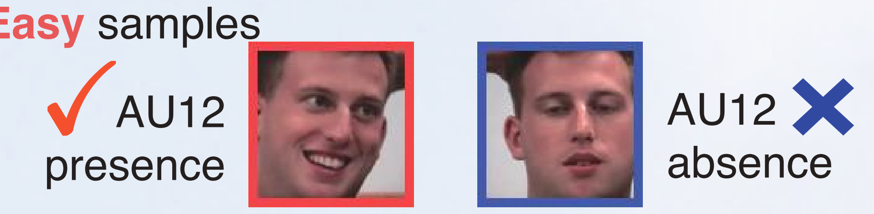
Confidence Preserving Machine for Facial Action Unit Detection

Jiabei Zeng[1], Wen-Sheng Chu[2], Fernando De la Torre[2], Jeffrey F. Cohn[2,3], and Zhang Xiong[1]

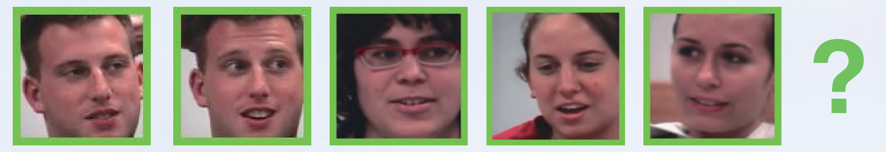
ICCV15 SANTIAGO, CHILE DECEMBER 11-18, 2015

Problem

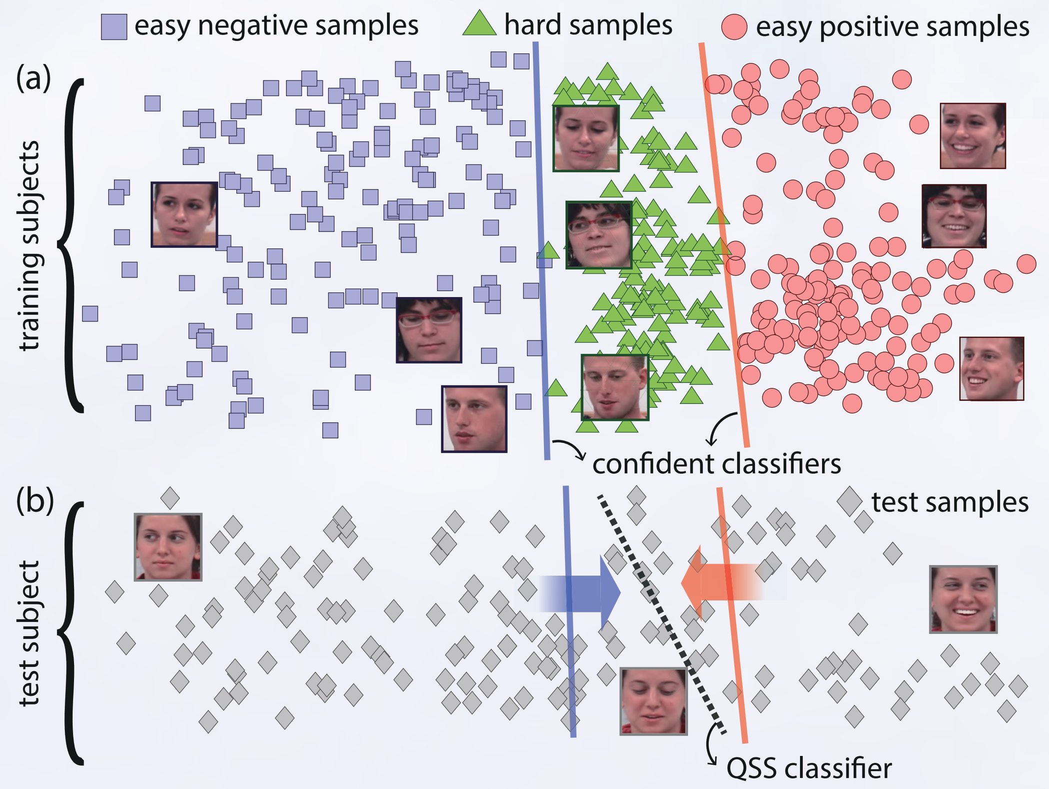
- Facial Action Unit (AU) detection
  - △ Easy samples



  - △ Hard samples



- Standard supervised methods
  - △ Single hyperplane
  - △ Perform well on easy samples, e.g., high intensity AUs, frontal head pose or on particular subjects
  - △ Perform badly on hard samples, e.g. , subtle AUs
- Main idea
  - △ Easy-to-hard strategy

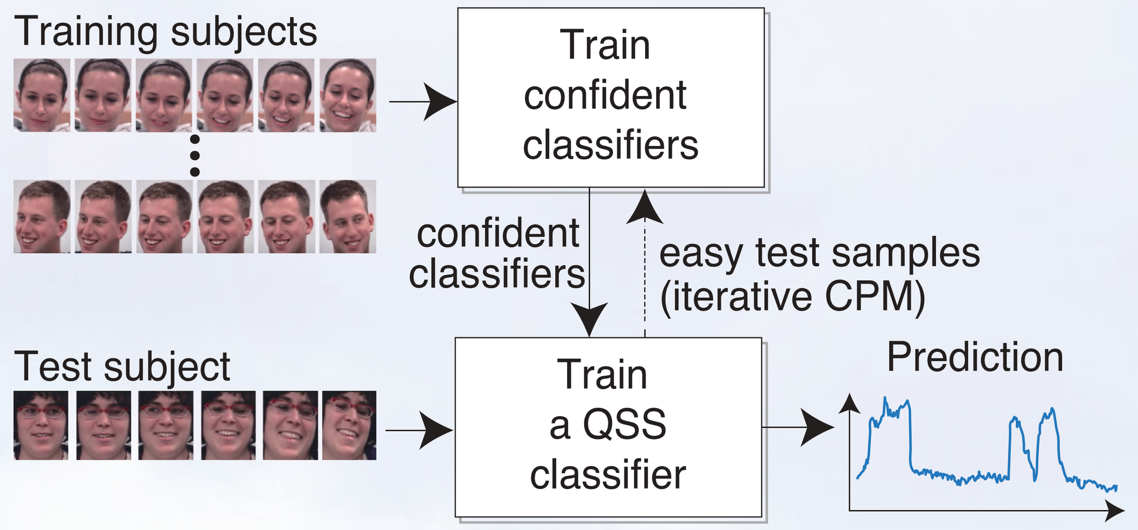


- Previous work
  - △ Address specific and known sources
    - E.g. subtle AUs, head pose, individual differences [1]
  - △ Semi-supervised learning methods (SSL) [2]
    - Smoothness/Cluster/Manifold assumptions on relationships between input and label space
    - Not suitable for AU
  - △ Transfer learning methods [3,4,5]
    - Only individual differences
    - Absence of spatial-temporal smoothness
    - Less efficient than CPM

[1] W.-S. Chu, et al. Selective transfer machine for personalized facial action unit detection. CVPR 2013.
[2] S. Melacci, et al. Laplacian support vector machines trained in the primal, JMLR, 12:1149–1184, 2011.
[3] Duan L, et al. Domain adaptation from multiple sources: A domain-dependent regularization approach, TNNLS, 23(3): 504-518, 2012.
[4] B. Gong, et al. Geodesic flow kernel for unsupervised domain adaptation, CVPR 2012.
[5] Q. Sun, et al. A two-stage weighting framework for multi-source domain adaptation, NIPS 2011.

Confidence Preserving Machine (CPM)

- Diagram of CPM



Confident classifiers

- ($\mathbf{w}_+$, $\mathbf{w}_-$), or $\mathbf{w}_y$ to confidently predict on positive and negative samples, respectively
- Identify easy and hard samples (1)
  - △ $\mathcal{E}$ easy: ($\mathbf{w}_+$, $\mathbf{w}_-$) have same predictions
  - △ $\mathcal{H}$ hard: ($\mathbf{w}_+$, $\mathbf{w}_-$) have different predictions
- Objective (2)

truth labels

$$\min_{\mathbf{w}_y,\mathcal{E}} \sum_{y\in\{+,-\}} \|\mathbf{w}_y\|^2 + \sum_i \xi_{i_y}^2 \quad \text{s.t. } \boxed{y_i}\mathbf{w}_y^\top \mathbf{x}_i \geq 1-\xi_{i_y}, \forall i \in \mathcal{E},$$
$$\boxed{\eta_i^y}\mathbf{w}_y^\top \mathbf{x}_i \geq 1-\xi_{i_y}, \forall i \in \mathcal{H}$$

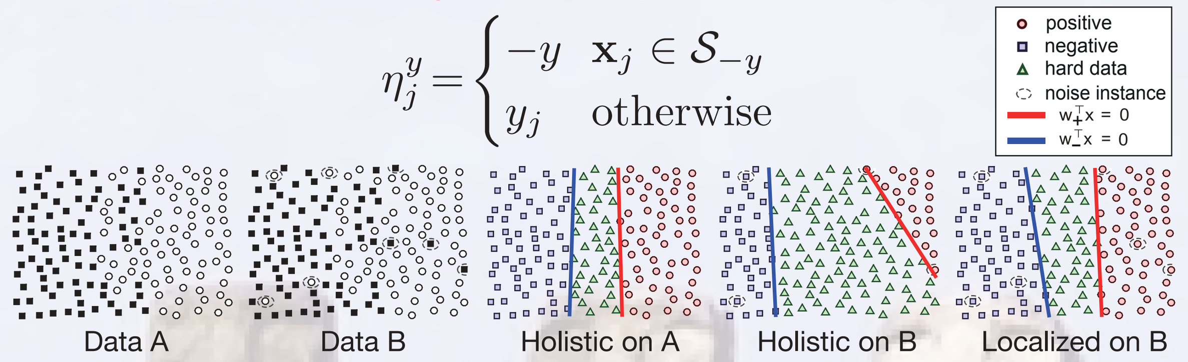
relabel

- Relabeling strategies
  - △ Holistic relabeling: Relabel all the hard samples

$$\eta_j^y = -y, \forall \mathbf{x}_j \in \mathcal{H}.$$

  - △ Localized relabeling: Relabel part of the hard samples

$$\eta_j^y = \begin{cases} -y & \mathbf{x}_j \in \mathcal{S}_{-y} \\ y_j & \text{otherwise} \end{cases}$$



- Alternating algorithm for training confident classifiers

Algorithm 1 Train confident classifiers

Input: Data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ and its index set $\mathcal{D} = \{1, 2, \ldots, n\}$

Output: Confident classifiers $(\mathbf{w}_+, \mathbf{w}_-)$, easy samples $\mathcal{E}$ and hard samples $\mathcal{H}$

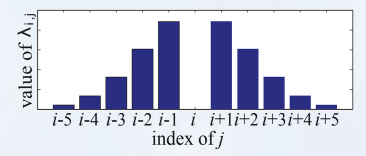
Initialization: $\mathcal{E} \leftarrow \mathcal{D}$; $\mathcal{H} \leftarrow \emptyset$;
repeat
  $(\mathbf{w}_+, \mathbf{w}_-) \leftarrow$ solve (2) with fixed $\mathcal{E}$ and $\mathcal{H}$;
  Update easy and hard samples $(\mathcal{E}, \mathcal{H})$ using (1);
  Update relabels $\eta_j^+, \eta_j^- \forall j \in \mathcal{H}$;
until convergence or reach maximum #iteration

Quasi-semi-supervised (QSS) learning

- Generalization on unseen test subjects
  - △ A person-specific classifier $\mathbf{w}_t$ for a test subject
- Quasi-semi-supervised (QSS) classifier
  - △ To get the person-specific classifier $\mathbf{w}_t$
  - △ Virtual labels for easy samples from ($\mathbf{w}_+$, $\mathbf{w}_-$)
  - △ Propagate virtual labels from easy to hard test samples under semi-supervided manner.
- Objective (4)

$$\min_{\mathbf{w}_t} \underbrace{\sum_{i\in\mathcal{E}_t} \ell(\hat{y}_i, \mathbf{w}_t^\top \mathbf{x}_i)}_{\text{loss}} + \gamma_s\|\mathbf{w}_t\|^2 + \gamma_I \underbrace{(\mathbf{X}^{te}\mathbf{w}_t)^\top \mathbf{D}^\top \mathbf{D}\mathbf{X}^{te}\mathbf{w}_t}_{\text{spatial-temporal smoothness}}$$

$$\sum_{i\in\mathcal{D}^{te}} \left(\mathbf{w}_t^\top \mathbf{x}_i - \frac{1}{Z_i} \sum_{j=i-T, j\neq i}^{i+T} \lambda_{ij} e_{ij} \mathbf{w}_t^\top \mathbf{x}_j \right)^2$$

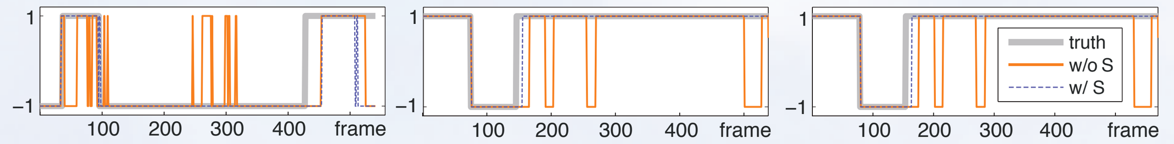


preserve temporal similarity

exclude spatial disimilarity

$$e_{ij} = \begin{cases} 1, d_{ij} < \epsilon \\ 0, \text{otherwise} \end{cases}$$

- Effectiveness of spatial-temporal smoothness



Iterative CPM (iCPM)

- ($\mathbf{w}_+$, $\mathbf{w}_-$) might not generalize well due to mismatch between training and test.
- iCPM jointly learns ($\mathbf{w}_+$, $\mathbf{w}_-$) and QSS.

Algorithm 2 Iterative Confidence Preserving Machine

Input: labeled training data $\{\mathbf{x}_i, y_i\}_{i=1}^n$ with index set $\mathcal{D} = \{1, 2, \ldots, n\}$, unlabeled test data $\{\mathbf{x}_i^{te}\}_{i=1}^m$ with index set $\mathcal{D}^{te} = \{1, 2, \ldots, m\}$

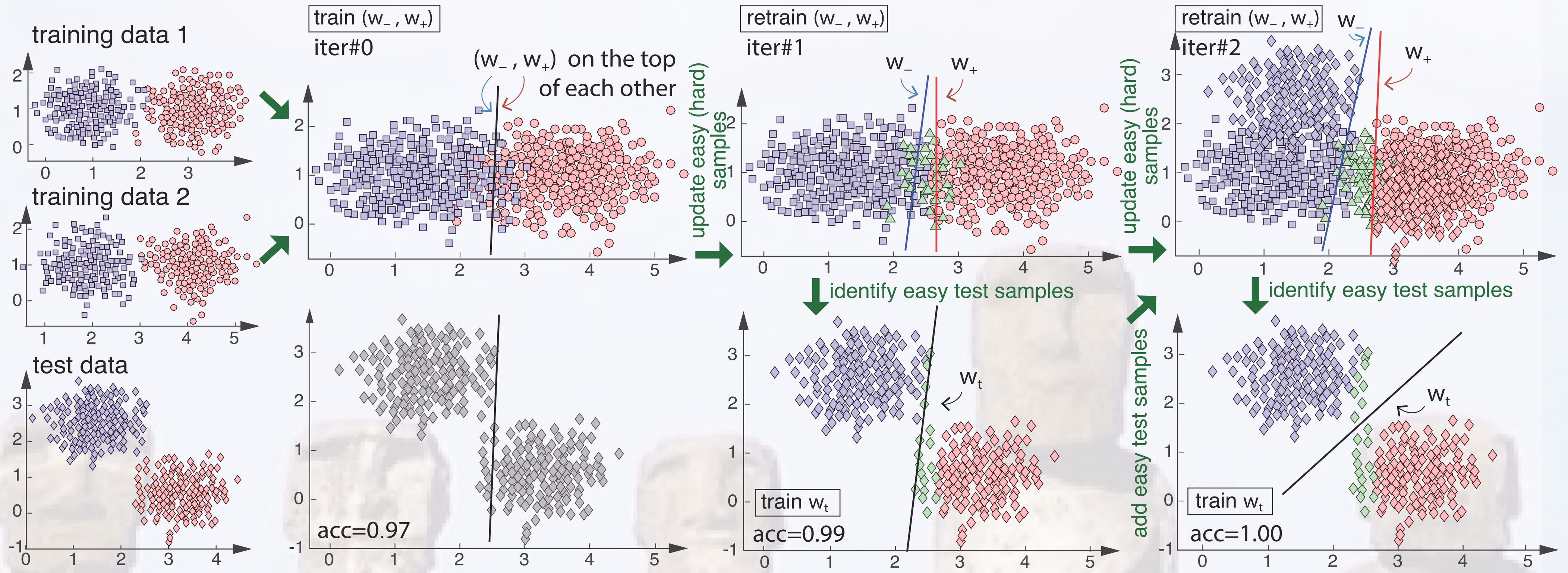
Output: QSS classifier $\mathbf{w}_t$

$\mathcal{E} \leftarrow \mathcal{D}, \mathcal{H} \leftarrow \emptyset$;
$(\mathbf{w}_+, \mathbf{w}_-) \leftarrow$ solve (2);
$(\mathcal{E}, \mathcal{H})$ using (1);
repeat
  Update relabels $\eta_j^+, \eta_j^-, \forall j \in \mathcal{H}$;
  $(\mathbf{w}_+, \mathbf{w}_-) \leftarrow$ solve (2) with fixed $\mathcal{E}$ and $\mathcal{H}$;
  Estimate virtual labels $\{\hat{y}_i\}_{i=1}^m$,

$$\hat{y}_i = \begin{cases} 1 & \mathbf{w}_y^\top \mathbf{x}_i^{te} > 0, \forall y \in \{-1, +1\}, \\ -1 & \mathbf{w}_y^\top \mathbf{x}_i^{te} < 0, \forall y \in \{-1, +1\}, \\ 0 & \text{otherwise.} \end{cases}$$

  $\mathcal{E}_t = \{i \in \mathcal{D}^{te} | \text{sign}(\mathbf{w}_+^\top \mathbf{x}_i^{te}) = \text{sign}(\mathbf{w}_-^\top \mathbf{x}_i^{te})\}$;
  if $\exists i, j \in \mathcal{E}_t$, s.t. $\hat{y}_i = -1, \hat{y}_j = 1$ **then**
    $\mathbf{w}_t \leftarrow$ solve (4) given $\mathbf{X}^{te}$ and $\{\hat{y}_i\}_{i=1}^m$;
  else
    $\mathbf{w}_t = \frac{1}{2}(\mathbf{w}_+ + \mathbf{w}_-)$;
  end if
  Update $\mathcal{E}_t = \{i \in \mathcal{D}^{te} | \hat{y}_i = \text{sign}(\mathbf{w}_t^\top \mathbf{x}_i^{te})\}$;
  Update $(\mathcal{E}, \mathcal{H}) \leftarrow (1)$;
  $\mathcal{E} \leftarrow \mathcal{E} \cup \mathcal{E}_t$;
until convergence

- iCPM on a **synthetic example**



Comparison with alternative methods

	Multiple classifiers	Indentify easy/hard	Unlabeled data	Different distribution	Smoothness assumption	Non-parallel hyperplane	Progressive labeling
Boosting	✓	×	×	×	×	×	×
TW-SVM [6]	✓	×	×	×	×	✓	×
Self-paced learning [7]	×	✓	×	×	×	×	×
RO-SVM [8]	×	✓	×	×	×	×	×
Co-training [9]	✓	×	✓	×	×	×	✓
Lap-SVM [2]	×	×	✓	×	✓	×	×
DAM [3]	✓	×	✓	✓	✓	×	✓
CPMs	✓	✓	✓	✓	✓	✓	✓

[6] R. Khemchandani, et al. Twin support vector machines for pattern classification, TPAMI, 29:905-910, 2007.
[7] M. P. Kumar, et al. Self-paced learning for latent variable models, NIPS 2010.
[8] Y. Grandvalet, et al. Support vector machines with a reject option, NIPS 2009.
[9] A. Blum, et al.Combining labeled and unlabeled data with co-training, CoLT 1998.

Experiments

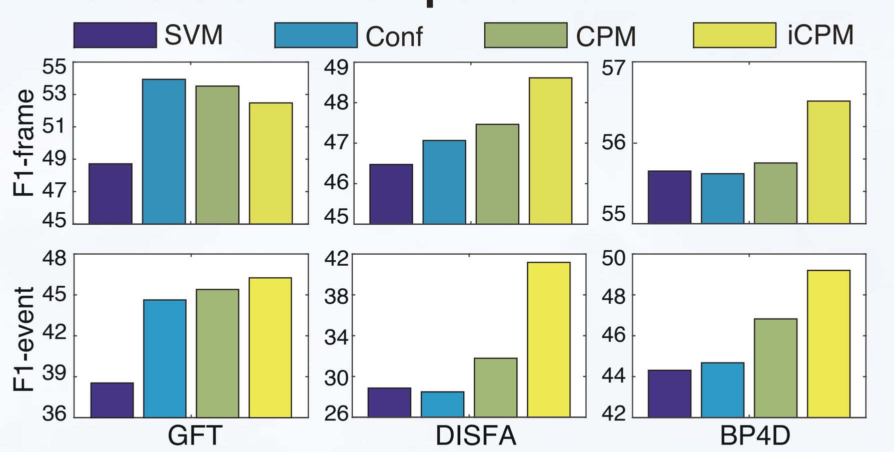
- **Settings**
  - △ SIFT descriptors on predetermined facial landmarks
  - △ 10-fold with disjoint training and test sets
- **Datasets**
  - △ GFT [10]: 50 2-min spontaneous videos from 50 participants
  - △ BP4D [11]: 328 spontaneous videos from 41 participants
  - △ DISFA [12]: 27 spontaneous videos from 27 participants
- **Metric**
  - △ Frame-based F1

$$\text{F1-frame} = \frac{2P \cdot R}{P+R}$$

  - △ Event-based F1

$$\text{F1-event} = \frac{2EP \cdot ER}{EP+ER}$$

- **Evaluate CPM components**



- **Comparisons**
  - △ baseline methods (SVM, Adaboost), SSL (Laplacian SVM [2]), transfer learning methods (DAM [3], MDA [4], GFK [5])

Table 1. Comparison on GFT. ("H" stands for an extra post-processing with HMM)

AU	F1-frame SVM|H	Ada|H	Lap|H	DAM	MDA	GFK	iCPM	F1-event SVM|H	Ada|H	Lap|H	DAM	MDA	GFK	iCPM
1	30.3|16.8	20.3|15.4	12.1|16.4	1.7	29.2	**30.9**	29.9	20.3|17.9	15.3|**28.2**	5.4| 9.7	2.1	21.3	21.6	27.1
2	25.6|18.4	14.8|21.8	26.0|19.3	5.3	25.8	**29.3**	25.7	20.2|21.1	12.2|**30.7**	18.2|16.6	4.7	21.3	22.5	24.8
6	66.2|66.4	62.1|47.3	2.7|40.7	58.0	63.8	66.1	**67.3**	49.1|**56.8**	47.5|43.4	4.4|37.5	50.0	47.0	50.2	**56.8**
7	70.9|72.2	69.6|50.0	24.0|50.3	66.0	66.6	72.2	**72.5**	50.4|59.8	50.7|44.0	21.6|48.3	41.7	49.2	52.1	**60.1**
10	65.5|65.5	65.5|43.7	56.7|61.2	64.9	65.4	**67.5**	67.0	50.2|57.8	50.2|46.6	46.5|57.5	53.1	51.6	54.3	**58.1**
12	74.2|**75.9**	73.0|54.5	64.8|69.0	72.9	71.9	72.7	75.1	56.3|**65.0**	54.7|59.9	54.9|64.4	61.9	52.0	54.3	**65.0**
14	79.6|78.1	77.7|59.2	76.7|51.2	79.5	74.0	79.8	**80.7**	63.8|70.8	62.3|59.9	**81.5**|61.2	64.6	63.7	64.8	74.7
15	34.1|17.5	20.3|20.5	19.3|13.9	1.4	31.8	31.7	**43.5**	28.1|20.1	17.7|41.8	15.9|20.2	2.3	25.4	26.8	**32.2**
17	49.2|**50.6**	48.2|38.6	42.5|21.2	34.6	47.4	48.9	49.1	42.9|**53.1**	37.1|38.5	36.4|25.9	29.6	41.4	41.3	52.3
23	28.3|29.8	19.4|20.7	27.1|25.1	2.8	26.0	26.7	**35.0**	27.7|35.9	16.8|**36.7**	9.5|19.5	4.4	26.7	27.1	25.9
24	31.9|21.0	22.3|25.8	25.7|16.9	3.0	31.8	**33.0**	31.6	30.3|21.8	20.8|26.4	21.7|13.9	4.9	30.0	30.5	**31.8**
Av.	48.7|46.6	44.8|36.1	32.8|35.0	35.5	48.5	48.6	**52.5**	38.6|43.7	35.0|41.5	27.3|34.1	29.0	39.1	38.9	**46.3**

Table 2. Comparison on BP4D. ("H" stands for an extra post-processing with HMM)

AU	F1-frame SVM|H	Ada|H	Lap|H	DAM	MDA	GFK	iCPM	F1-event SVM|H	Ada|H	Lap|H	DAM	MDA	GFK	iCPM
1	46.0|43.4	41.5|37.7	43.8|29.0	38.2	39.6	42.4	**46.6**	29.2|38.1	29.8|**41.7**	29.2|27.8	26.7	30.5	29.7	35.3
2	38.5|38.4	12.4|25.5	17.6|27.8	27.3	37.0	35.8	**38.7**	29.3|**36.1**	12.9|32.4	24.8|27.1	12.3	28.2	28.9	32.5
4	**48.5**|41.6	39.4|30.4	27.2|26.1	29.1	45.7	47.3	46.5	33.5|37.4	28.9|28.3	30.5|26.5	22.3	32.8	32.8	**39.4**
6	67.0|62.0	**71.7**|61.2	71.5|26.1	67.5	69.2	71.2	68.4	53.7|37.4	54.4|58.5	53.7|26.5	55.4	52.9	54.4	**60.9**
7	72.2|56.5	**74.7**|53.7	71.6|52.2	72.6	70.2	72.5	73.8	59.0|55.3	55.2|49.2	56.2|57.6	61.1	58.4	54.9	**62.1**
10	72.7|54.6	**75.7**|62.1	72.8|55.3	74.4	71.0	74.2	74.1	61.3|52.8	59.3|67.8	60.7|60.6	**68.6**	57.5	59.7	65.1
12	83.6|65.4	84.3|62.6	84.3|55.3	76.4	81.8	83.9	**84.6**	62.5|52.8	63.9|60.8	64.2|60.6	60.8	59.9	65.6	**71.4**
14	59.9|49.2	61.0|50.9	**62.6**|26.3	59.9	57.8	57.2	62.2	49.5|46.3	51.7|**56.7**	51.9|26.9	53.3	50.2	48.7	55.9
15	41.1|39.9	30.6|30.4	35.2|25.5	15.9	41.4	40.6	**44.3**	33.7|**39.0**	24.4|**39.0**	25.4|25.4	12.7	28.2	31.1	37.4
17	55.6|57.8	56.6|47.8	**59.1**|46.3	52.9	50.1	55.4	57.5	46.0|**56.1**	44.0|51.5	44.0|41.7	51.5	39.6	44.0	49.9
23	40.8|39.4	33.0|32.8	33.6|27.6	3.9	36.2	39.9	**41.7**	36.4|**44.0**	28.2|41.4	27.2|22.2	5.8	30.7	33.3	41.9
24	**42.1**|19.3	34.2|26.7	40.5|16.9	4.9	41.1	41.7	39.7	37.7|16.0	30.9|35.7	34.8|13.8	3.6	35.4	35.6	**38.7**
Av.	55.7|47.3	51.3|43.5	54.7|36.9	42.6	53.4	55.2	**56.5**	44.3|44.8	40.3|46.9	41.9|36.5	36.2	42.0	43.2	**49.2**

Table 3. Comparison on DISFA. ("H" stands for an extra post-processing with HMM)

AU	F1-frame SVM|H	Ada|H	Lap|H	DAM	MDA	GFK	iCPM	F1-event SVM|H	Ada|H	Lap|H	DAM	MDA	GFK	iCPM
1	26.5|14.4	17.1|12.4	13.1|16.2	7.9	19.0	23.2	**29.5**	14.5|17.6	16.1|**21.2**	9.6|11.6	5.4	11.6	18.1	18.7
2	24.0|15.3	20.1|10.5	6.4|12.6	13.1	9.5	16.3	**24.8**	10.8|17.6	17.3|17.0	11.4|11.0	12.1	16.4	17.4	**19.2**
4	56.1|48.5	59.8|26.4	21.1|23.4	40.4	59.3	**60.3**	56.8	31.6|37.2	32.5|27.7	15.9|16.2	32.4	28.6	28.3	**41.8**
6	40.9|34.9	31.9|22.1	22.1|19.9	19.2	21.1	**41.9**	41.7	30.3|29.2	28.3|25.9	23.7|13.5	22.6	30.8	30.6	**36.9**
9	30.5|10.9	29.3|17.4	12.1|10.9	11.9	7.6	30.3	**31.5**	23.4|13.2	22.7|**39.9**	7.8| 8.3	14.3	27.4	14.6	31.7
12	65.6|70.1	69.4|46.3	33.7|32.2	50.9	63.1	69.6	**71.9**	49.9|57.1	51.9|61.6	33.2|20.7	44.3	42.1	46.2	**56.6**
25	78.3|**84.1**	83.9|70.5	35.3|30.3	56.2	81.3	80.0	81.6	31.9|**76.7**	38.0|58.8	42.5|21.2	56.4	46.5	36.1	**76.7**
26	50.0|51.5	**59.6**|50.5	18.9|25.5	43.2	51.1	54.6	51.3	38.6|**51.7**	38.7|49.5	48.7|18.4	38.4	36.4	37.3	47.7
Av.	46.5|41.2	46.6|32.0	20.3|21.4	30.4	39.0	47.0	**48.6**	28.9|37.5	30.7|37.7	24.1|15.1	28.2	30.0	28.6	**41.2**

[10] J. F. Cohn, et al. Spontaneous facial expression in a small group can be automatically measured: An initial demonstration, Behavior Research Methods, 42(4):1079–1086, 2010.
[11] X. Zhang, et al. A high-resolution spontaneous 3d dynamic facial expression database, FG 2013.
[12] S. Mavadati, et al. Disfa:A spontaneous facial action intensity database,TAC,4:151–160, 2013.

Supported in part by NSFC 61272350, 2013AA01A603