Random forests for facial expression analysis

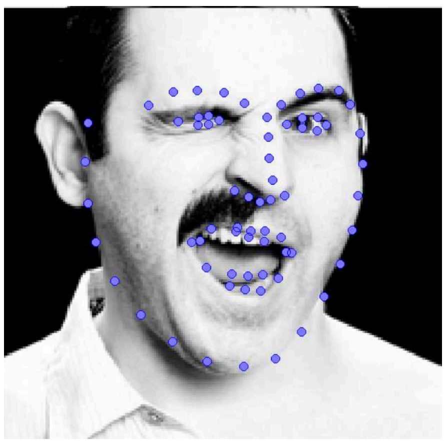
Kévin Bailly (ISIR UPMC)
Joint work with **A. Dapogny** & **S. Dubuisson**

SMART School -- 5/9/2017

What are we trying to measure?

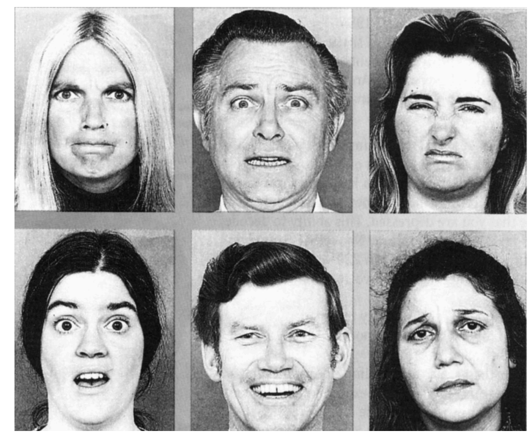
Objective

Subjective



Facial landmark localisation

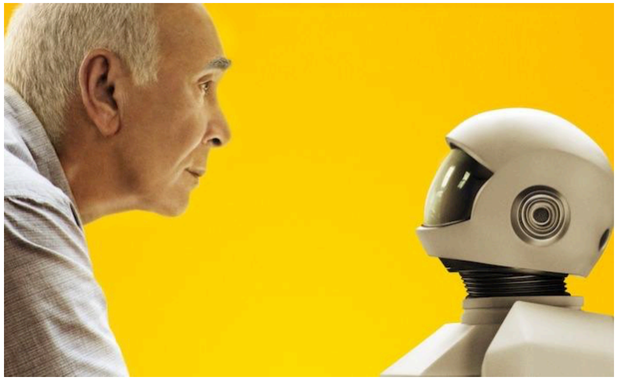
Facial Action Coding System (FACS)



Ekman prototypical emotions

Applications

Human-Machine Interaction



Safety

Medical

Marketing

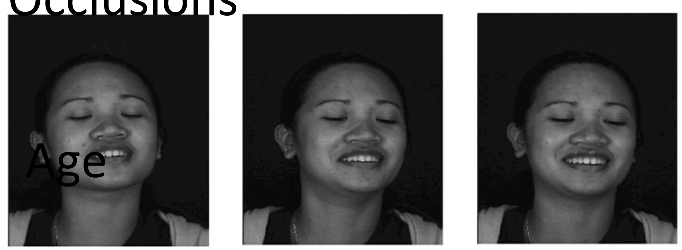
Motion capture

Challenges

Data complexity

Occlusions

Age

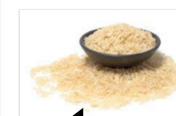


1 2 3

Head pose

Precision of the tasks

needed data



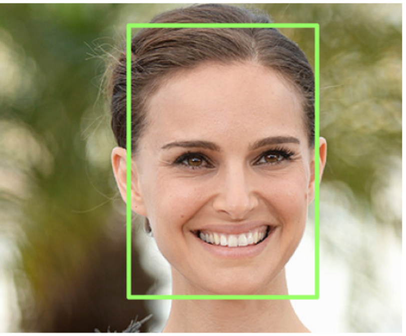
available data

Limited amount of data

Presentation Outline

- **Classical Pipeline**
- Random forests for FER
- Pairwise conditional random forests
- Local subspace random forests
- Conclusion and perspectives

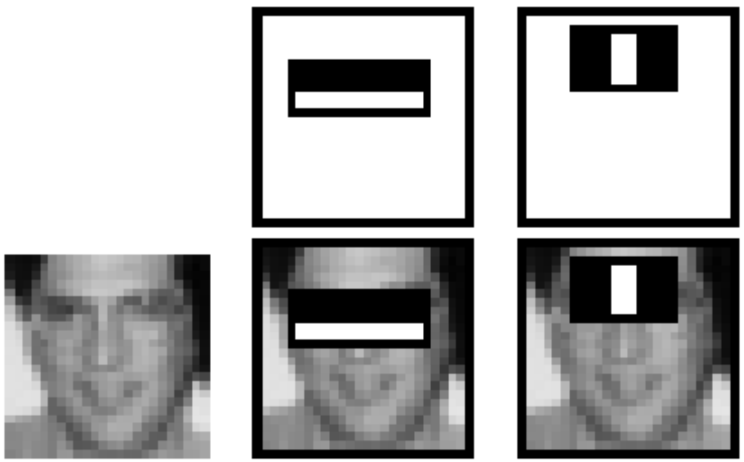
Classical Pipeline



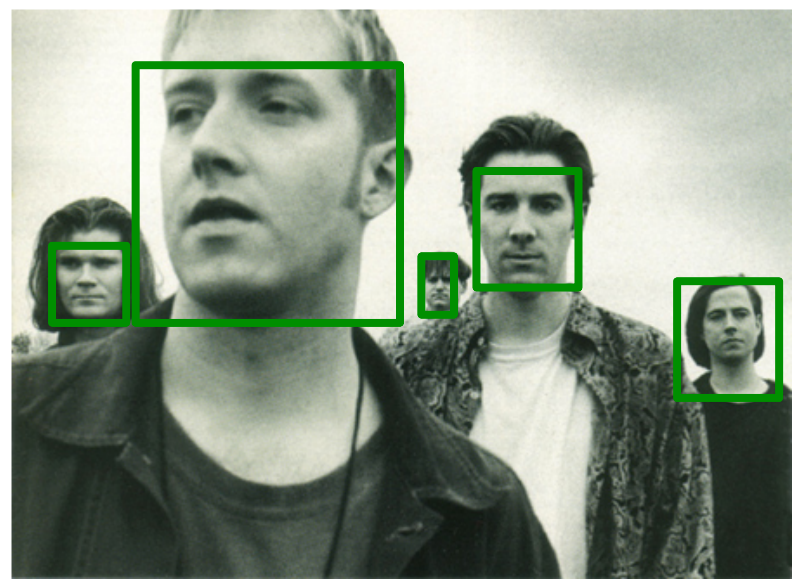
Face detection

Face detection

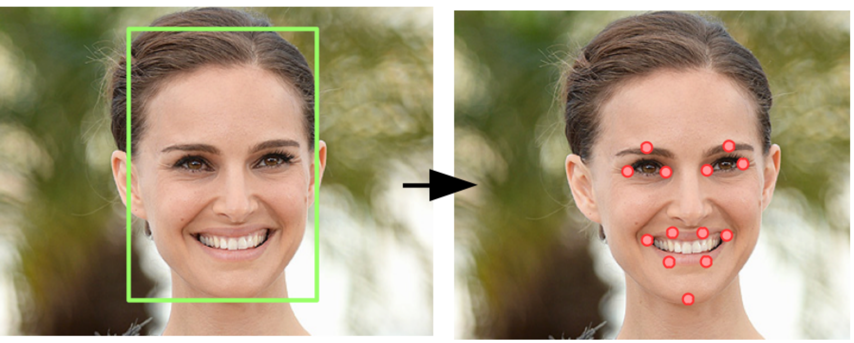
- Viola & Jones Algorithm [Viola 2001]
  - Multi-scale sliding windows
  - Haar like features + AdaBoost
  - Accurate and Real time
- Widely used (OpenCV, Matlab vision tbx…)



$$F(x) = \alpha_1 h_1(x) + \alpha_2 h_2(x) + \ldots$$



Classical Pipeline

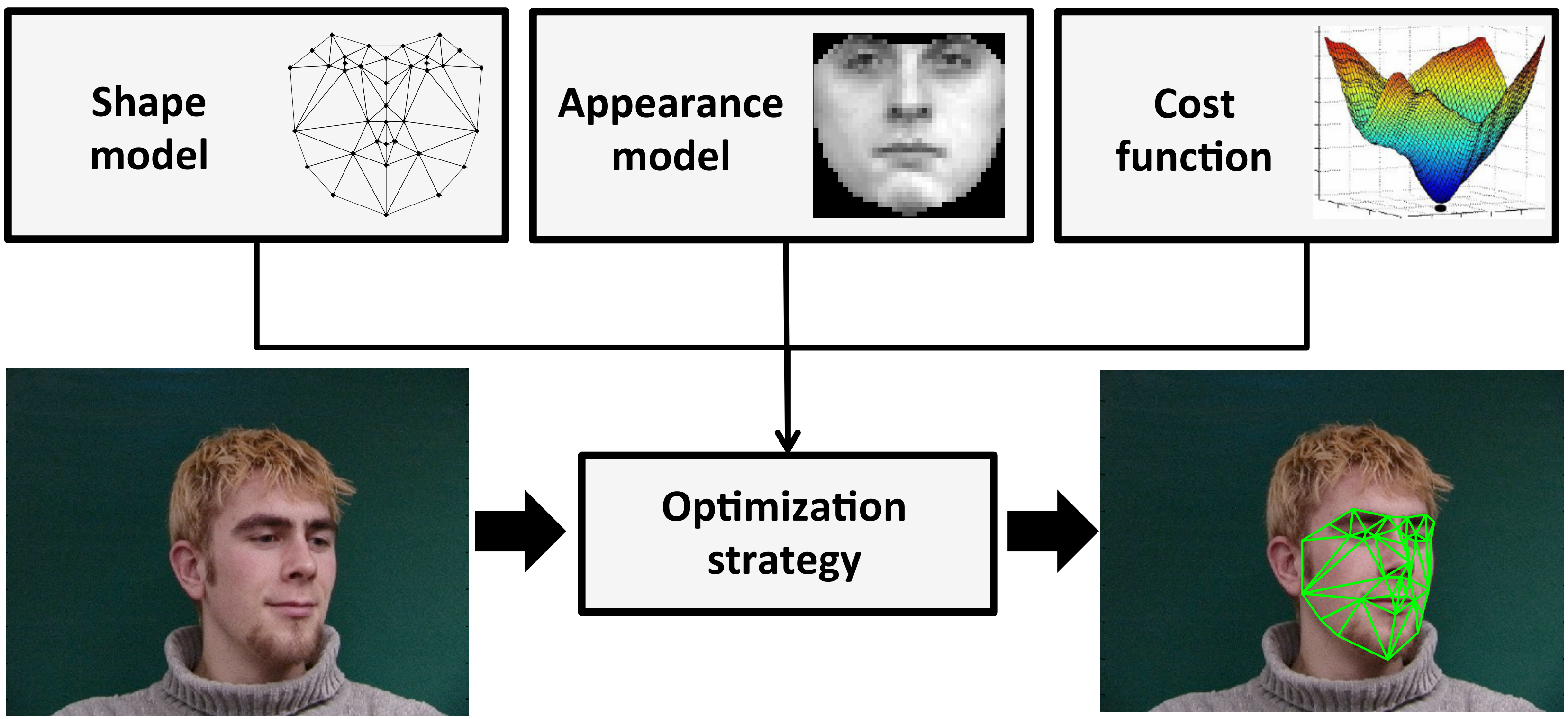


Face detection

Feature point alignment

Feature point alignment

- Main parts of feature point alignment techniques

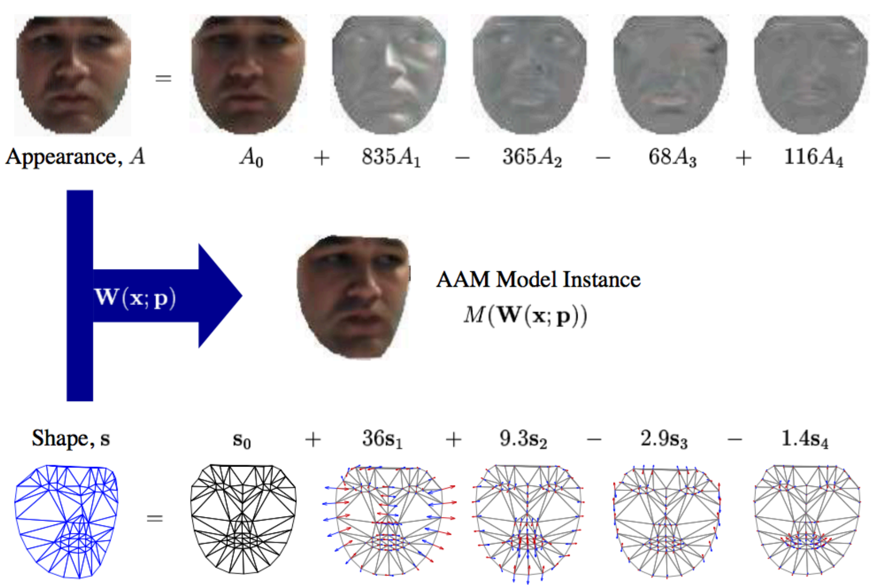


Input image

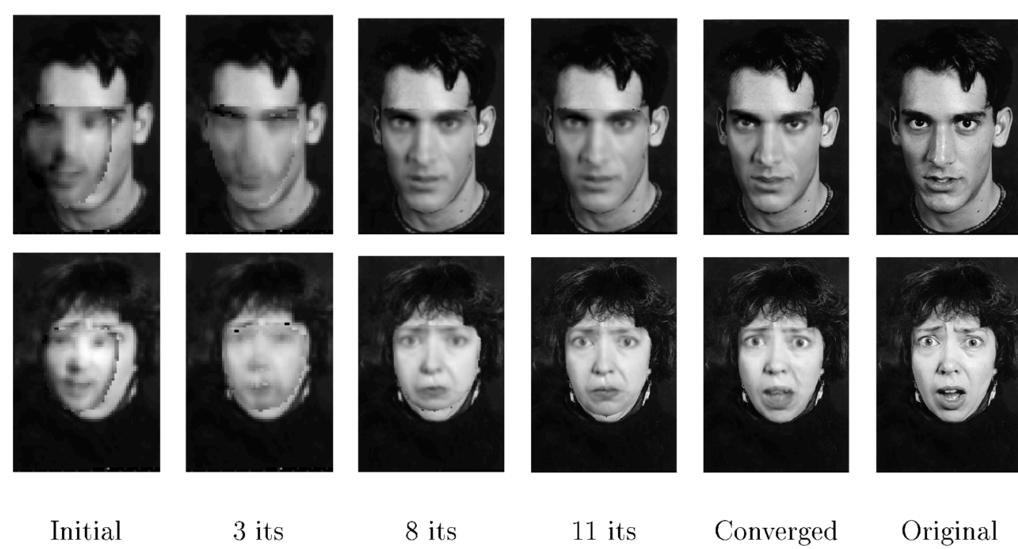
Aligned model

Feature point alignment

- Active Appearance Models (AAM)
  - Parametric models of shape and appearance
  - Estimation of parameters that best describe the current instance of the face in the image
  - Lot of improvements since 2001



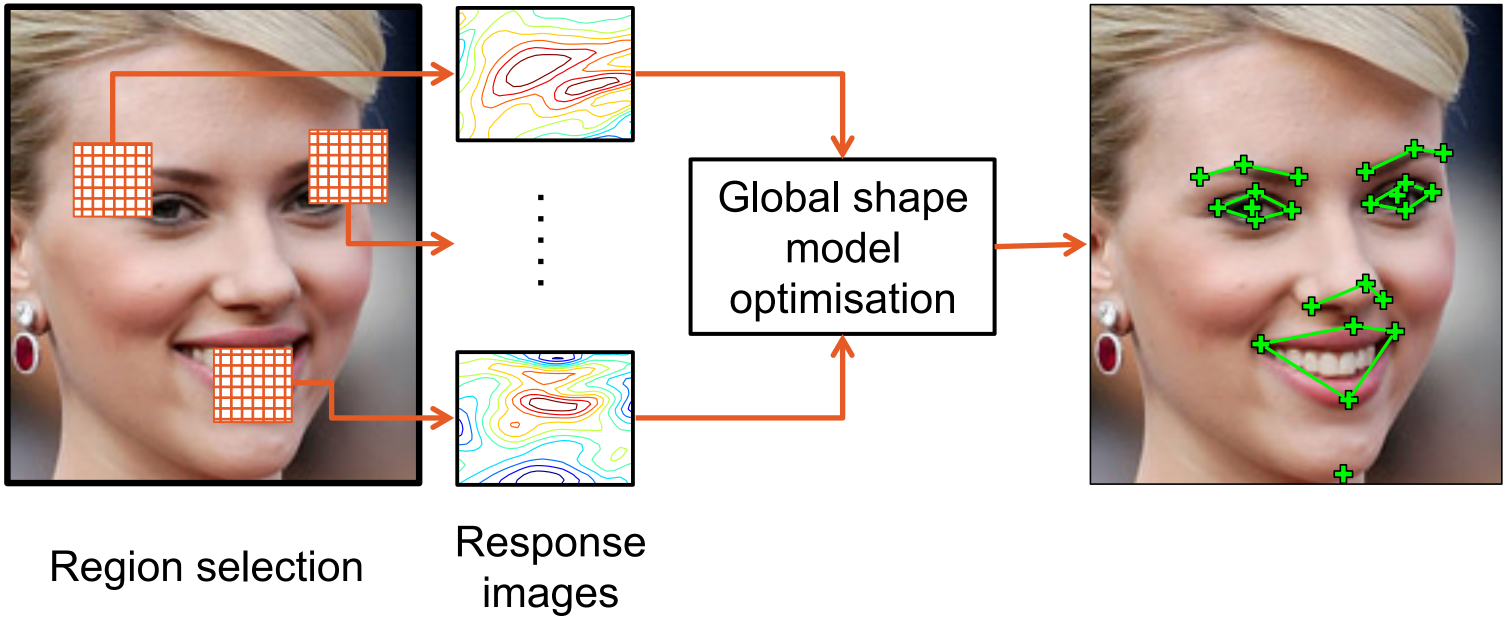
[Matthews et al. 04]



[Cootes et al. 01]

Feature point alignment

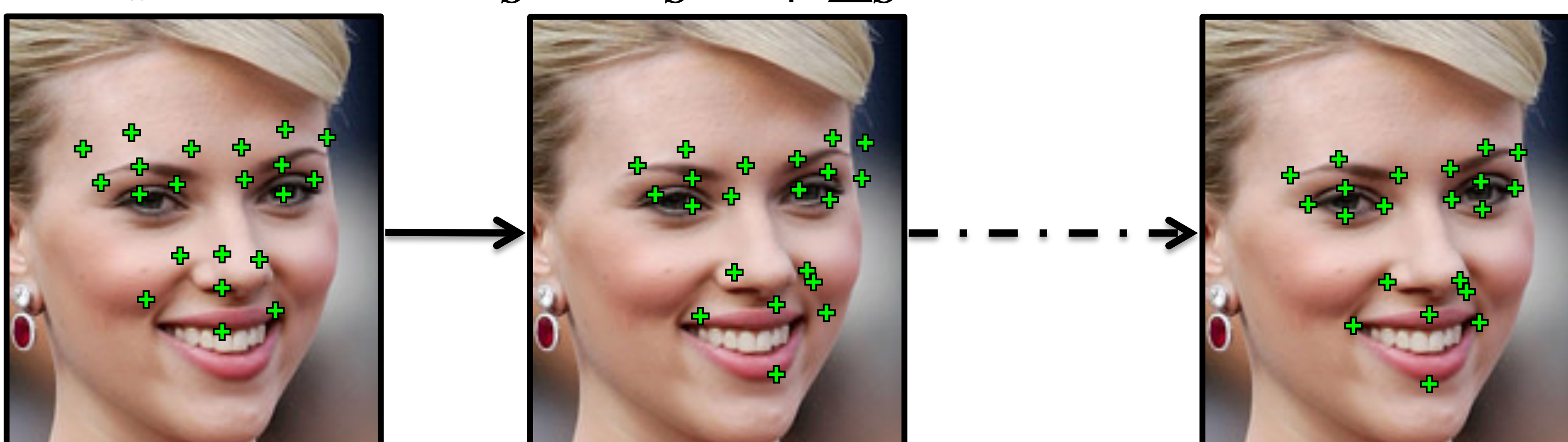
- Constrained Local Models (CLM)
  - Based on local texture analysis around landmarks



Feature point alignment

- Cascaded regression-based techniques
  - State of the art results
  - Iteratively update the landmark location (cascade of regressors)
  - Require a lot of training samples

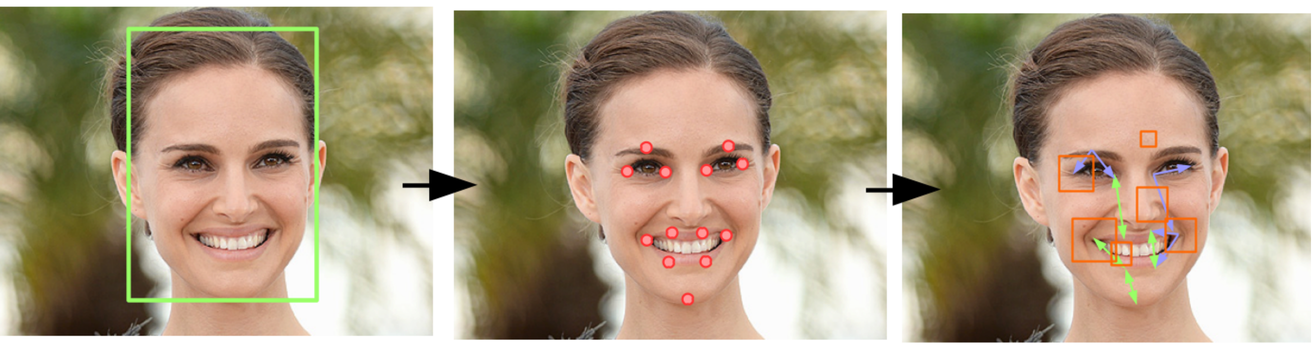
$s^{(0)}$ $\qquad$ $s^{(1)} = s^{(0)} + \Delta s^{(1)}$ $\qquad$ $s^{(n)} = s^{(n-1)} + \Delta s^{(n)}$



Recent results

[arxiv 17]
[PRL (under review)]

Classical Pipeline



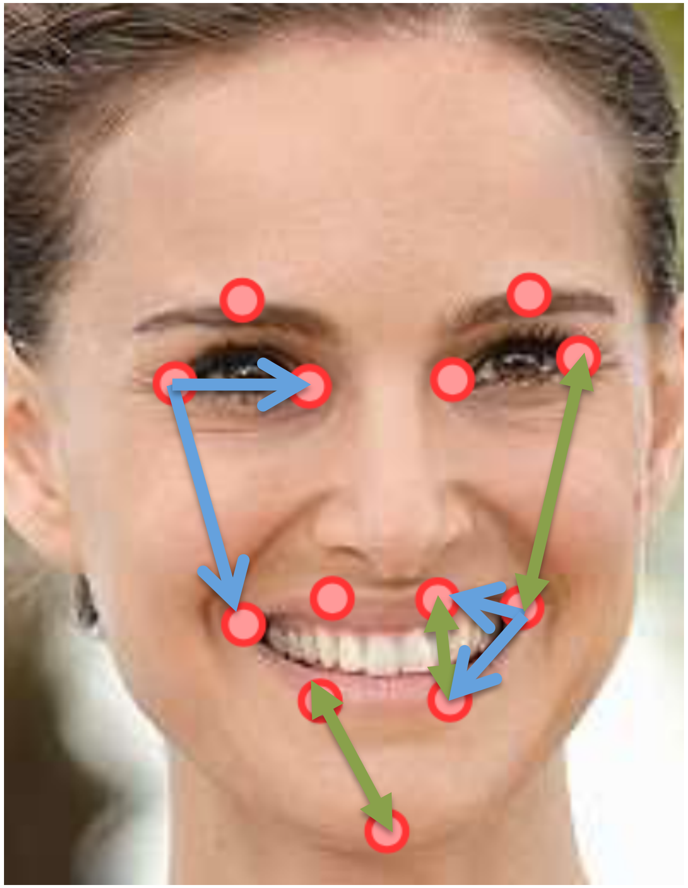
Face detection

Feature point alignment

Feature extraction

Feature extraction

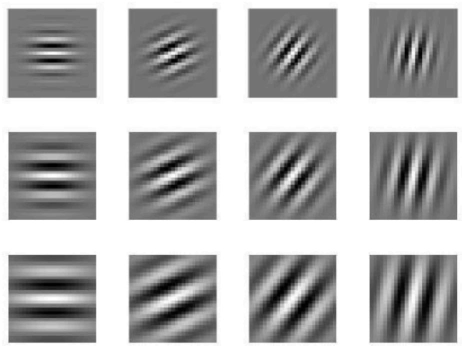
- Geometric features
  - normalized **distance** between feature points
  - **angle** between triplets of points



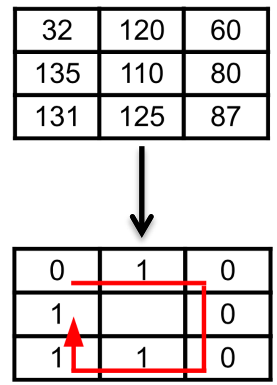
Feature extraction

- Geometric features
- Appearance features

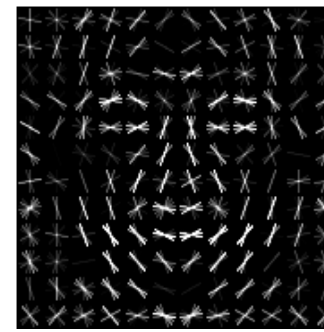
Raw pixels



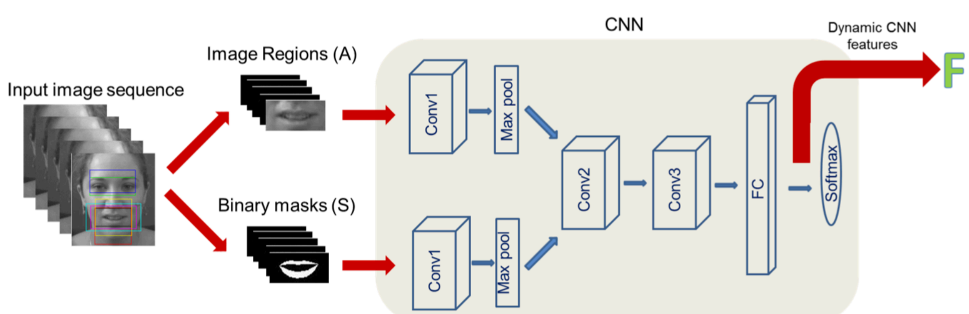
Gabor filter bank



Local Binary Patterns



SIFT / HOG

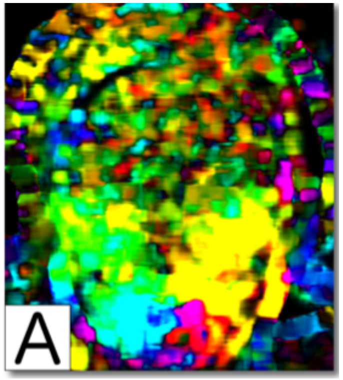


Deep learning features [Jaiswal 16]

Feature extraction

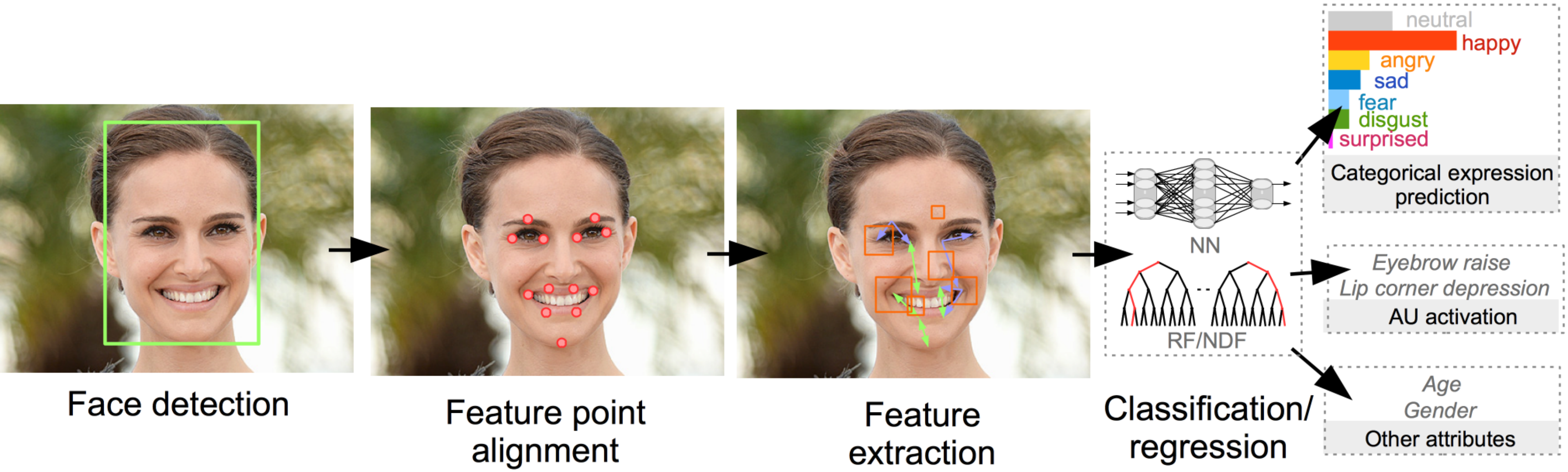
- Geometric features
- Appearance features
- Temporal features

Motion History Image (MHI)
(e.g. [Koelstra et al. 10])



(dense) optical flow
(e.g. [Allaert et al. 17])

Classical Pipeline



Contributions

Challenges	Method	Task		
		Landmarks	Action Units	Emotions
Heterogenous Features	MK-SVM	[FG 11] [IVC 13]	[FERA 11] 1st (with Supelec)	[PAA 14]
Limited amount of data	HMT-MLKR	[These Nicolle]	[FERA 15] 1st [IVC 16]	
Temporal information	PCRF			[ICCV 15] [CVPR 16 demo]
Head pose	MV-PCRF			[TAC 17]
Occlusions	LS-RF		[IJCV 17]	
Online Learning	NDF	[Arxiv. 17] [PRL submitted]		

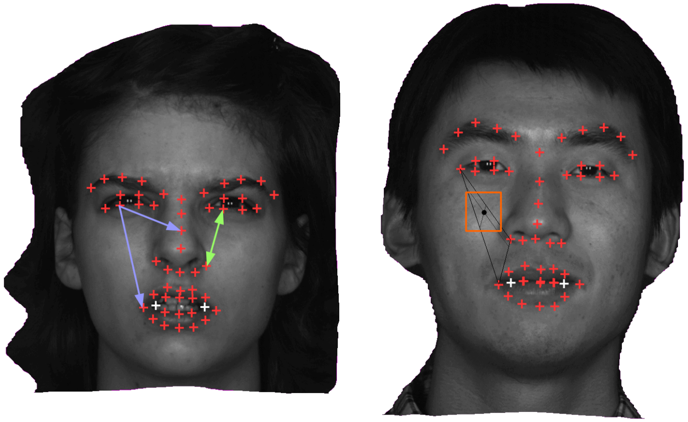
Presentation Outline

- Classical Pipeline
- **Random forests for FER**
- Pairwise conditional random forests
- Local subspace random forests
- Conclusion and perspectives

Random Forests for FER

Features from heterogeneous templates

- normalized **distance** between feature points
- **angle** between triplets of points
- integral **HoG** from barycentric coordinates

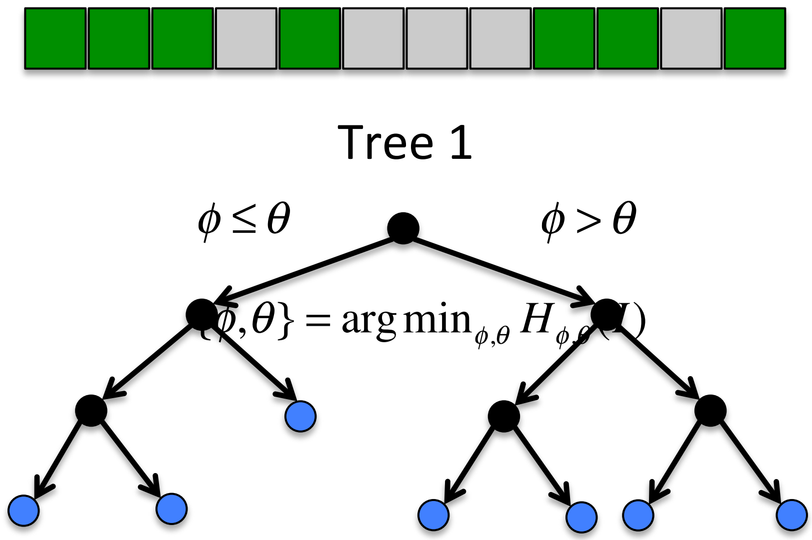


Constraints on the ML framework:

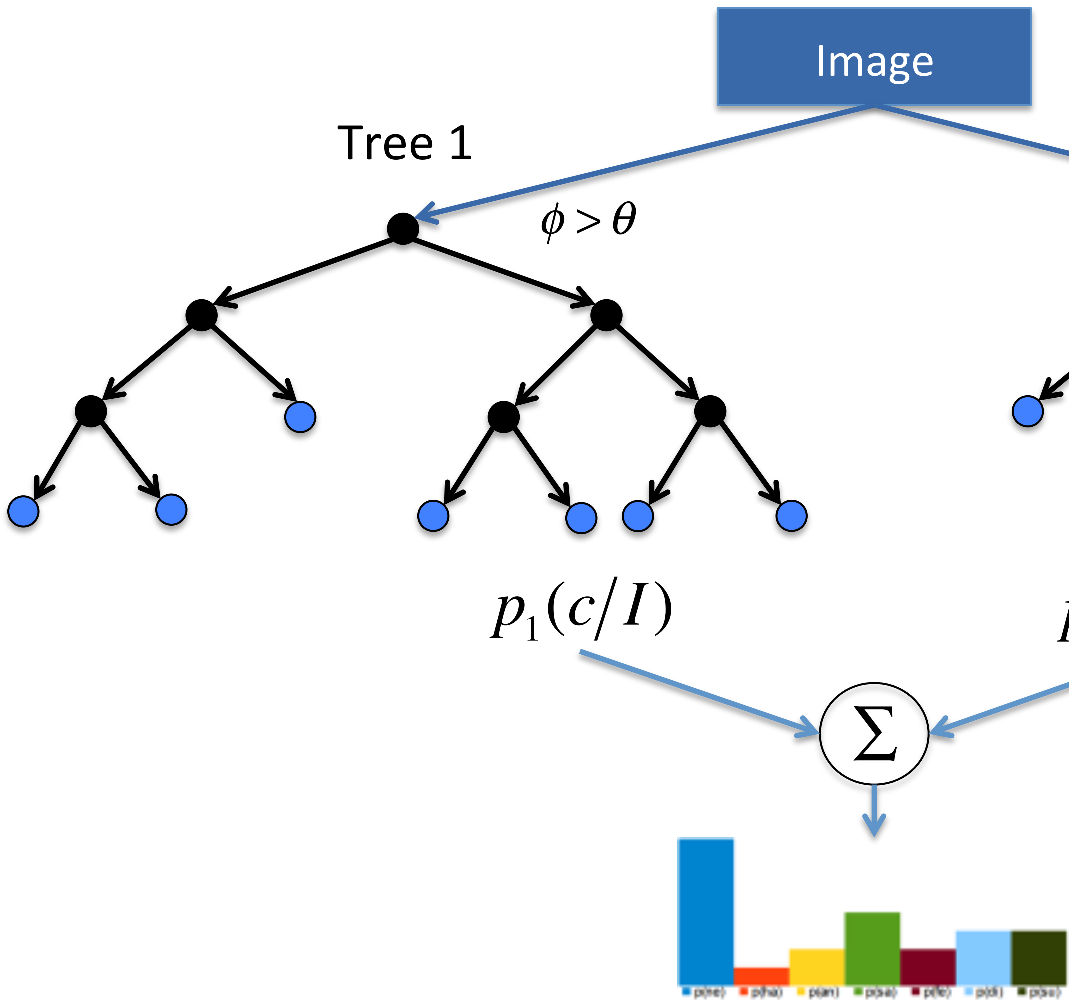
- **Very high dimensional**, potentially noisy data
- Generalize well to **unseen morphologies** and contexts
- Recognition in **real-time**

➔ **Random Forests (RF)**

Random Forests for FER

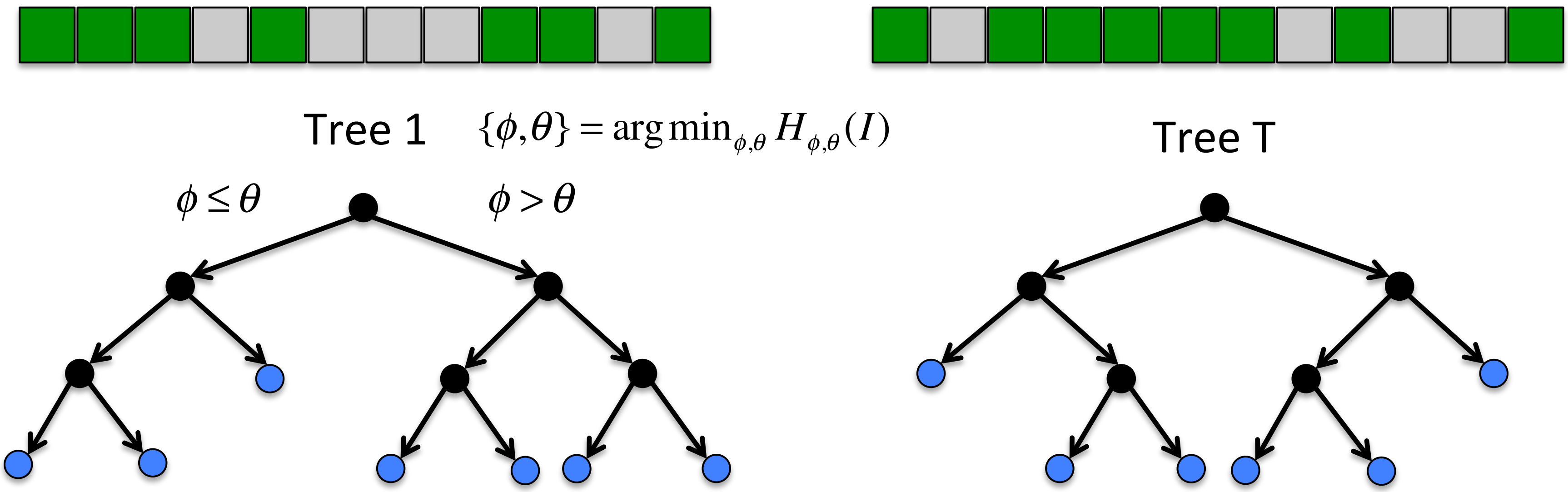


Random Forests for FER



Votes are averaged among all trees in the forest

Random Forests for FER

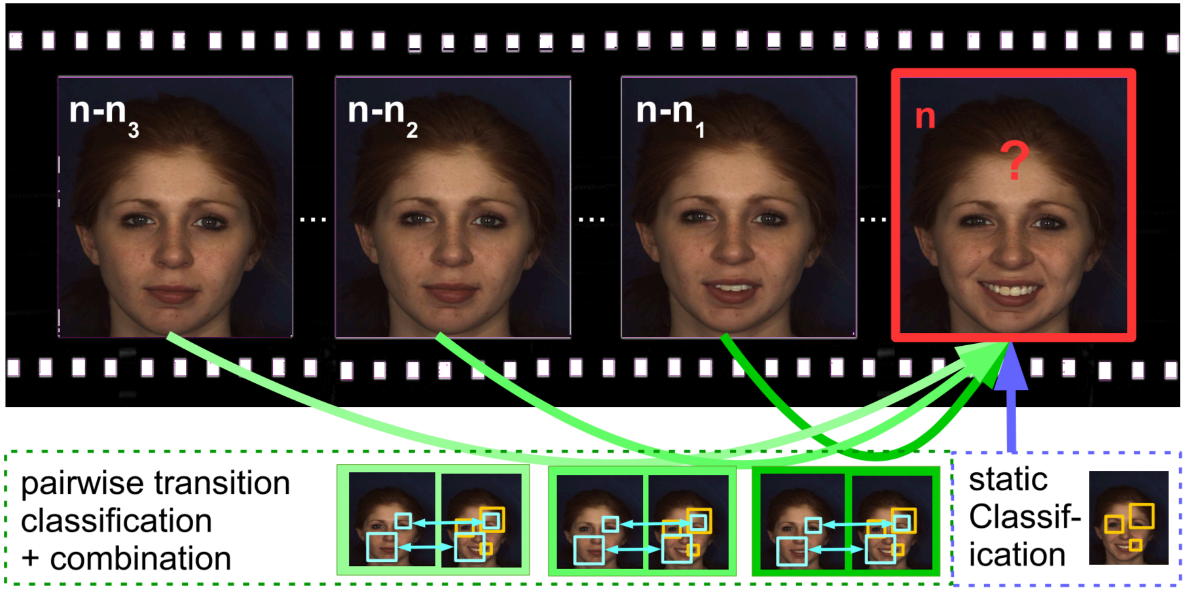


- **Problems**
  - Dynamic of the expression is not used
  - Robustness w.r.t. head pose variation is not ensured
  - Presence of occlusions can dramatically affect the performance

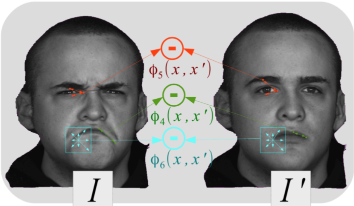
Presentation Outline

- Classical Pipeline
- Random forests for FER
- **Pairwise conditional random forests**
  - **PCRF for Transition pattern modelling**
  - Multi-view extension of PCRF
  - Experiments
- Local subspace random forests
- Conclusion and perspectives

PCRF for Transition pattern modelling



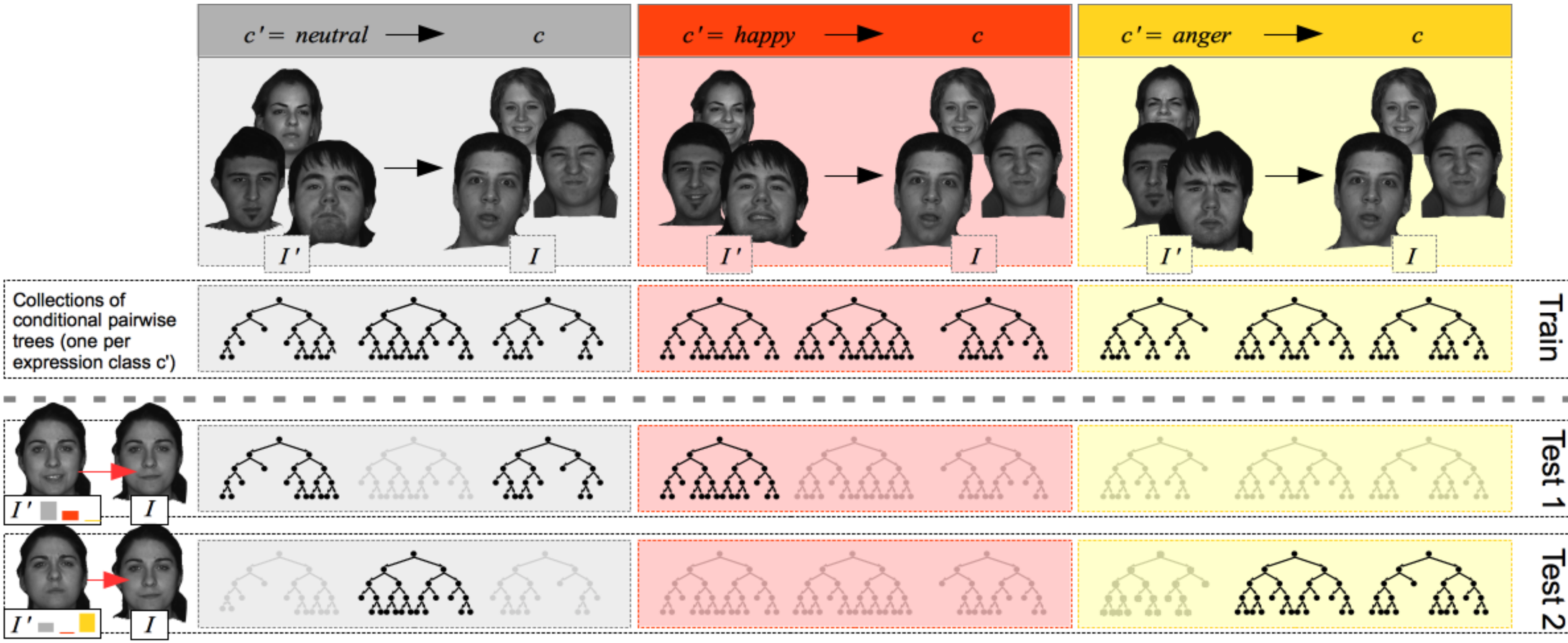
- Trade-of between strength and decorrelation of trees
- Key ideas:
  1) **increase strength** with pairwise derivatives feature templates
     – static features
     – pairwise derivatives of static features



  2) **increase decorrelation** by considering multiple transitions

A Dapogny, K Bailly, S Dubuisson. *Pairwise Conditional Random Forests for facial expression recognition*. **ICCV** 2015

PCRF for Transition pattern modelling



Example pairwise query of a PCRF

A Dapogny, K Bailly, S Dubuisson. *Pairwise Conditional Random Forests for facial expression recognition*. **ICCV** 2015

PCRF for Transition pattern modelling

Static RF classification

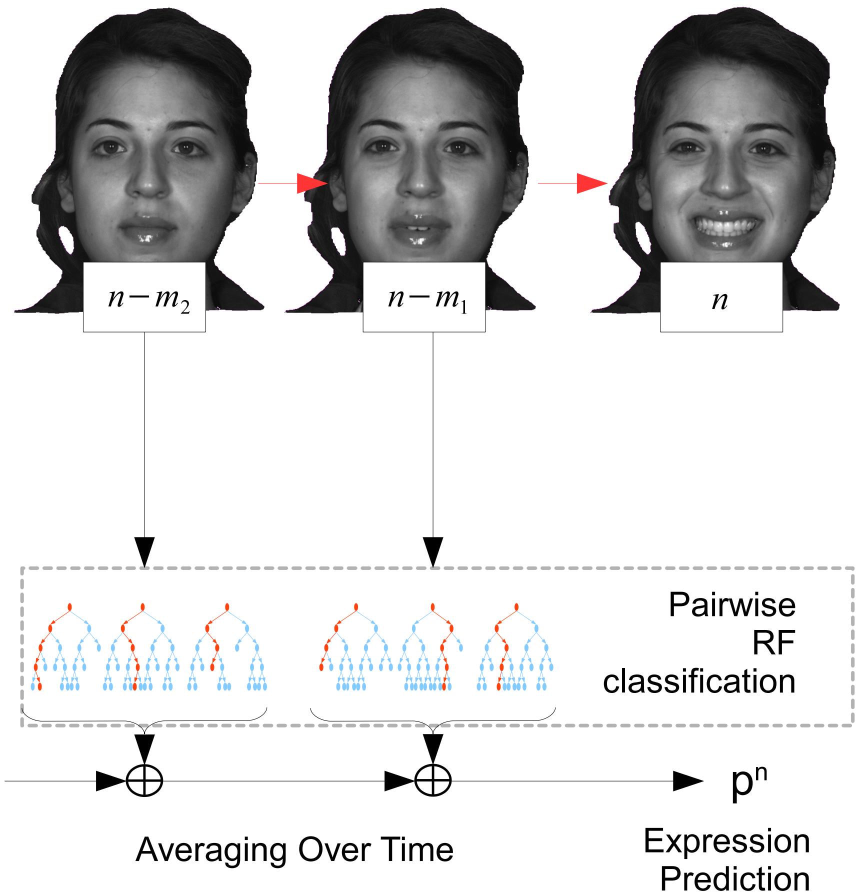
p^n

Expression Prediction

$$p^n(c) = \frac{1}{T} \sum_{t=1}^{T} p_t(c|\mathcal{I}^n)$$

A Dapogny, K Bailly, S Dubuisson. *Pairwise Conditional Random Forests for facial expression recognition*. **ICCV** 2015

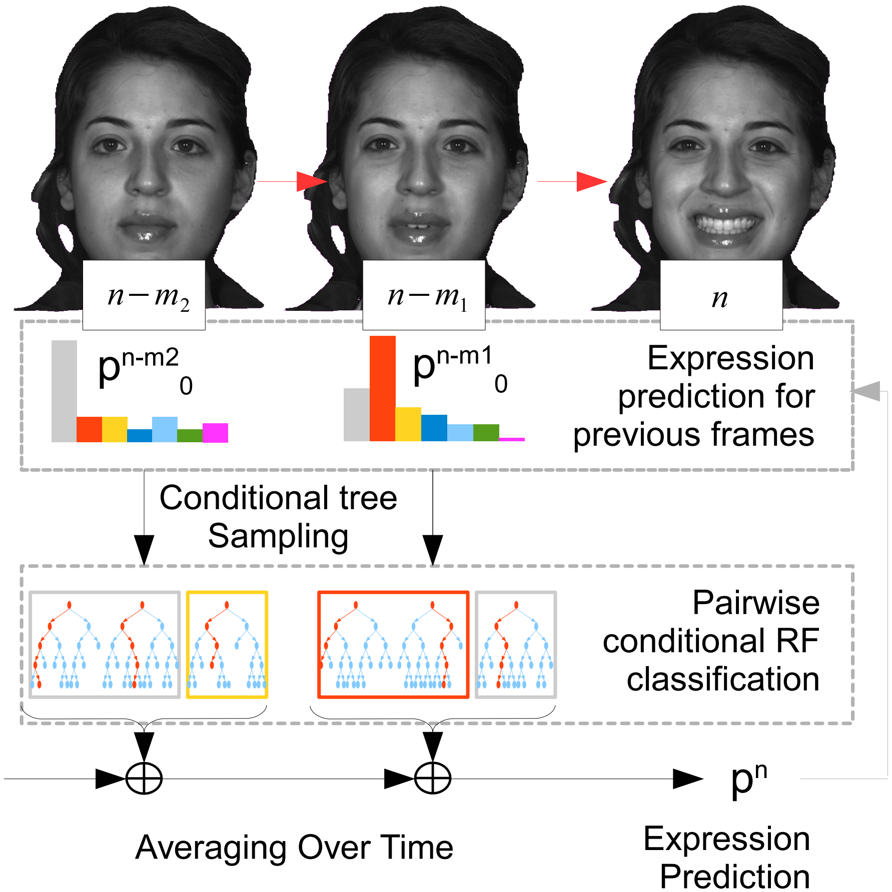
PCRF for Transition pattern modelling



$$p^n(c) = \frac{1}{NT} \sum_{m=n-1}^{n-N} \sum_{t=1}^{T} p_t(c|\mathcal{I}^m, \mathcal{I}^n)$$

A Dapogny, K Bailly, S Dubuisson. *Pairwise Conditional Random Forests for facial expression recognition*. **ICCV** 2015

PCRF for Transition pattern modelling



$$p^n(c) = \frac{1}{NT} \sum_{m=n-1}^{n-N} \sum_{c' \in \mathcal{C}} \sum_{t=0}^{\mathcal{N}(c')} p_t(c|\mathcal{I}^m, \mathcal{I}^n, c')$$

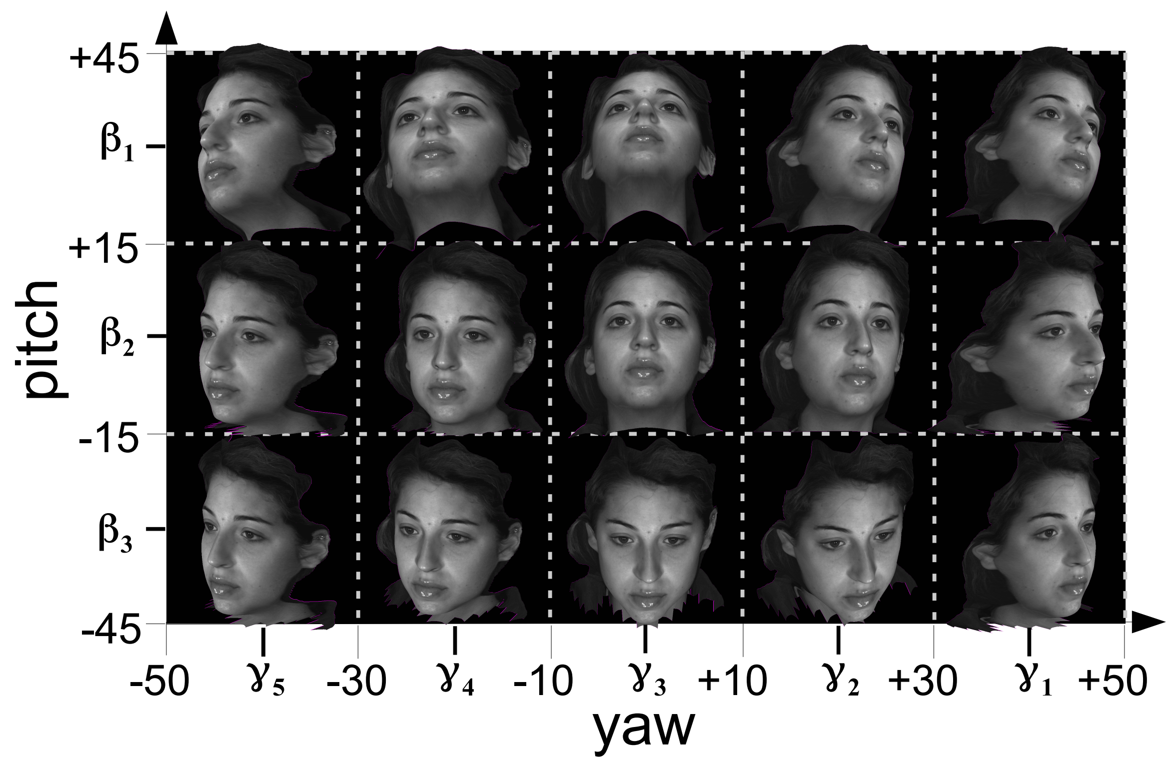
A Dapogny, K Bailly, S Dubuisson. *Pairwise Conditional Random Forests for facial expression recognition*. **ICCV** 2015

Presentation Outline

- Classical Pipeline
- Random forests for FER
- **Pairwise conditional random forests**
  - PCRF for Transition pattern modelling
  - **Multi-view extension of PCRF**
  - Experiments
- Local subspace random forests
- Conclusion and perspectives

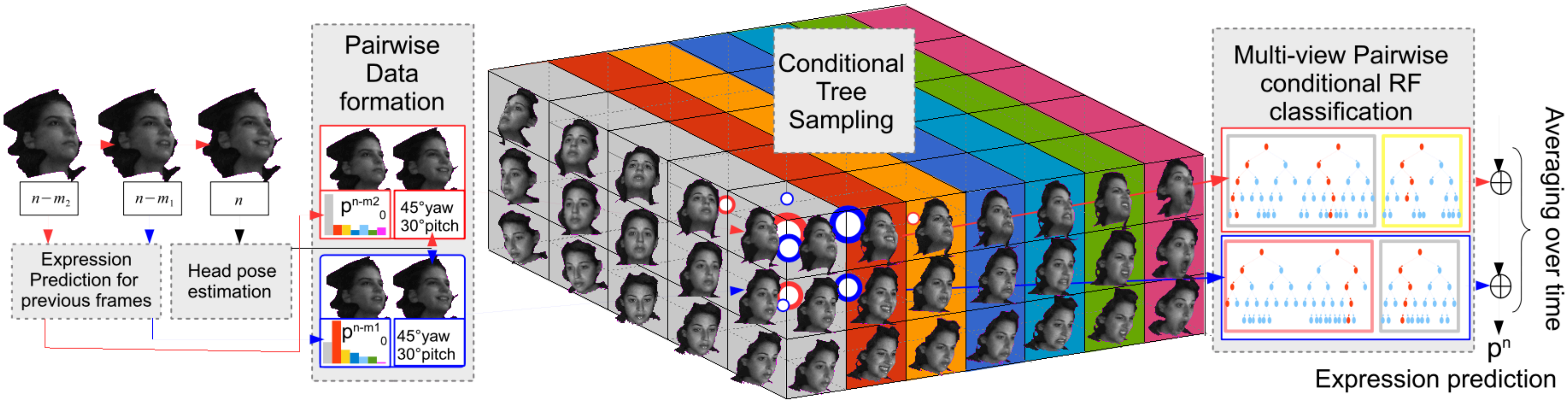
Multi-view extension of PCRF

- Conditioning on head pose estimate $\omega(\mathcal{I}^n) \in \Omega$
- Pose space quantized in 15 pose bins $\{\Omega_i = \Omega_{\gamma_i,\beta_i}\}_{i=1,\dots,k}$



A Dapogny, K Bailly, S Dubuisson *Dynamic Pose-Robust Facial Expression Recognition by Multi-View Pairwise Conditional Random Forests*. **IEEE Trans. on Aff. Comp.** 2017

Multi-view extension of PCRF



Static Multi-view model (MVRF)

$$p^n(c) = \frac{1}{T} \sum_{\Omega_i \in \Omega} \sum_{t=1}^{\mathcal{N}(\Omega_i)} p_t(c|\mathcal{I}^n, \Omega_i)$$

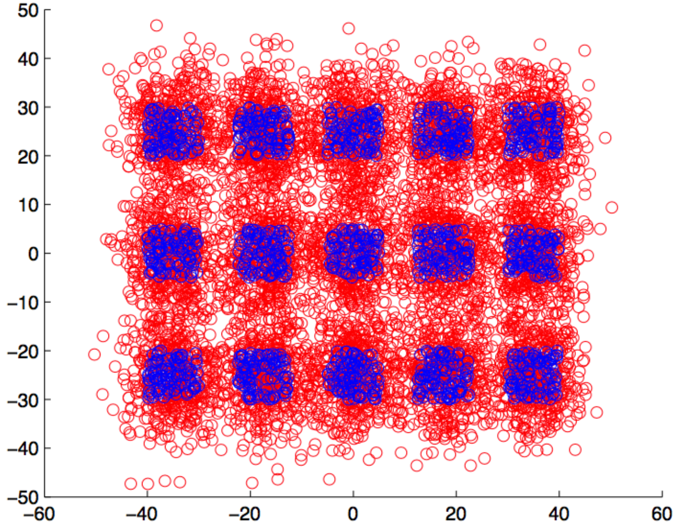
Multi-view PCRF

$$p^n(c) = \frac{1}{T} \sum_{m=n-1}^{n-N} \sum_{\Omega_i \in \Omega} \sum_{c' \in \mathcal{C}} \sum_{t=1}^{\mathcal{N}(c',\Omega_i)} p_t(c|\mathcal{I}^n, \mathcal{I}^m, \Omega_i, c')$$

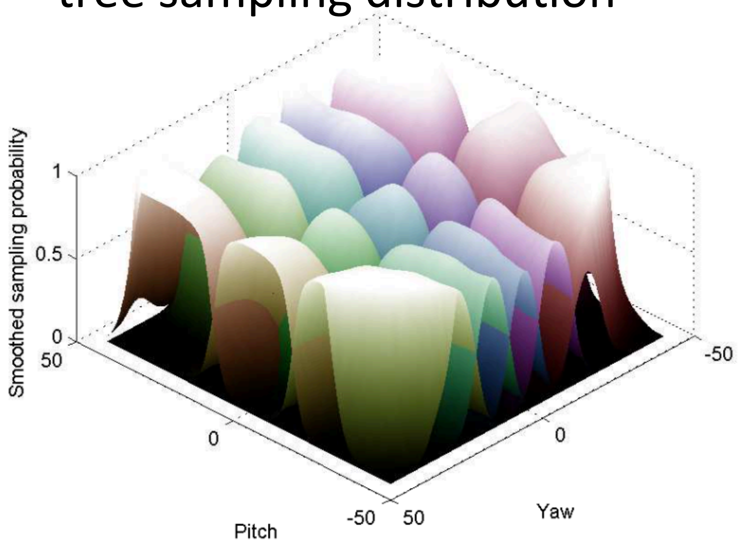
A Dapogny, K Bailly, S Dubuisson *Dynamic Pose-Robust Facial Expression Recognition by Multi-View Pairwise Conditional Random Forests*. **IEEE Trans. on Aff. Comp.** 2017

Multi-view extension of PCRF

data repartition



tree sampling distribution



- Non-frontal head poses generated from high-resolution 3D face scans
- Useful pose range of ±45 deg yaw, ±30 deg pitch
- Tree sampling distribution constructed from the data
- 122623 images for training, 906030 images for test

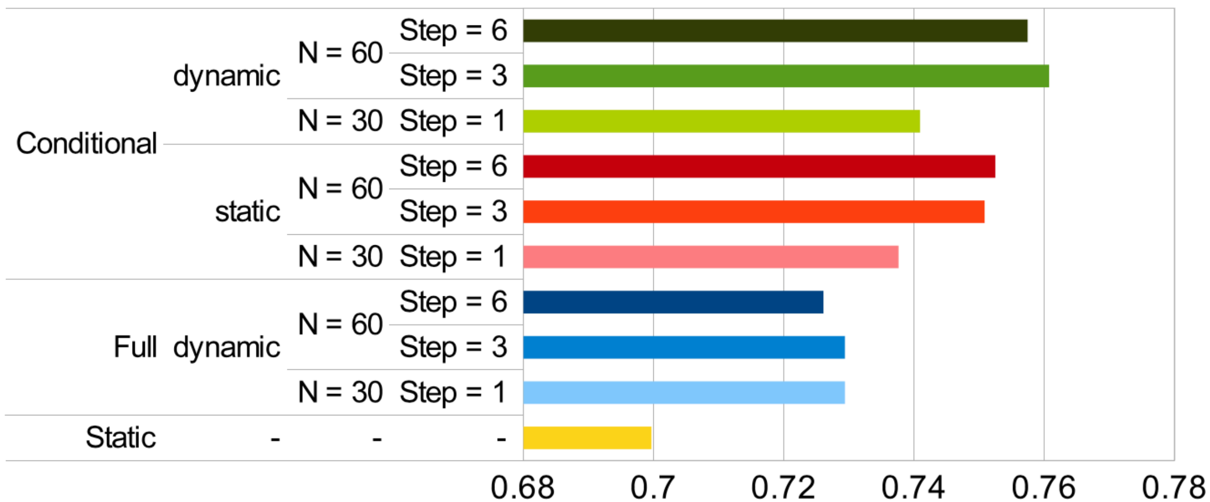
A Dapogny, K Bailly, S Dubuisson *Dynamic Pose-Robust Facial Expression Recognition by Multi-View Pairwise Conditional Random Forests*. **IEEE Trans. on Aff. Comp.** 2017

Presentation Outline

- Classical Pipeline
- Random forests for FER
- **Pairwise conditional random forests**
  - PCRF for Transition pattern modelling
  - Multi-view extension of PCRF
  - **Experiments**
- Local subspace random forests
- Conclusion and perspectives

Experiments

- Varying the temporal integration parameters (in frames, at 30 fps) on BU-4DFE database

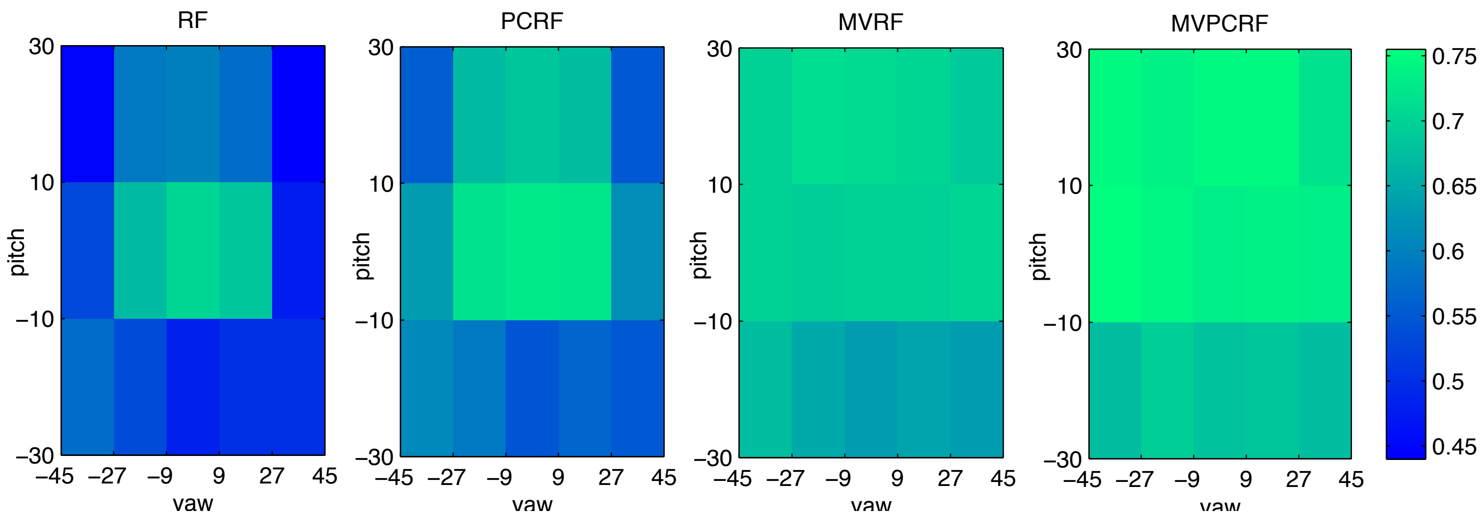


- Static < Full < Conditional
- Accuracy is better when we look at more frames with less correlation between them

A Dapogny, K Bailly, S Dubuisson *Dynamic Pose-Robust Facial Expression Recognition by Multi-View Pairwise Conditional Random Forests*. **IEEE Trans. on Aff. Comp.** 2017

Experiments

- Robustness w.r.t. head pose variations



- Runtime evaluation for 500 trees: 2.4 ms

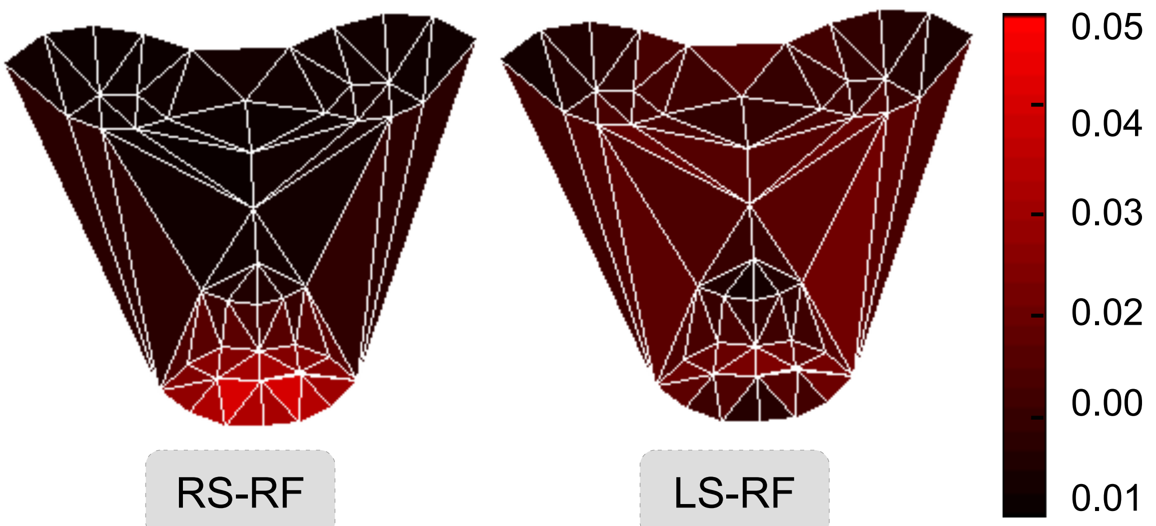
A Dapogny, K Bailly, S Dubuisson *Dynamic Pose-Robust Facial Expression Recognition by Multi-View Pairwise Conditional Random Forests*. **IEEE Trans. on Aff. Comp.** 2017

Presentation Outline

- Classical Pipeline
- Random forests for FER
- Pairwise conditional random forests
- **Local subspace random forests**
  - **Local expression predictions**
  - Weighted local subspace random forests
  - Confidence-Aware AU prediction using LEP features
- Conclusion and perspectives

Local expression predictions

- **Problem**: global RF (RS-RF) is **sensible to occlusions** of targeted areas
- No available data for realistic self-occlusions
- Idea: Grow trees on **spatially-constrained local face subspaces**
  - increased decorrelation
  - better repartition of the features and robustness to occlusions



A Dapogny, K Bailly, S Dubuisson. *Confidence-Weighted Local Expression Predictions for Occlusion Handling in Expression Recognition and Action Unit Detection*. **International Journal of Computer Vision,** 2017

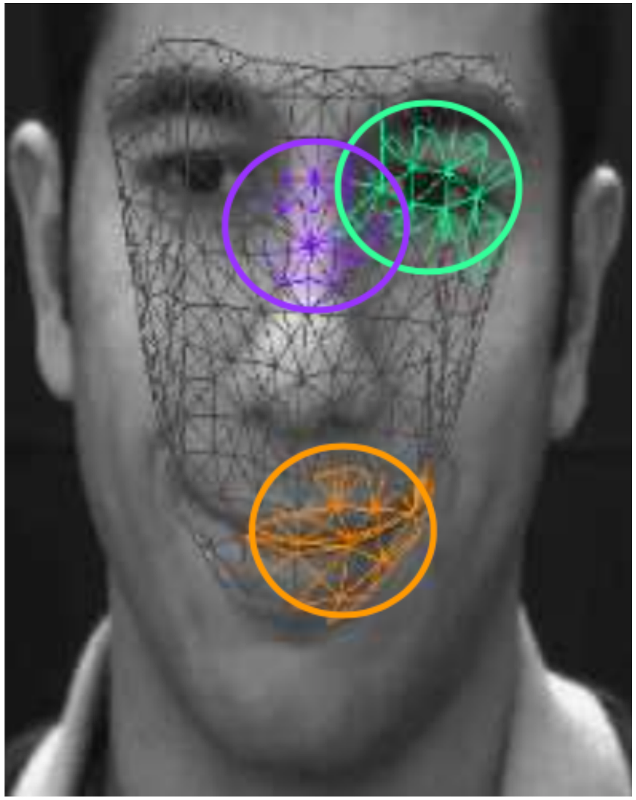
Local expression predictions

- Mesh tessellation via adaptative splitting
- Random masks M_t: Surface R, triangles τ
- Local Expression Predictions (LEP) :

$$p(c|\mathcal{I},\tau) = \frac{1}{Z_\tau}\sum_{t=1}^{T}\frac{\delta(\tau \in M_t)p_t(c|\mathcal{I})}{|M_t|}$$

- Local Subspace RF (LS-RF):

$$p(c|\mathcal{I}) = \frac{1}{T}\sum_{\tau} Z_\tau p(c|\mathcal{I},\tau)$$



A Dapogny, K Bailly, S Dubuisson. *Confidence-Weighted Local Expression Predictions for Occlusion Handling in Expression Recognition and Action Unit Detection*. **International Journal of Computer Vision,** 2017

Presentation Outline

- Classical Pipeline
- Random forests for FER
- Pairwise conditional random forests
- **Local subspace random forests**
  - Local expression predictions
  - **Weighted local subspace random forests**
  - Confidence-Aware AU prediction using LEP features
- Conclusion and perspectives

Weighted Local subspace random forests

- Weighted Local Subspace RF (WLS-RF):

$$p(c|\mathcal{I}) = \frac{\sum_{\tau} \alpha^{(\tau)} Z_{\tau} p(c|\mathcal{I}, \tau)}{\sum_{\tau} \alpha^{(\tau)} Z_{\tau}}$$

- $\alpha^{(\tau)}$ triangle-wise confidence measurement

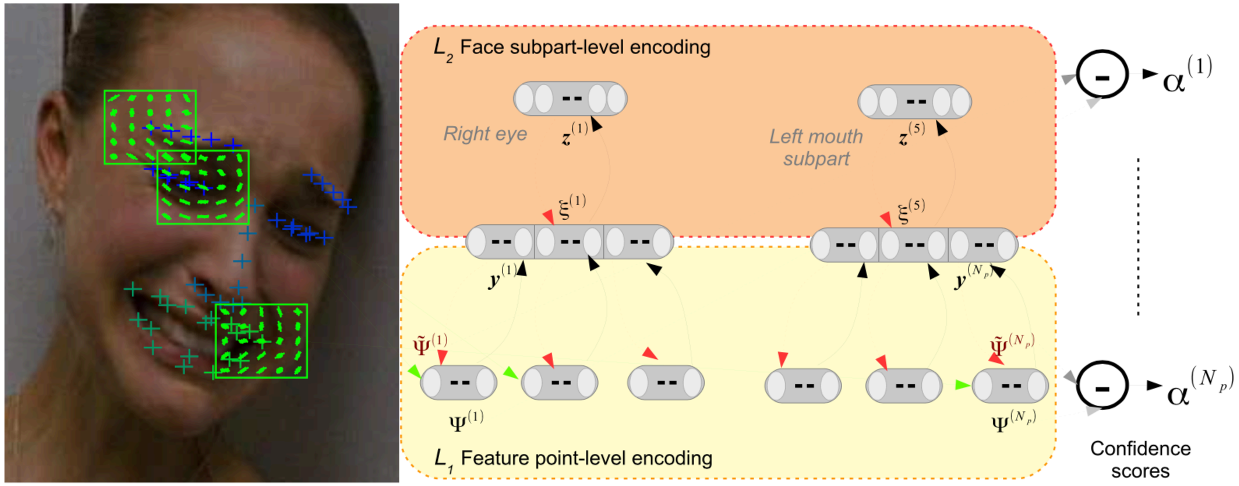
$$\alpha^{(\tau)}(\mathcal{I}) = \min(\alpha^{(k_1)}(\mathcal{I}), \alpha^{(k_2)}(\mathcal{I}), \alpha^{(k_3)}(\mathcal{I}))$$

- $\alpha^{(k)}$ feature point-wise confidence provided by a hierarchical autoencoder network

A Dapogny, K Bailly, S Dubuisson. *Confidence-Weighted Local Expression Predictions for Occlusion Handling in Expression Recognition and Action Unit Detection*. **International Journal of Computer Vision,** 2017

Weighted Local subspace random forests

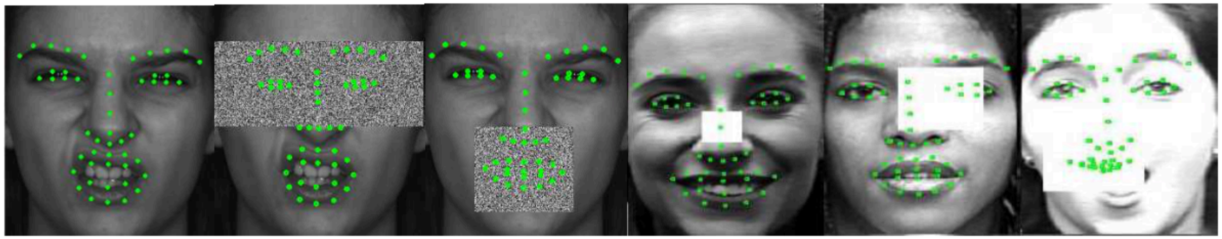
- Hierarchical autoencoder network
- Manifold learning of local texture with denoising autoencoders

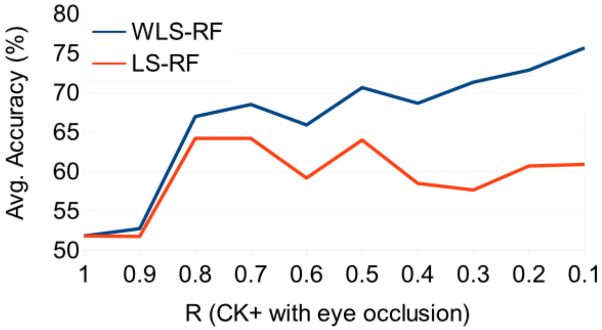


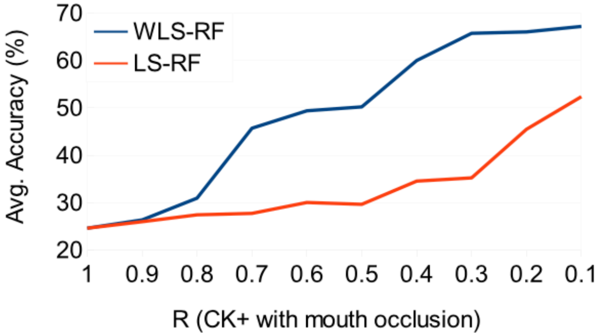
$$\alpha^{(k)}(\mathcal{I}) = 1 - \frac{||\boldsymbol{\Psi}^{(k)} - g^1 \circ g^2 \circ h^2 \circ h^1(\boldsymbol{\Psi}^{(k)})||^2}{(||\boldsymbol{\Psi}^{(k)}|| + ||g^1 \circ g^2 \circ h^2 \circ h^1(\boldsymbol{\Psi}^{(k)})||)^2}$$

A Dapogny, K Bailly, S Dubuisson. *Confidence-Weighted Local Expression Predictions for Occlusion Handling in Expression Recognition and Action Unit Detection*. **International Journal of Computer Vision,** 2017

Experiments on occluded data



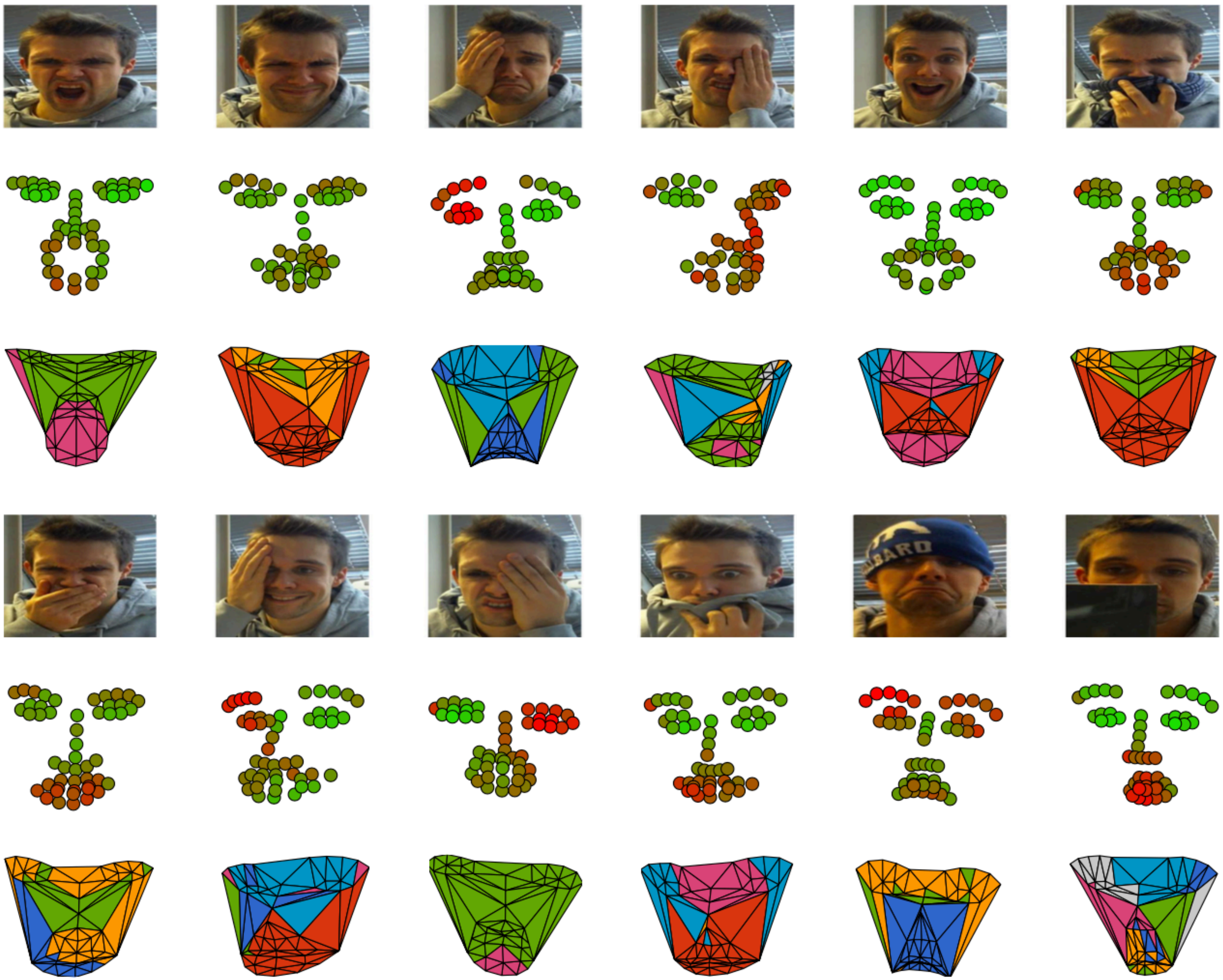




Occlusion	WLS-RF	RGBT[13]
None	**94.3**	**94.4**
R8	**92.2**	**92**
R16	**86.4**	82
R24	**74.8**	62.5
Eyes	**87.9**	**88**
Mouth	**72.7**	30.3

[13] "Random Gabor based templates for facial expression recognition in images with facial occlusion". In: *Neurocomputing* (2014).

Qualitative results

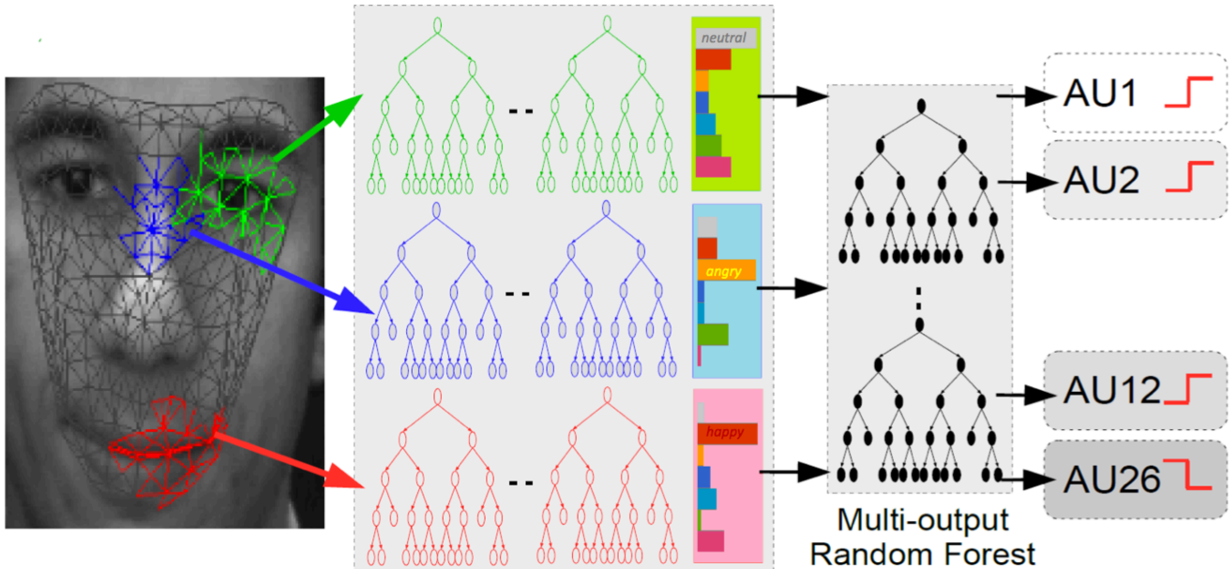


Presentation Outline

- Classical Pipeline
- Random forests for FER
- Pairwise conditional random forests
- **Local subspace random forests**
  - Local expression predictions
  - Weighted local subspace random forests
  - **Confidence-Aware AU prediction using LEP features**
- Conclusion and perspectives

Confidence-Aware AU prediction using LEP features

- Using LEPs as features to describe AU occurrence



Local expression predictions

- Use LEP feature relevance for AU-wise confidence measurement
- Single-output vs Multi-output
- Random task assignment at each split node

A Dapogny, K Bailly, S Dubuisson. *Confidence-Weighted Local Expression Predictions for Occlusion Handling in Expression Recognition and Action Unit Detection*. **International Journal of Computer Vision,** 2017

Confidence-Aware AU prediction using LEP features

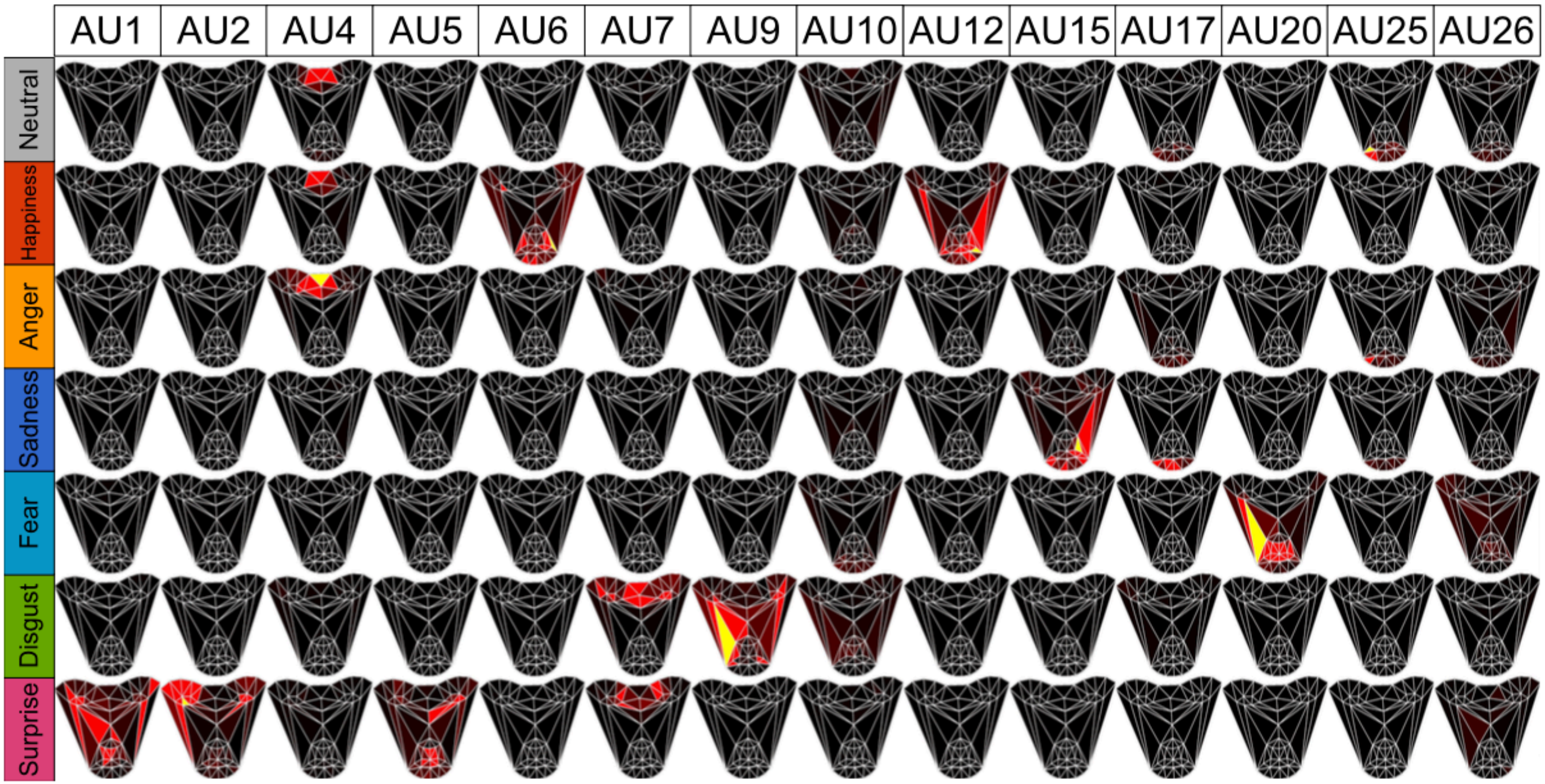
AU	MO-500	MO	SO	MLCNN[14]
1	**0.728**	0.674	0.684	0.702
2	0.619	0.526	0.552	**0.715**
4	**0.778**	0.752	0.667	0.704
6	**0.905**	0.897	0.860	0.855
9	**0.849**	0.825	0.770	0.829
12	**0.962**	**0.962**	0.929	0.913
25	**0.958**	0.957	0.950	0.838
26	0.793	0.770	**0.814**	0.757
Avg	**0.824**	0.795	0.778	0.789

- LEPs learned on multiple datasets yield high predictive power
- Multi-output formulations bring extra tree decorrelation
- Using a refined mesh allows better prediction of subtle AUs

[14] "A Multi-label Convolutional Neural Network Approach to Cross-Domain Action Unit Detection". In: *ACII*. 2015.

Confidence-Aware AU prediction using LEP features

- "reverse FACS": LEP heat maps



A Dapogny, K Bailly, S Dubuisson. *Confidence-Weighted Local Expression Predictions for Occlusion Handling in Expression Recognition and Action Unit Detection*. **International Journal of Computer Vision,** 2017

JEMImE Project

- **ANR JEMImE** (Jeu Educatif Multimodal d'Imitation Emotionnelle)
- Teach ASD children how to produce adequate emotions
- Assess **the quality of emotions** produced by children
- Requires analysis of subtle **spontaneous facial behaviors**
- **Robust and Real-time**

JEMImE Project

Conclusion

- Solutions for automatic facial expression analysis
  - Random-forest based approach
  - Prototypical emotions and AU prediction
  - Temporal integration, pose and occlusion handling

- Future work
  - New conditional networks architectures (e.g. deep neural decision forests) for robust and real time face processing systems
  - Preliminary results for facial landmark tracking