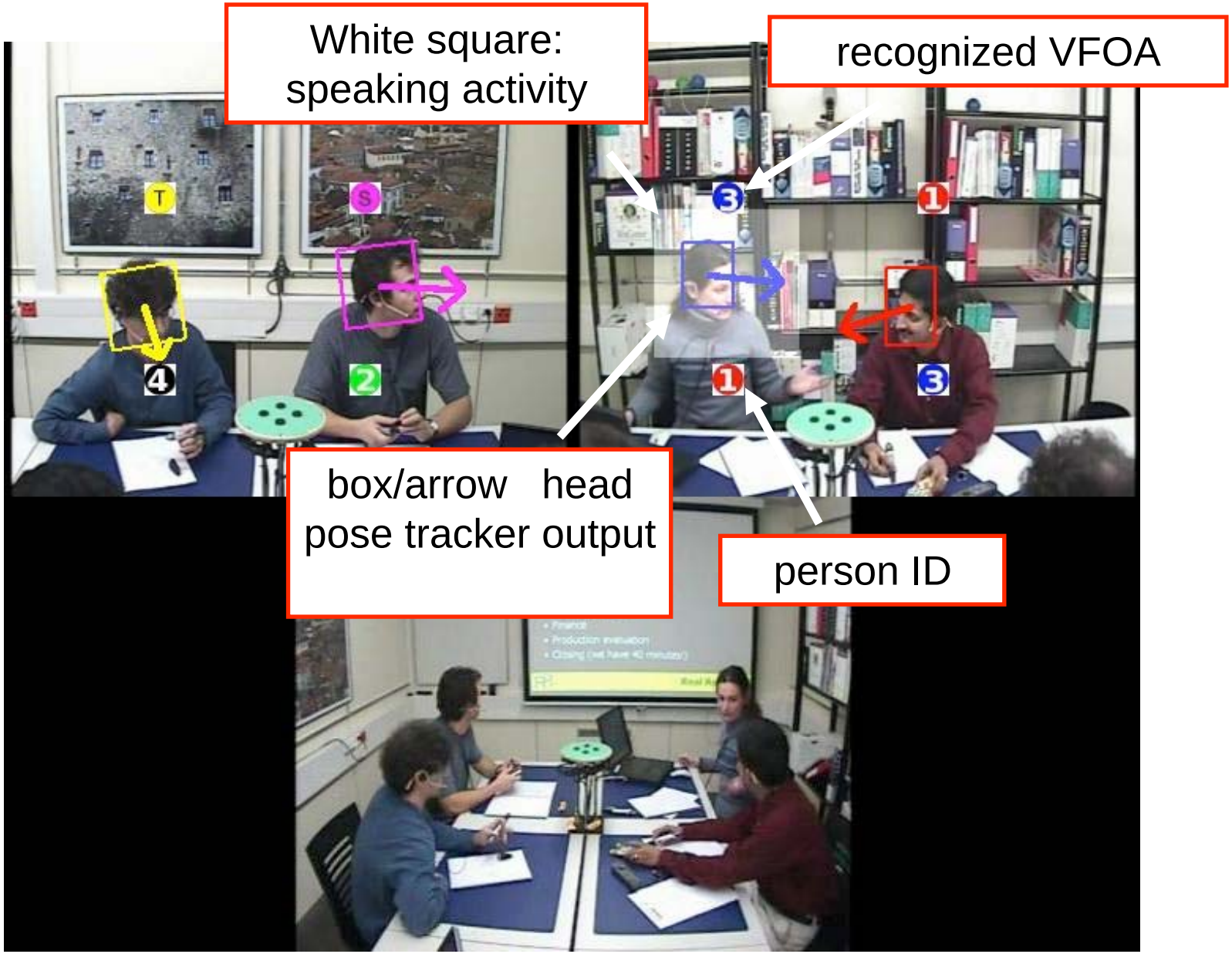
- **multimodal interaction** modeling (gaze, speech, visual activity of all people)
- **contextual** recognition (presentation activity, conversation status, people location)
- **contextual model parameters adaptation**
- handling **moving** people

Evaluation on IM2/AMIDA datasets

5 hours of natural meeting - 55% VFOA recognition rate – improvement due to context modeling > 10%

S. Ba and J.M. Odobez, **Recognizing Visual Focus of Attention from Head Pose in Natural Meetings**, IEEE Trans. SMC, Feb 2009.
S. Ba and J.M. Odobez, **Multi-Person Visual Focus of Attention from Head Pose and Meeting Contextual Cues**, PAMI, second revision, 2009.

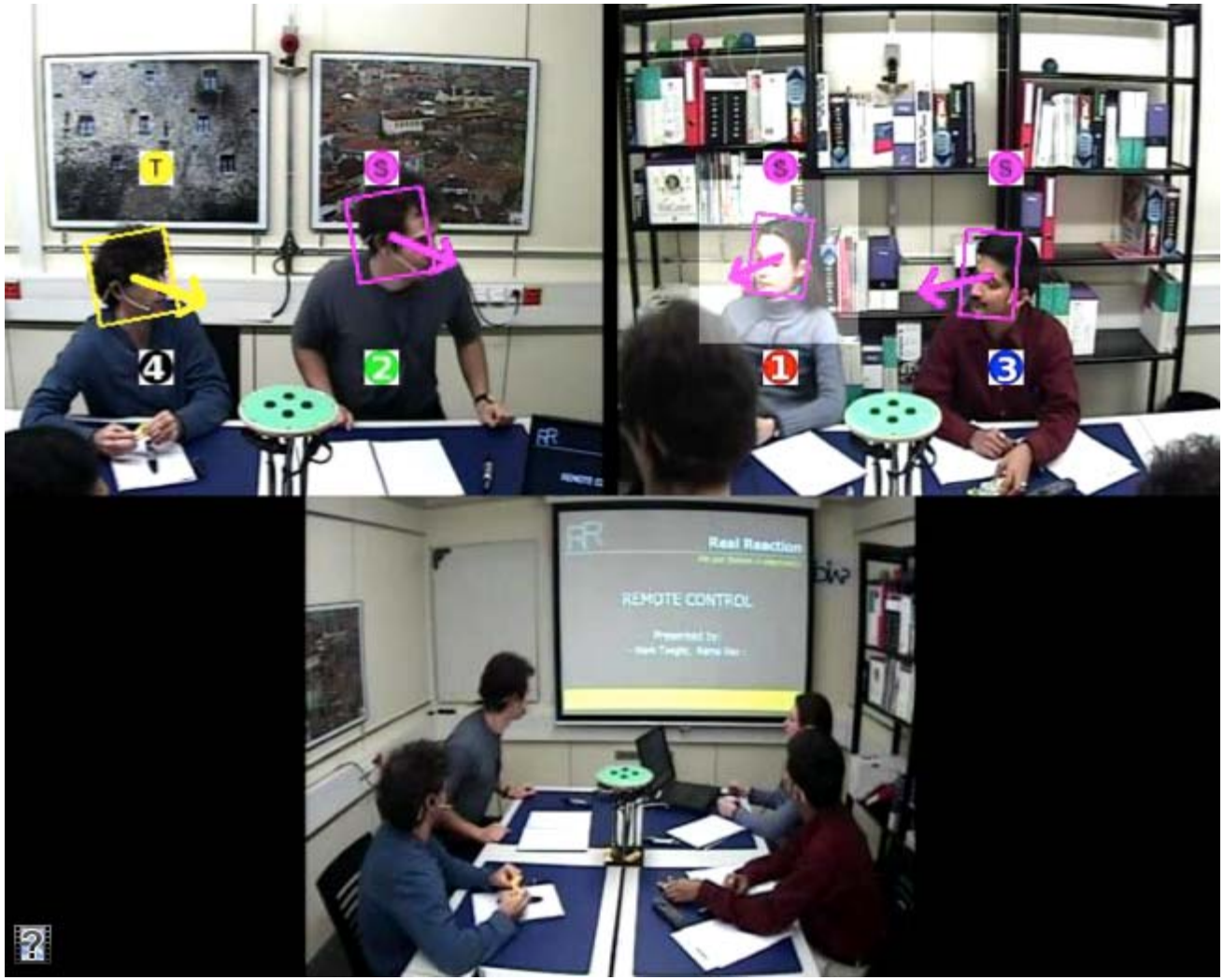
Multimodal multiparty VFOA recognition



Multimodal multiparty VFOA recognition

Notice

- Liveliness, difficulty of data
- person 3 focus changes according to context
  (between looking at person1, slide, standing person)
- slide changes favor looking at slides
- person 4 erroneous VFOA estimation (mainly due to head pose estimation problems)



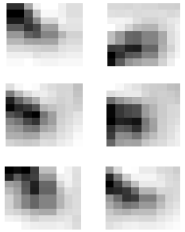
Robust 3D Facial Actions Tracking

Goals : tracking head pose and facial actions on high resolution video sequences

- not limited to near frontal poses
- investigate head activities in natural conversation

Approach

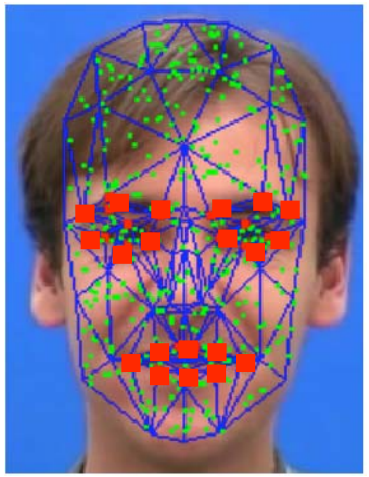
- 3D deformable mesh model
- likelihood modeling of different features
  - **Structural (salient) features**
    local, illumination and view invariant
    model learned from virtual samples
    => stable representation

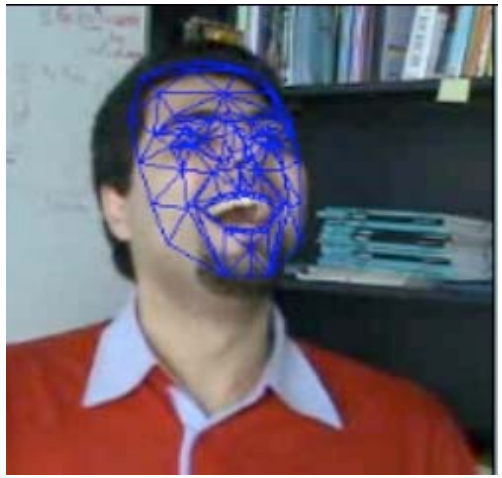


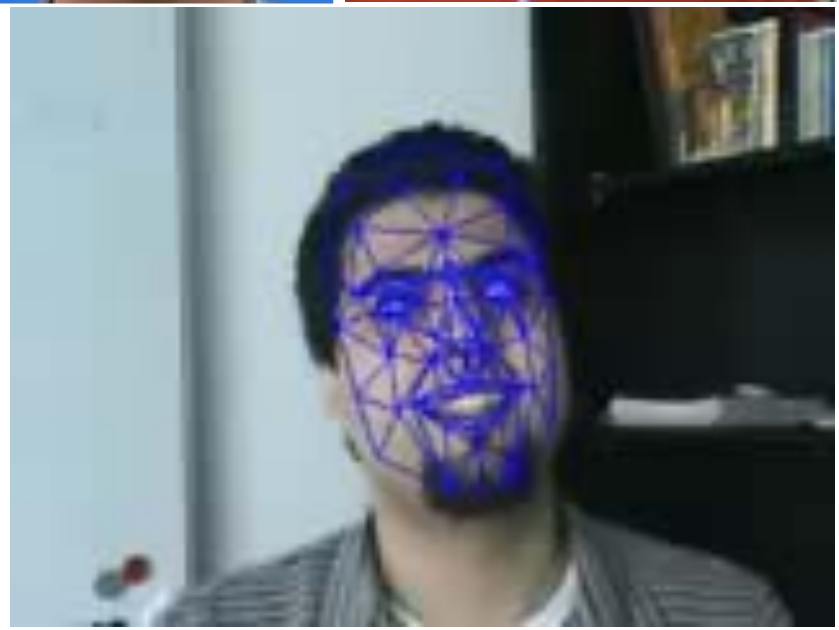
  - **Appearance facial texture**
    dense, illumination sensitive
    online learned view-based templates
    => good for frame-to-frame smooth tracking

Evaluation

- Head pose, Boston Univ. database: 3^o for pan angle; better performance than state of the art trackers
- stability on long sequences (interviews)

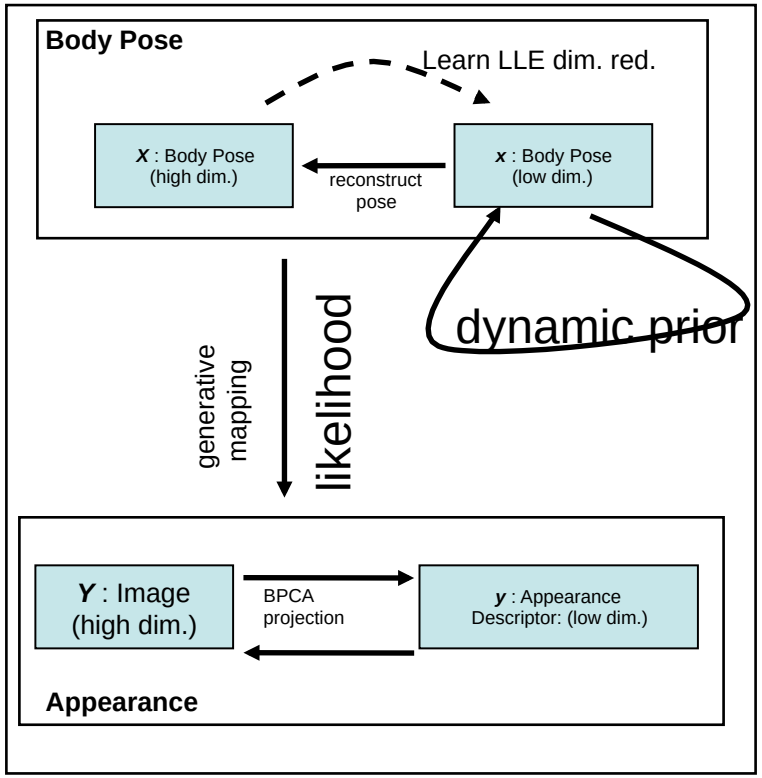






[Lefevre 2009] S. Lefèvre and J.M. Odobez, **Stucture and Appearance Features for Robust 3D Facial Actions Tracking**, Best student paper award at the IEEE ICME 2009 conference.

Monocular Body Pose Estimation (ETHZ)

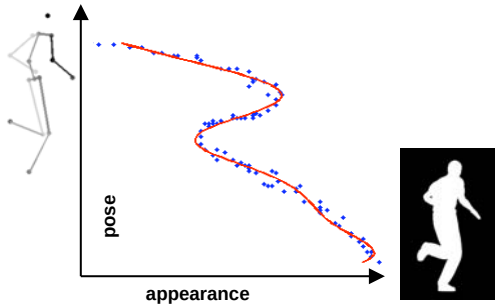


Learning Generative Models for Monocular Body Pose Estimation
T.Jaeggli, E. Koller-Meier, L.Van Gool
ACCV, November 2007
Best paper award

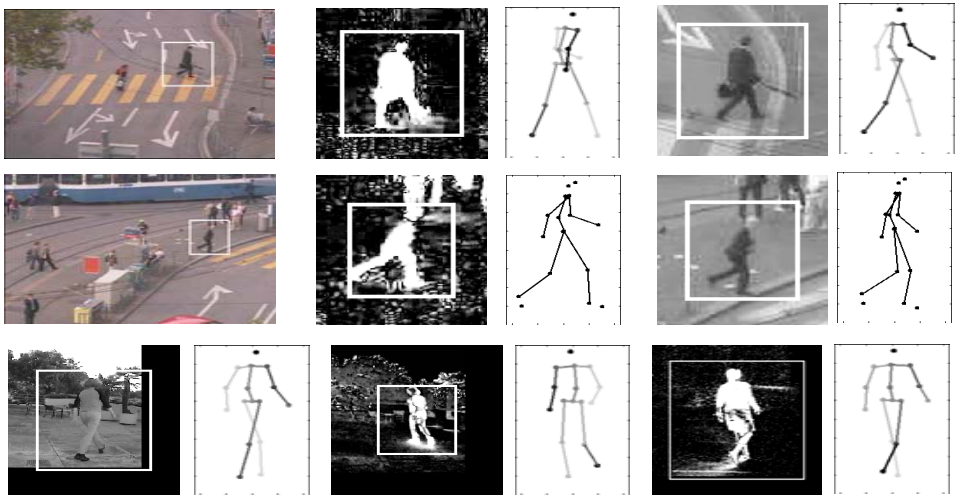
Approach

Problem:
- viewing ambiguities
- non-functional relation appearance-pose
- one-to-many discriminative mapping

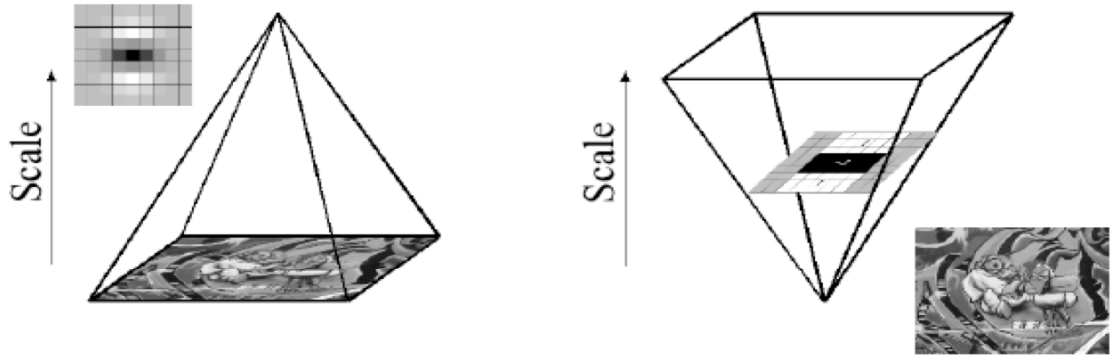
Our Method
- learn generative mapping pose-appearance
- functional relationship
- nonlinear kernel regression (RVM)



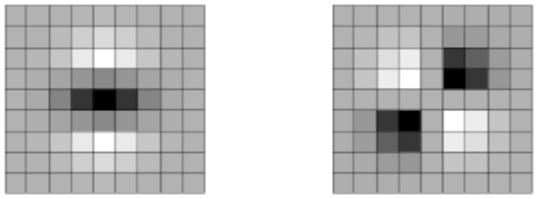
Tracking Human Locomotion



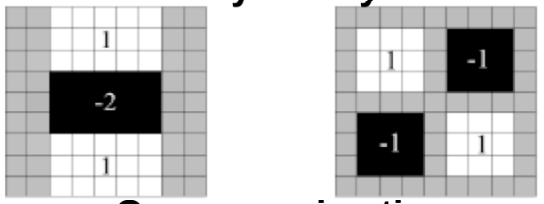
Speeded-Up Robust Features (SURF)



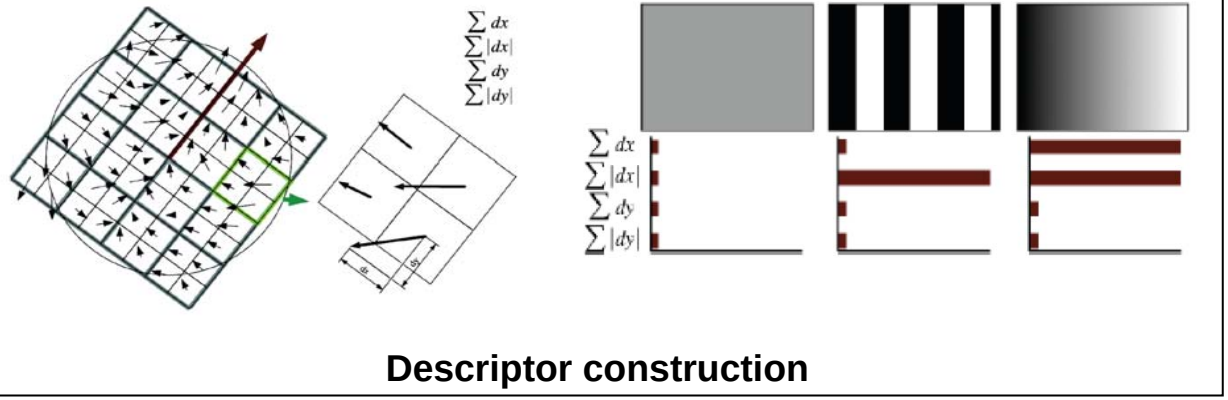
Instead of reducing the image size (left), integral images allow the up-scaling of the filter at constant cost (right).



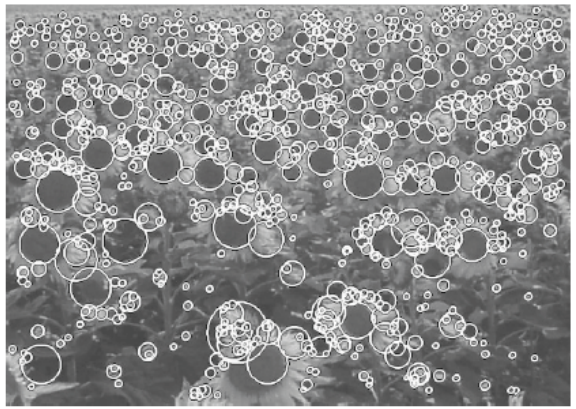
Gaussian second order partial derivative in y and xy-direction



Our approximation



Descriptor construction



Interest points for a sunflower field

SURF: Speeded Up Robust Features,
H.Bay, T. Tuytelaars, L. Van Gool, ECCV 2006
465 citations (google scholar)

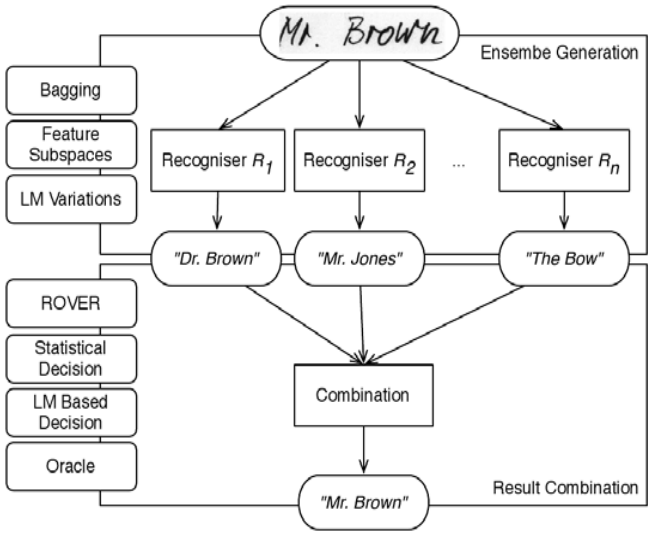
Online Handwriting Recognition (Prof. H. Bunke)

- Recognition of Whiteboard Notes: Online, Offline and Combination
  - Evaluation of Features
  - BLSTM and HMM Recognizers
  - Combining Online and Offline Information
- Text vs. Non-Text distinction in Online Handwritten Documents. (in progress)
  - Segmentation followed by Classification
  - Classification of Individual Strokes

M. Liwicki and H. Bunke. *Recognition of Whiteboard Notes – Online, Offine and Combination*. World Scientific, 2008

Online Handwriting Recognition (Prof. H. Bunke)

- Writer Identification and Verification
  - Evaluation of confidence measures for HMM and GMM based systems
  - Evaluation of feature selection methods
- Ensemble Methods for Handwritten Text Line Recognition
  - Improved recognition via combinations of automatically generated recognizers
- Semi-Supervised Learning for Handwriting Recognition (in progress)
  - Investigating self-learning for HMM and NN-based recognizers

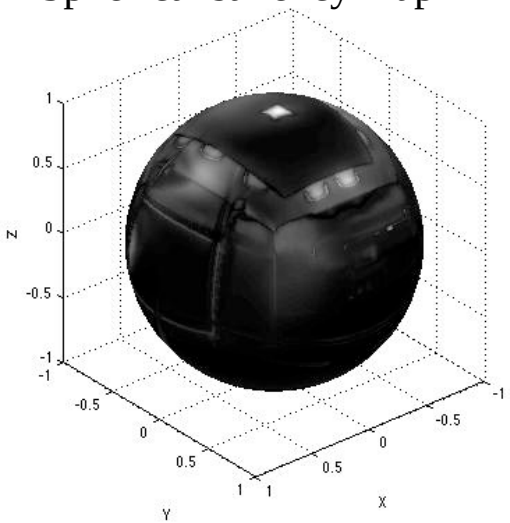


R. Bertolami and H. Bunke. Hidden Markov model based ensemble methods for offine handwritten text line recognition. *Pattern Recognition*, 41(11):3452-3460, 2008

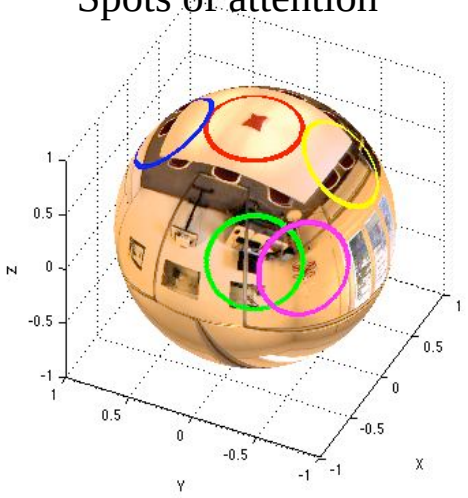
A. Schlapbach, H. Bunke, A Writer Identification and Verification System Using HMM Based Recognizers, *Pattern Analysis & Applications*, 10(1): 33-43, 2007

Visual Attention in Omnidirectional

Spherical saliency map



Spots of attention



MAIN RESULTS (May 2007-May 2009):

- Development of static visual attention algorithm on the sphere with direct application to omnidirectional images:

- Development of dynamic visual attention for detecting spots of attention in omnidirectional video

References:

«Visual Attention on the Sphere», IEEE TIP, November 2008,

« Dynamic Attentive System for Omnidirectional Video», PCS, May 2009

IM2.VP - conclusions

- IM2.VP developed and provided new methods and tools for visual information analysis
  - Face image processing
  - VFOA
  - Gesture
  - Object recognition
  - Handwriting
- Largely applied to IM2 corpus and other reference datasets
- Contributed in enriching the IM2 corpus
- Integrated in multimodal analysis (IM2.MPR)
- Impressive publication record
  - Journal papers (IEEE TIP, IEEE PAMI, etc)
  - Conference papers
- Spin-off companies
  - Kooaba
  - nViso
  - …