‘Exemplar Hidden Markov Models for Classification of Facial Expressions in Videos’

Workshop on Analysis and Modeling of Face and Gesture
CVPR 2015

Karan Sikka

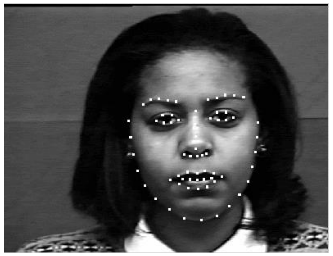
Machine Perception Lab

UCSD

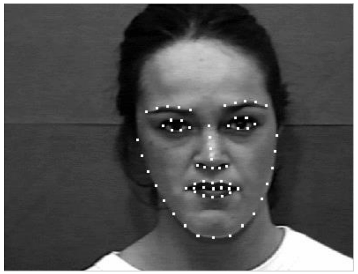
Joint work with Dr. Abhinav Dhall and Dr. Marian Bartlett

Automatic Facial Expression Recognition

- Classify underlying expressions in a video.
- Emotions, Pain, engagement level.



Smile
Disgust
Surprise



Smile
Disgust
Surprise

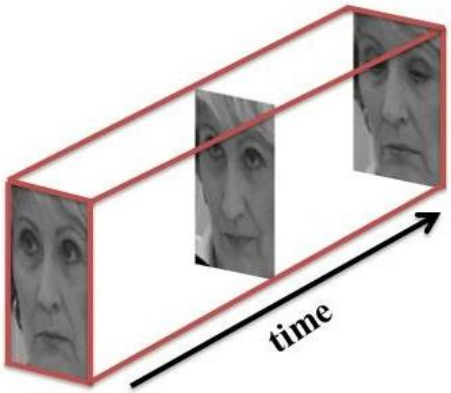
Previous Art
Image Based Approaches

- Spatial features + Classifier.
- Issues
  - Require key-shots (apex frames).
  - No explicit dynamics.
- Gabor, LBP, SIFT.

- Image - > Video based approaches.

Previous Art
Video-based approaches

- Space-Time
  - Extract localized S-T features across entire video.
  - Feature pooling + Classifier.
  - LBPTOP, BoW, facial point time-series.



- Issues
  1. Pooling from multiple expressions.
     - Loss of discriminative power (unsegmented videos).
  2. Loss of temporal information.
     - No temporal correspondence between facial states.

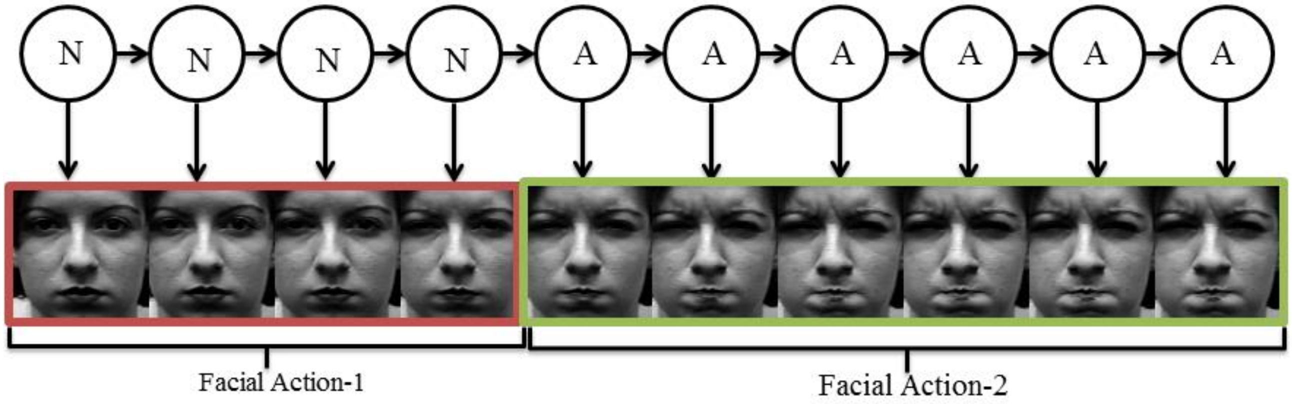
Previous Art for AFER
Video based approaches

- Sequential
  - Analyze an expression as a sequence of features.
  - Explicitly model spatio-temporal aspects.

- Focus on HMMs.
  - Desirable properties for modeling expressions.

Why HMM

- HMMs model expression dynamics.



- Hidden states: Temporal Segmentation (variable length)
- Model per state: Model behavior for each facial state (local states).
- Transition probabilities: Temporal dynamics

HMMs -> Exemplar HMMs

- In PRACTICE *Accuracy*(HMM) < *Accuracy* (Discriminative)
  - Generative model.
  - Modeling decision boundary is easier than modeling classes.

- Solution Proposed
  - Structural advantages of HMM + discriminative ability SVMs.
  - Probabilistic kernels.

- Probabilistic kernels.
  - Recognition of dynamics textures, handwritten text, shapes.
  - Jaakkola et al., Jebara et al., Vasconcelos et al.

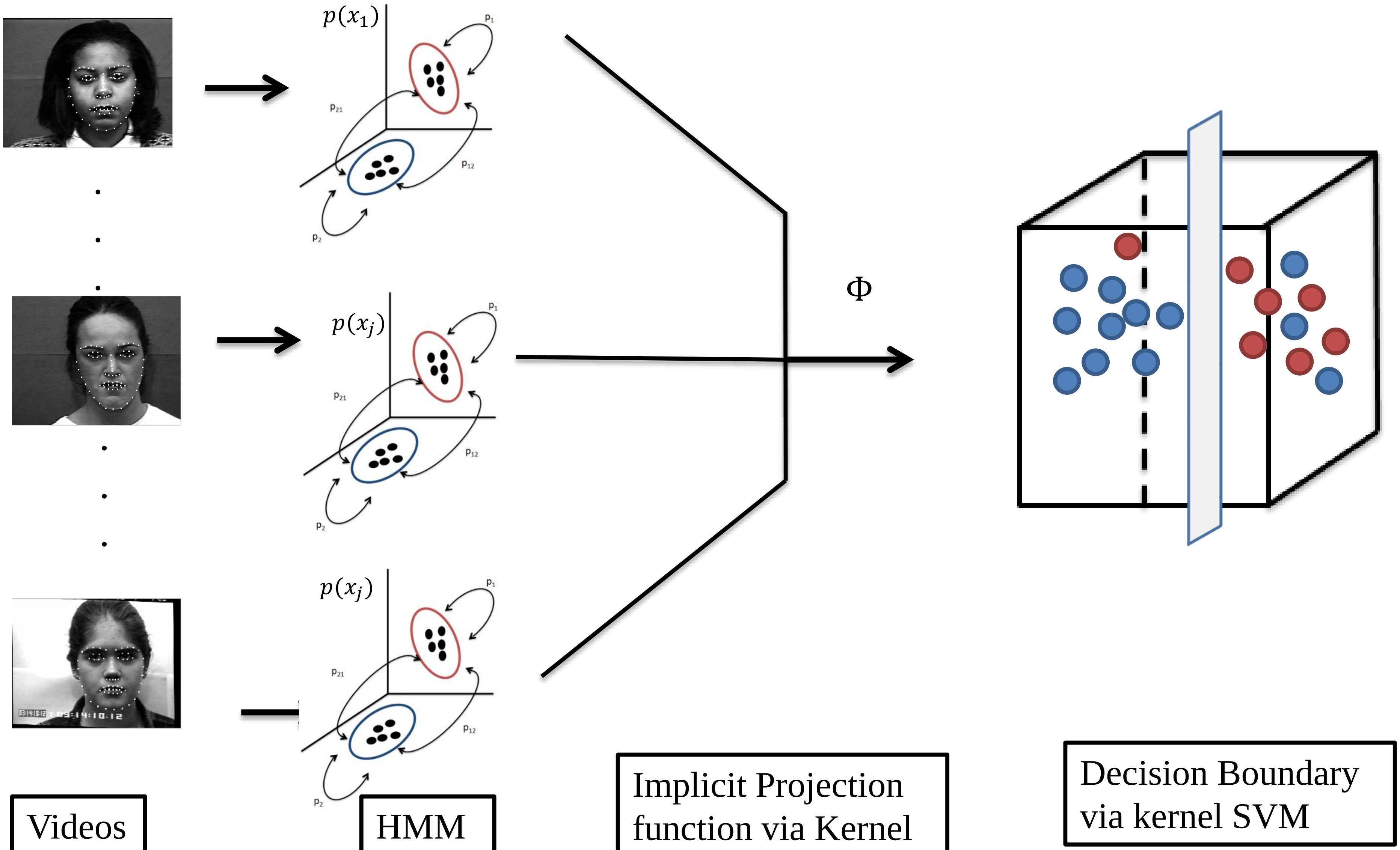
Kernels and implicit space

- Dot products in SVM can be replaced by Kernel functions (Kernel SVM)

$$\mathrm{K}(x_i, x_j) = < \Phi(x_i), \Phi(x_j) >$$

- Possible to compute Dot products indirectly for points in non-Eucledian (implicit) space
  - Φ maps HMM models to a vector space.
  - $\mathrm{K}(p_i, p_j) = < \Phi(p_i), \Phi(p_j) >$

Exemplar-HMMs



Probabilistic Product Kernel (PPK)

- PPK (Jebara et al.) to compute distance bw two HMMs

$$k(p_1, p_2) = \int p_1(x)\, p_2(x) \mathrm{dx}$$

- Closed form solution of HMM (Exponential family).
- Intuitive: Compares all states from two HMMs while using transition probabilities.

Experiments
Basic Emotions

Dataset	Type	Videos/subjects	Target
CK+	Posed	327 (118 subj)	7 emotions (leave-one-subject)
Oulu-CASIA VIS	Posed	480 (80 sub)	6 emotions (10 fold)
FEEDTUM	Spontaneous	320 (19 subj)	6 emotions (leave-one-subject)

- Time-series of facial landmarks points ($49*2$ dim).
- PCA.
- Metric: Average accuracy across all classes.

Competing algorithms

1. Global S-T
   - Landmark features + pooling +SVM
   - LBPTOP: Local Binary Patterns (texture) from XYT planes.
     - S-T Histograms + Pooling

2. Baseline generative model
   - HMM generative classifier

3. State of the art- Expressionlets (STM-Explet, Liu et al, CVPR’14)
   - Explicit temporal info inside texture features.
   - Universal GMM (UGMM) learned.
   - Video-> Align localized S-T features with UGMM.

Experiments
Basic Emotions- Posed

Method	Accuracy **(CK +)**	Accuracy **(Oulu)**
Geom. + Mean-pooling	93.00 (±1.55)	70.83 (±2.84)
Geom. + Max-pooling	92.85 (±1.67)	69.16 (±1.80)
LBPTOP	91.30 (±1.79)	72.08 (±2.22)
HMM	85.35 (±2.16)	63.54 (±3.10)
STM-ExpLet	94.19 (N/A)	74.59 (N/A)
ITBN	86.3 (±N/A)	NA
Exemplar-HMMs	**94.60** (±1.55)	**75.00** (±2.12)

- Significant improvement compared to S-T approaches.

Experiments
Basic Emotions- Posed

Method	Accuracy **(CK +)**	Accuracy **(Oulu)**
Geom. + Mean-pooling	93.00 (±1.55)	70.83 (±2.84)
Geom. + Max-pooling	92.85 (±1.67)	69.16 (±1.80)
LBPTOP	91.30 (±1.79)	72.08 (±2.22)
HMM	85.35 (±2.16)	63.54 (±3.10)
STM-ExpLet	94.19 (N/A)	74.59 (N/A)
ITBN	86.3 (±N/A)	NA
Exemplar-HMMs	**94.60** (±1.55)	**75.00** (±2.12)

- Advantages of discriminative modeling over generative modeling. .

Experiments
Basic Emotions- Spontaneous

Method	Accuracy **(FEEDTUM)**
Geom. + Mean-pooling	48.91 (±3.70)
Geom. + max-pooling	53.87 (±2.59)
LBPTOP	48.17 (±3.31)
HMM	48.23 (±3.88)
Exemplar-HMMs	**54.14** (±3.72)

AMFED Dataset

- Videos of participants watching 3 superbowl commercials.

- Video responses collected over the internet along with self-ratings describing:
  - Like/not-like
  - Watch again or not

- Public Dataset
  - 242 videos
  - Expert annotations for : AU 2, 4, 5 9, 12, 14, 15, 17, 18, 26 + Smile + Expressability.
  - Annotations in form of agreement between annotators.

AMFED

- 2 Binary self-report prediction tasks
  - Predict whether a video is rated liked/not-liked.
  - Predict whether a video will be watched again or not.
- Using time-series of AU annotations.
  - Threshold to 0 (<50%) and 1 (>=50%) based on agreement.
- 3 Fold
  - AUC

Results
AMFED

Method	**Like/Don't Like**	**Watch-again/Don't Watch-again**
AU + Mean-pooling	.66	.87
AU + max-pooling	.61	.89
HMM	.58	.84
Exemplar-HMMs	**.84**	**.92**

Devil in the details

- Bayesian HMMs avoid overfitting and lead to better results.
- Cross-validation necessary to select the kernel parameters.
- Gaussian assumption limits dimensionality.
  - To be extended for texture (high dim.) features.

Summary

- Explored approach for using HMMs within a discriminative framework for AFER.

- Exemplar-HMMs for temporal modeling
  - Temporal segmentation + model expression states
  - Model dynamics
  - Maintains specificity of each example

- PPK for model-based similarity
  - Comprehensively compares states from two HMM.
  - Takes into account temporal information.

Questions?



Karan Sikka

Abhinav Dhall

Dr. Marian S. Bartlett

Machine Perception Lab, UCSD

Thanks