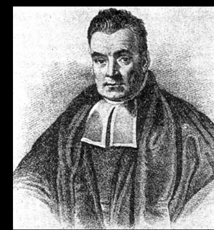
Department of Electrical & Computer Engineering

My Adventures with Bayes

Searching for Bayes optimal solutions in machine learning, statistics, computer vision, neuroscience and beyond

Aleix M. Martinez
Computational Biology
and Cognitive Science Lab
aleix@ece.osu.edu

Bayes versus Least Squares



VS.

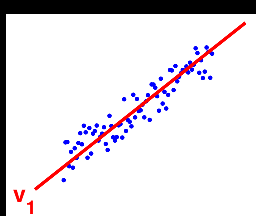
T. Bayes
1702-1761

J.C.F. Gauss
1777-1855

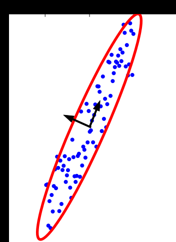
Martinez & Kak, "PCA versus LDA,"
IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001.

PCA: Least squares

- PCA is generally used to find convenient ways to represent our data and identify correlations.



Identify (linear) correlations.

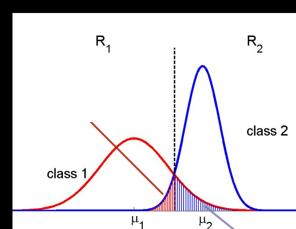


Data ~ N(μ,Σ)

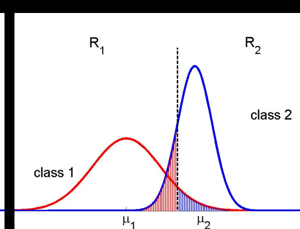
Univariate analysis.

Bayes Criterion

- In classification Bayes optimality is generally preferred over other criteria.



Bayes classifier



Another classifier

$$\int_{R_1} P_2 p(\mathbf{x} \mid \text{class 2})\, d\mathbf{x} + \int_{R_2} P_1 p(\mathbf{x} \mid \text{class 1})\, d\mathbf{x}$$

Multivariate analysis.

Discriminant Analysis

- **Goal**: Search for those linear combination of features in $\mathbf{R}^p$ that best classify the data.
- **Problem**: It is impossible to check for all possible solutions.
- **Solution**: Use criteria that can be easily minimized (or maximized). One of the most known is the Fisher criterion given by

$$\arg\max_{\mathbf{v}} \frac{\left|\mathbf{v}^T \mathbf{M}_A \mathbf{v}\right|}{\left|\mathbf{v}^T \mathbf{M}_B \mathbf{v}\right|}$$

Metrics

Least-squares solution.

$$\mathbf{M}_A \mathbf{V} = \mathbf{M}_B \mathbf{V} \Lambda.$$

Martinez & Zhu, PAMI, 2005

Linear Discriminant Analysis

Between-class scatter matrix:

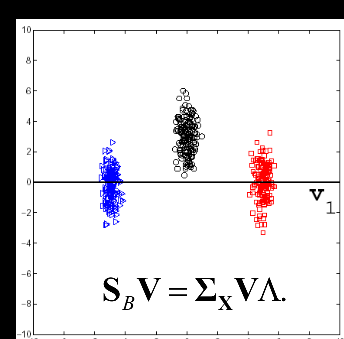
$$\mathbf{M}_A = \mathbf{S}_B = \frac{1}{C} \sum_{i=1}^{C} (\mu_i - \mu)^T (\mu_i - \mu).$$

Scatter matrix:

$$\mathbf{M}_B = \Sigma_X = \sum_{i=1}^{C} \Sigma_i.$$

Scatter matrix of class i:

$$\Sigma_i = \sum_{j=1}^{n_i} (\mathrm{x}_{ij} - \mu)^T (\mathrm{x}_{ij} - \mu).$$



$$\arg\max_{\mathbf{v}} \frac{\left|\mathbf{v}^T \mathbf{M}_A \mathbf{v}\right|}{\left|\mathbf{v}^T \mathbf{M}_B \mathbf{v}\right|} \Leftrightarrow \mathbf{M}_A \mathbf{V} = \mathbf{M}_B \mathbf{V} \Lambda$$

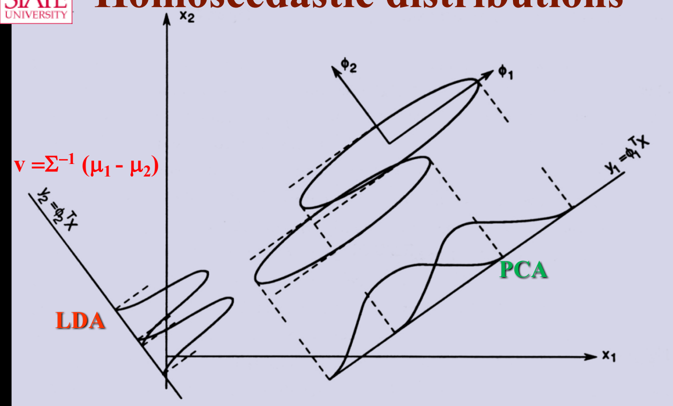
Martinez & Zhu, PAMI, 2005

Fisher's Insight

- **Theorem:** Let the samples of two classes be Normally distributed in $\mathbf{R}^p$, with common covariance matrix. Then, the classification errors in the p-dimensional space and that in the one-dimensional subspace given by $\mathbf{v} = \Sigma^{-1}(\mu_1 - \mu_2)/\|\Sigma^{-1}(\mu_1 - \mu_2)\|$, are the same; where $\|\mathbf{x}\|$ is the Euclidean distance (2-norm) of the vector $\mathbf{x}$.
- That is, there is **no** loss in classification when reducing from p dimensions to **one** – Bayes optimal.

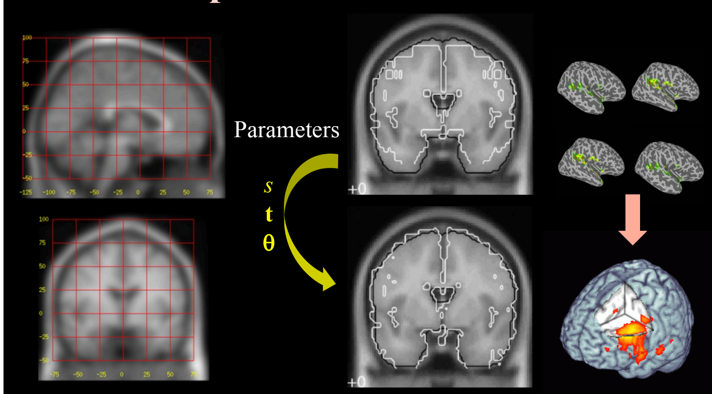
Martinez & Zhu, PAMI, 2005

Homoscedastic distributions



An example of feature extraction for classification.

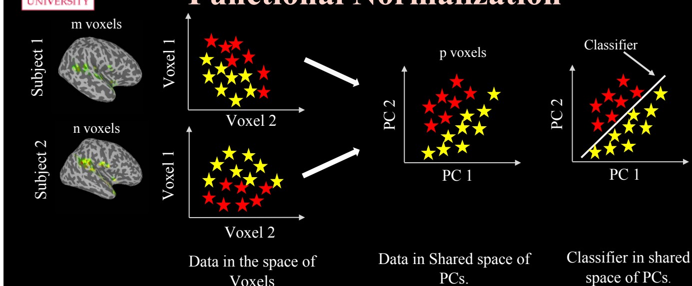
Standard Brain Alignment Spatial Normalization



MNI 305

Least-squares solution

PCA Alignment Functional Normalization

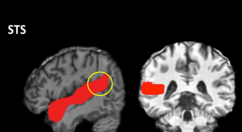


- Assumption – pattern of neural activation across subjects has the same variance.
- Map the activation pattern onto the space defined by the principal components (PCs).
- Train a linear classifier to decode AUs (e.g., LDA).

Srinivasan, Golomb, Martinez, JNeuro 2016

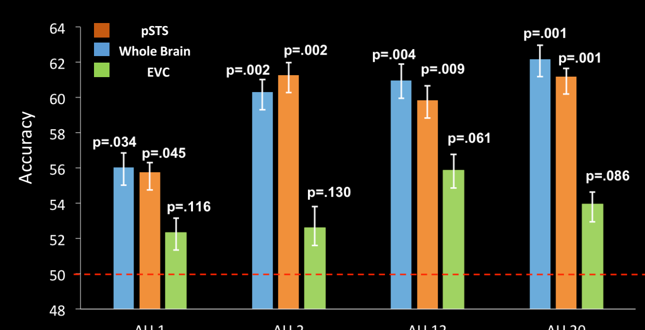
A Neural Basis of Facial Actions

- Action Units (AUs) – visible anatomical changes.
- **Hypothesis – posterior Superior Temporal Sulcus** (pSTS) is dedicated to the visual representation / interpretation of facial actions.



Srinivasan, Golomb, Martinez, JNeuro 2016

MVPA RESULTS

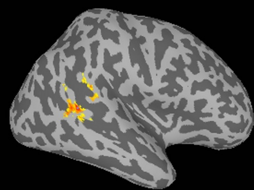


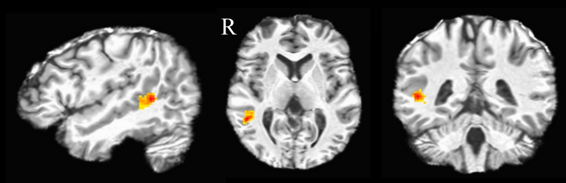
Leave-one-subject out cross-validation (10 subjects).

Srinivasan, Golomb, Martinez, JNeuro 2016

MOST DISCRIMINANT VOXELS
Whole brain analysis

- Since we used linear transformations – PCA and LDA – we can now invert these transformations to identify the voxels that most contribute to decoding AUs.





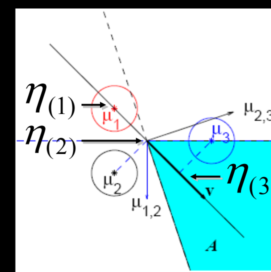
Srinivasan, Golomb, Martinez, JNeuro 2016

Bayes Optimal Homoscedastic LDA for *C* Classes

- There is a sequences of projected means that minimize the Bayes error.
- Every sequence defines a convex region.
- The Bayes error is given by

cdf

$$g(\mathbf{v}) = 2C^{-1}\sum_{i=1}^{C-1}\Phi\left(\frac{\eta_{(i)}-\eta_{(i+1)}}{2}\right)$$



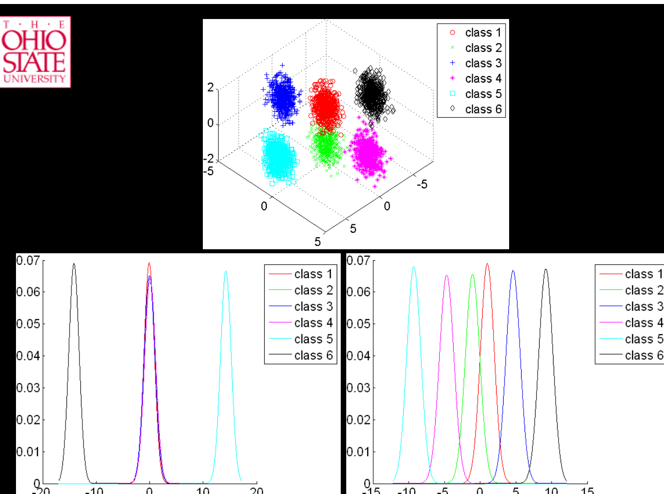
Hamsici & Martinez, PAMI, 2008.

Theorem

- Define a constrained region A where all vectors $\mathbf{v}$ sampled from it generate the same ordered sequence. Let $g(\mathbf{v})$ be the Bayes error function of the C homoscedastic Gaussian distributions in A. Then, the region A is a convex polyhedron, and the Bayes error function $g(\mathbf{v})$ for all $\mathbf{v}$ in A is also convex.

We can use convex optimization algorithms

Hamsici & Martinez, PAMI, 2008.

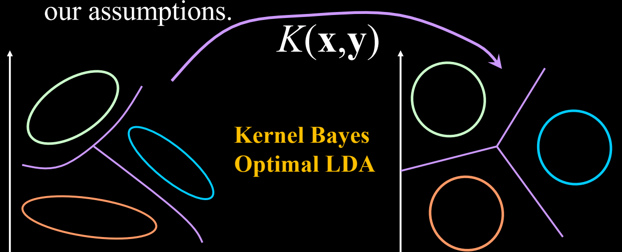


LDA (least squares)

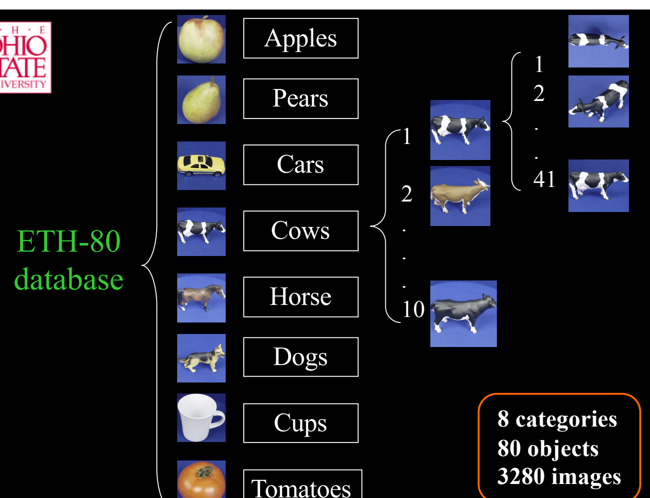
Bayes Optimal

Heteroscedastic case

- In general the data distributions are not homoscedastic.
- We can first map the data into a space of larger dimensionality where the new densities conform to our assumptions.



Hamsici & Martinez, PAMI, 2008.



Hamsici & Martinez, PAMI, 2008.

Experimental Results ETH-80

Dimensionality (d)	1	2	3	4	5	6	7
Bayes Opt. LDA	53.38	66.25	71.46	72.29	76.65	77.87	81.68
Kernel Opt. LDA	**81.62**	**84.70**	**86.01**	**88.29**	**89.21**	**89.48**	**92.04**
Linear app. LDA	52.65	63.84	73.02	72.20	77.44	80	81.68
Linear app. LDA opp.	53.38	67.99	70.55	72.04	76.40	79.63	81.68
LDA	46.16	61.10	68.99	69.51	73.32	78.08	81.68
LDA_{opt}	53.54	68.23	70.98	73.20	75.15	77.10	81.67
$aPAC$	44.70	62.59	68.81	70.73	74.45	80.15	81.68
$aPAC_{opt}$	53.54	68.93	71.31	73.72	76.52	78.51	81.68
$FLDA$	41.49	60.82	67.53	74.33	77.35	79.24	81.68
$FLDA_{opt}$	49.24	62.32	69.97	73.87	77.53	79.89	81.68
$DFLDA$	28.05	39.73	48.51	53.93	56.07	58.9	63.29
$DFLDA_{opt}$	40.7	54.15	60.15	61.92	63.26	62.44	63.29
$PCA-LDA$	46.19	62.93	63.08	64.27	64.42	66.25	66.07
$PCA-LDA_{opt}$	46.34	62.96	63.81	64.33	64.57	66.77	66.07
$KLDA$	62.80	74.18	73.32	78.69	79.97	85.85	**91.83**

Subclass Discriminant Analysis (SDA)

Between-subclass scatter matrix:

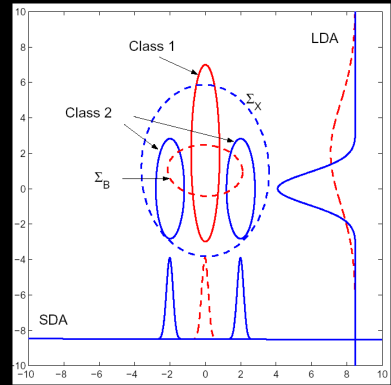
$$\Sigma_B = \sum_{i=1}^{C} \sum_{j=1}^{H_i} p_{ij} (\mu_{ij} - \mu)^T (\mu_{ij} - \mu)$$

Basis vectors:

$$\Sigma_B V = \Sigma_X V \Lambda.$$

How many subclasses (H):

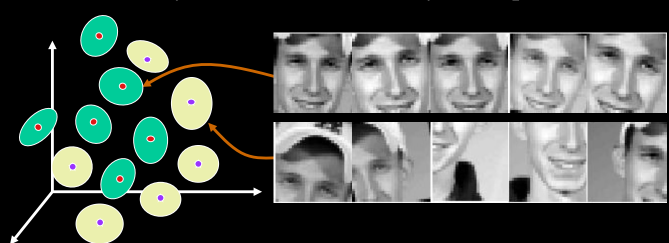
Minimize conflict criterion.



Zhu & Martinez, PAMI, 2006

Features vs. context

Observation: Most detections are near the correct location – they are not incorrect, they are *imprecise*.

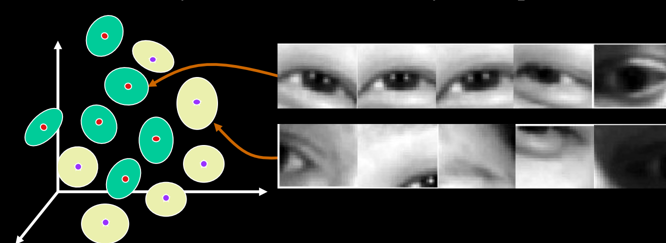


Key idea: Use context information to train where **not** to detect faces and facial features.

Ding & Martinez, CVPR, 2008; PAMI, 2010

Features vs. context

Observation: Most detections are near the correct location – they are not incorrect, they are *imprecise*.

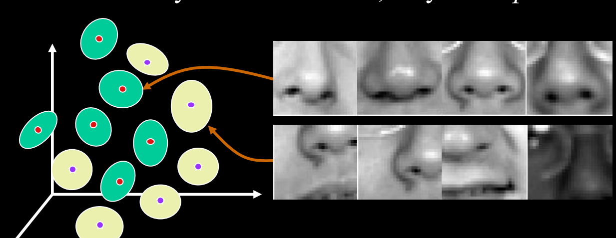


Key idea: Use context information to train where **not** to detect faces and facial features.

Ding & Martinez, CVPR, 2008; PAMI, 2010

Features vs. context

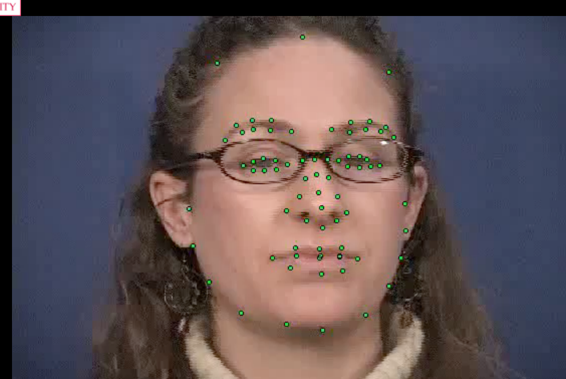
Observation: Most detections are near the correct location – they are not incorrect, they are *imprecise*.



Key idea: Use context information to train where **not** to detect faces and facial features.

Ding & Martinez, CVPR, 2008; PAMI, 2010

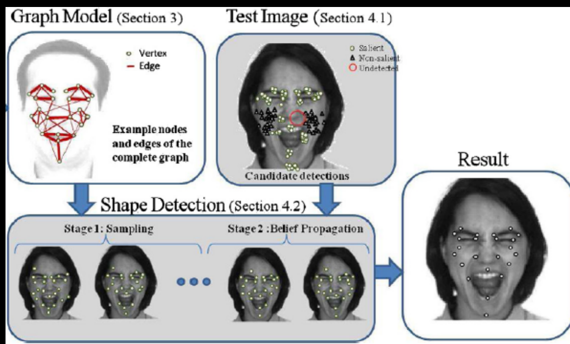
Precise Detailed Detection



Error: 6.2 pixels (2%) **vs** Manual: 4.2 (1.5%)

Ding & Martinez, CVPR, 2008; PAMI, 2010

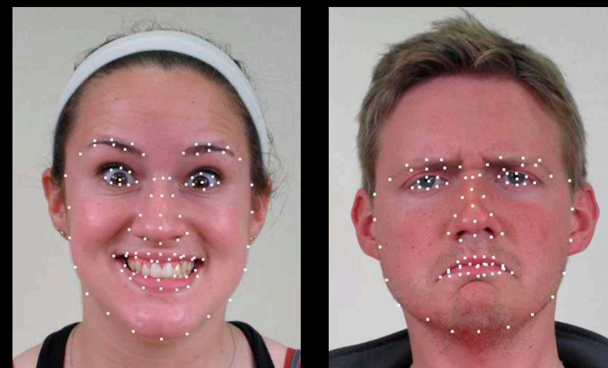
Conditional Probabilities



Probability of the location of fiducial point *i* given the location of *j*.

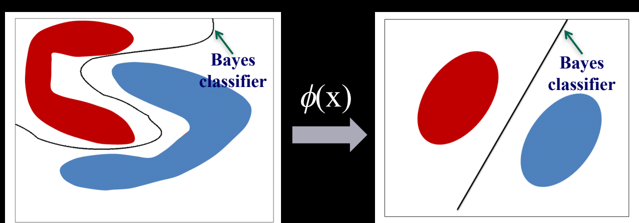
Benitez-Quiroz et al., PR 2014; Du et al., PNAS 2014

Conditional Probabilities



Du et al., PNAS 2014

Kernel SDA (KSDA)



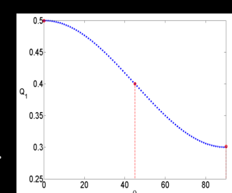
- If the problem is not linearly separable, we can use a kernel.
- We need to determine the kernel map (metric), $k(\theta)$.

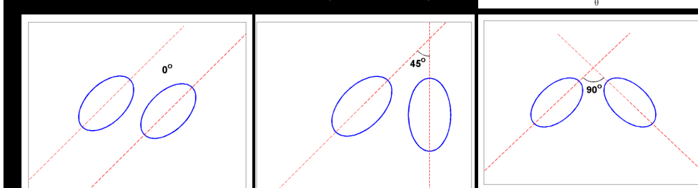
You , Hamsici & Martinez, PAMI, 2011; You & Martinez, CVPR, 2010.

Homoscedastic Criterion

- Recall that DA algorithms (e.g., BDA, LDA, SDA) are Bayes optimal when the class distributions are homoscedastic:

$$Q_1(\mathbf{x}) = \frac{2}{C(C-1)}\sum_{i=1}^{C-1}\sum_{k=i+1}^{C}\frac{tr(\Sigma_i^{\phi}\Sigma_k^{\phi})}{tr(\Sigma_i^{\phi^2})+tr(\Sigma_k^{\phi^2})}.$$

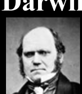




You , Hamsici & Martinez, PAMI, 2011.

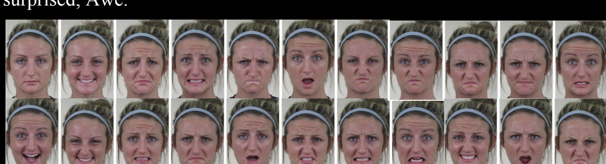
Compound Facial Expressions

- **Darwin:** 6 emotion categories.



Charles Darwin
1809—1882

- **21 emotion categories:** Happy, Sad, Angry, Fear, Surprise, Disgust, Angrily disgusted, Appalled, Hatred, Angrily surprised, Disgustedly sad, Disgustedly surprised, Fearfully angry, Fearfully disgusted, Fearfully sad, Happily disgusted, Happily surprised, Sadly angry, Sadly surprised, Fearfully surprised, Awe.

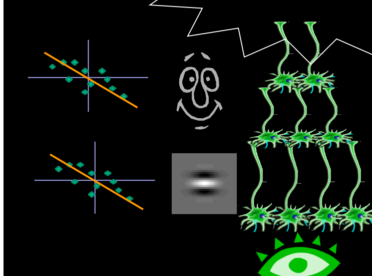


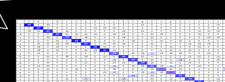
Du, Tao & Martinez, PNAS 2014

Computational Model

Classification

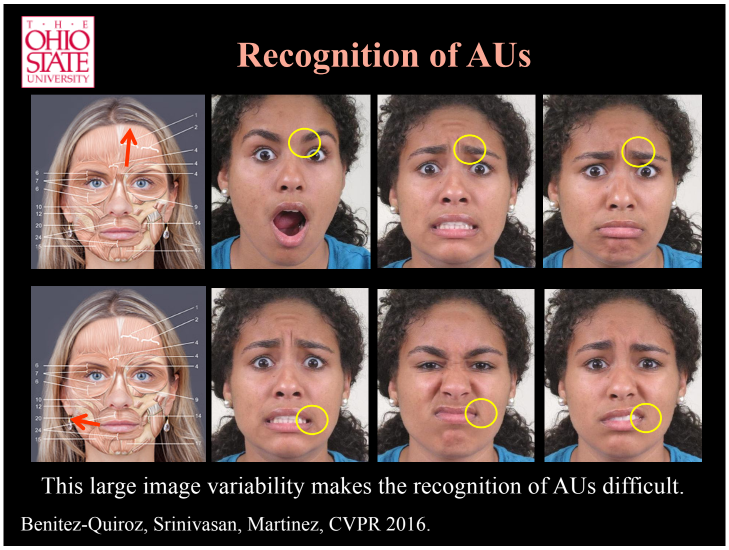
KSDA: ~77%
SVM: ~50%

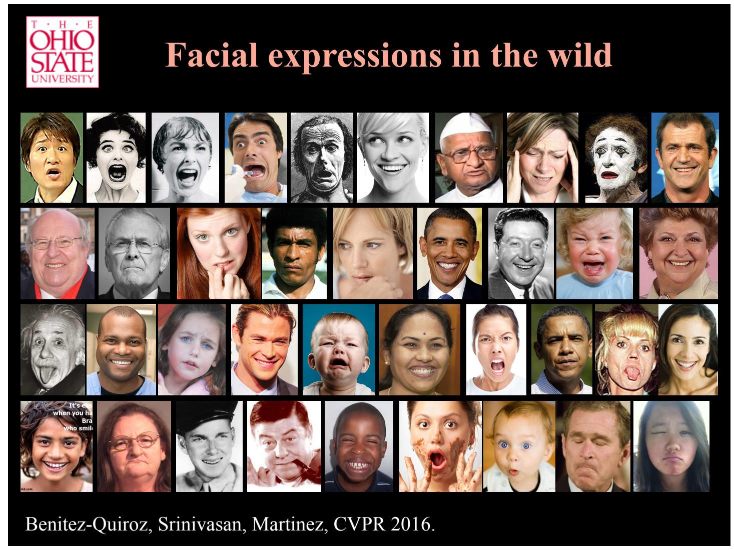


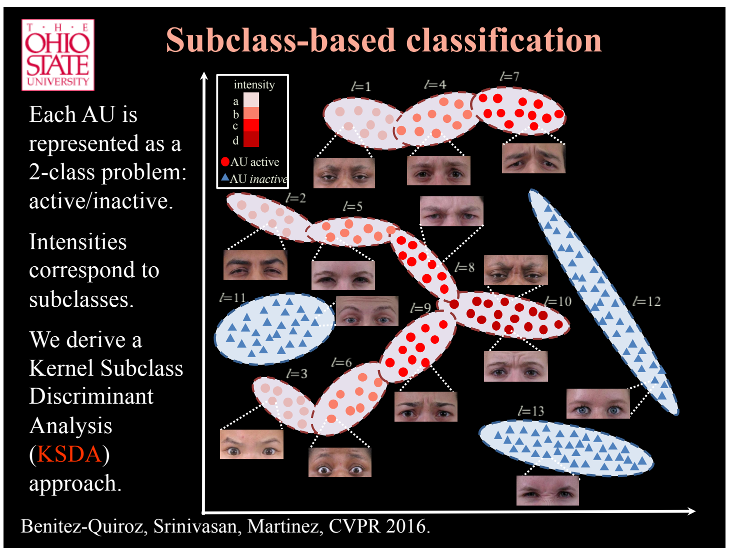


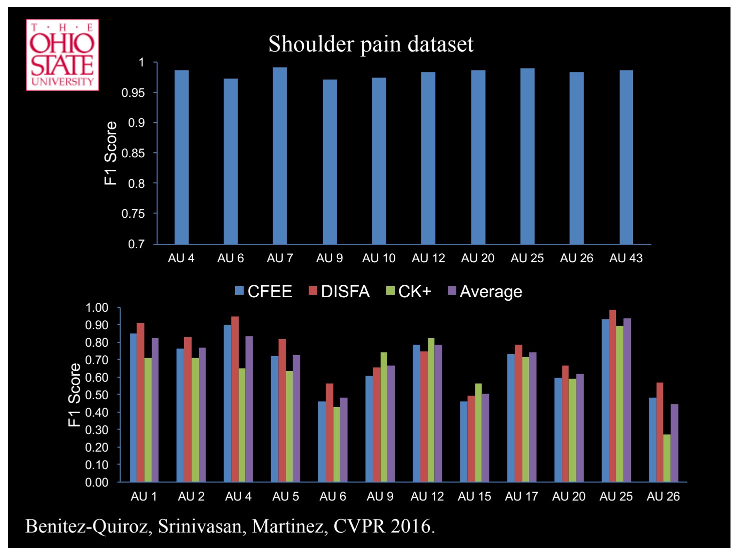
- Results show all categories can be visually discriminated.
- Recognition correlates with human results (0.8).
- Results correlate with AU similarity (0.7).

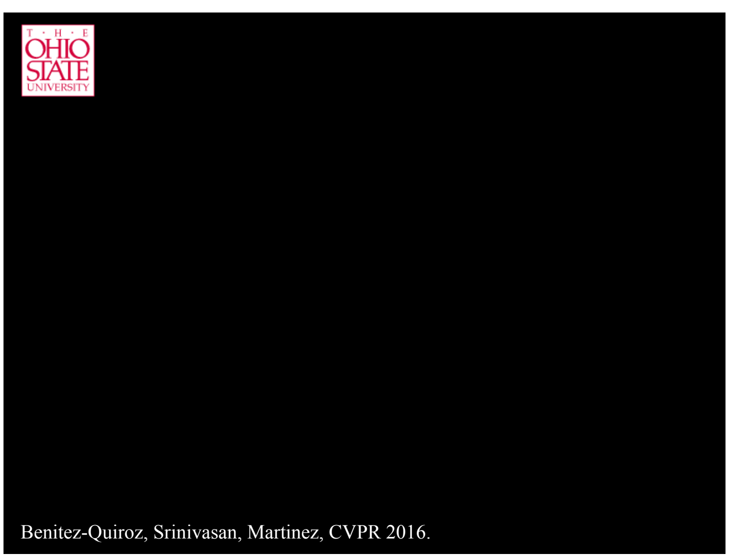
Du, Tao & Martinez, PNAS 2014

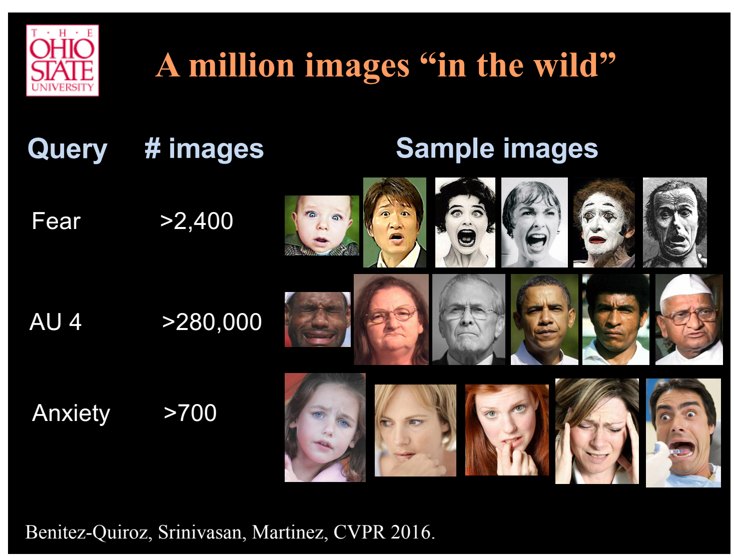

THE OHIO STATE UNIVERSITY
Recognition of AUs
This large image variability makes the recognition of AUs difficult.
Benitez-Quiroz, Srinivasan, Martinez, CVPR 2016.


THE OHIO STATE UNIVERSITY
Facial expressions in the wild
Benitez-Quiroz, Srinivasan, Martinez, CVPR 2016.


THE OHIO STATE UNIVERSITY
Subclass-based classification
Each AU is represented as a 2-class problem: active/inactive.
Intensities correspond to subclasses.
We derive a Kernel Subclass Discriminant Analysis (KSDA) approach.
intensity
a
b
c
d
AU active
AU inactive
Benitez-Quiroz, Srinivasan, Martinez, CVPR 2016.

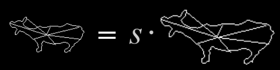

THE OHIO STATE UNIVERSITY
Shoulder pain dataset
F1 Score
AU 4
AU 6
AU 7
AU 9
AU 10
AU 12
AU 20
AU 25
AU 26
AU 43
CFEE
DISFA
CK+
Average
F1 Score
AU 1
AU 2
AU 4
AU 5
AU 6
AU 9
AU 12
AU 15
AU 17
AU 20
AU 25
AU 26
Benitez-Quiroz, Srinivasan, Martinez, CVPR 2016.


THE OHIO STATE UNIVERSITY
Benitez-Quiroz, Srinivasan, Martinez, CVPR 2016.


THE OHIO STATE UNIVERSITY
A million images "in the wild"
Query
images
Sample images
Fear
>2,400
AU 4
>280,000
Anxiety
>700
Benitez-Quiroz, Srinivasan, Martinez, CVPR 2016.

Spherical Distributions

- **Shape-based object recognition:** we would like our algorithm to be invariant to scale and in-plane rotations.



- **Appearance-based recognition:** brightness intensity should not affect recognition.

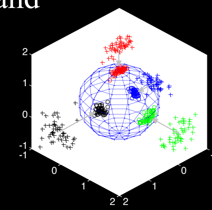
Norm normalization

Hamsici & Martinez, Journal of Machine Learning Research, 2007.

"Spherical" Kernels

- Many kernel functions result in a spherical representation.
- For example, the well-known and commonly used RBF:

$$k(\mathbf{x},\mathbf{y}) = \exp\left(\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\varsigma}\right).$$

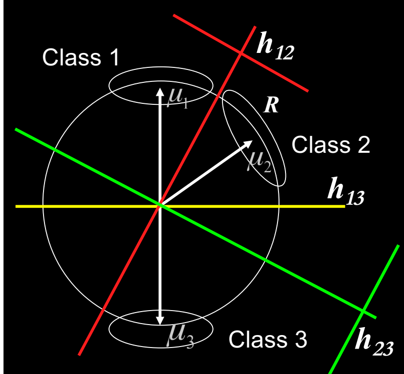


- Also, the Mahalanobis kernel:

$$k(\mathbf{x},\mathbf{y}) = \exp\left(\frac{(\mathbf{x}-\mathbf{y})^T\Sigma^{-1}(\mathbf{x}-\mathbf{y})}{2\varsigma^2}\right).$$

Hamsici & Martinez, Journal of Machine Learning Research, 2007.

Homoscedastic and Spherical-homoscedastic



$N_i(\mu_i,\Sigma_i)$ and $N_j(\mu_j,\Sigma_j)$ are *Homoscedastic* if $\Sigma_i = \Sigma_j$.

Definition *Spherical-homoscedastic*: Any two pdf where $\lambda_{ik} = \lambda_{jk}$ for all k and $\mathbf{h}_{ij}$ are hyperplanes.

Hamsici & Martinez, Journal of Machine Learning Research, 2007.

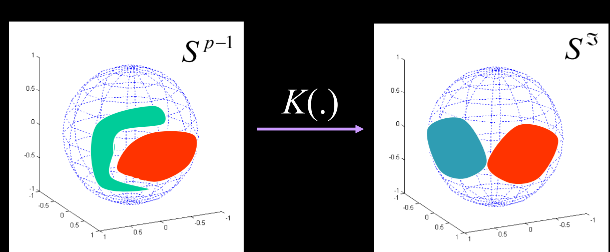
Theorem

If we estimate two spherical-homoscedastic (e.g., vMF, Bingham, Kent) distributions using Gaussian pdf instead, then the **Bayes classifier** obtained with these Gaussians is the same as the Bayes classifier calculated with the original spherical pdf. They correspond to two hyperplanes. One hyperplane partitions the hypersphere in two, the other is outside it and is hence irrelevant.

Hamsici & Martinez, Journal of Machine Learning Research, 2007.

Kernel Spherical-Homoscedastic

- As we did earlier, we can define a kernel, $K(\mathbf{x},\mathbf{y})$, and optimize its parameters to adapt to the spherical-homoscedastic case.

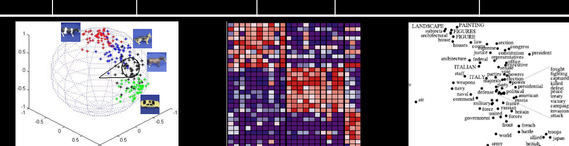


This allows us to define Bayes optimal linear classifiers

Hamsici & Martinez, Journal of Machine Learning Research, 2007.

Experimental Results

	vMF	Bingham	LDA	K-SH-vMF	K-SH-Bingham
ETH	13.75	73.11	62.9	**79.24**	**78.84**
CNS tumors	28.57	14.29	21.4	**88.10**	**90.48**
Text	38.64	N/A	76.4	**90.21**	**88.92**



Hamsici & Martinez, Journal of Machine Learning Research, 2007.

Rotation Invariant Kernels

- We can define a kernel which carries the much needed rotation invariance.
- This eliminates the requirement of working with complex symmetric distributions.
- It can even be used to represent **3D shapes**:

$$k(z_j,z_k)=\exp\left(-\frac{\left\|z_j-z_k\exp(i\theta_{z_jz_k})\right\|^2}{2\sigma^2}\right)=\exp\left(-\frac{1-\left\|z_j^*z_k\right\|}{\sigma^2}\right).$$

Optimize SH

Hamsici & Martinez, PAMI, 2009.

Results

- Kernel criterion: optimize spherical-homoscedastisity (SH).
- ETH-80 (2D shapes):

	Proc. NM	Proc. NN	Proc. TS	Kernel Proc.	*Kents' Hybrid*	*complex Bingham*	*complex Normal*	RIK_{CV}	*RIK*
Recognition Rate	79.02	82.10	79.12	86.34	79.66	86.95	87.5	91.22	**92.29**
Training Time (in seconds)	0.21	N/A	1.25	1680.4	1.06	18.34	**0.95**	3049.8	89
Testing Time (in seconds)	**0.01**	7.05	0.16	2.16	0.05	**0.02**	**0.02**	3.07	3.50

- Face recognition (FRGC 3D shapes):

	Proc. NM	Proc. NN	Proc. TS	Kernel Proc.	*Kents' Hybrid*	*complex Bingham*	*complex Normal*	RIK_{CV}	*RIK*
Recognition Rate	47.50	68.90	46.75	93.25	42.28	9.33	41.73	**94.78**	**94.13**
Training Time (in seconds)	36.97	N/A	1.34	40.38	5.60	70.33	5.60	2594.4	263.4
Testing Time (in seconds)	2.97	35.53	**0.02**	0.09	1.86	0.11	0.46	**.01**	**.01**

Hamsici & Martinez, PAMI, 2009.

3D AAMs with RIK

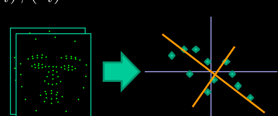
- A main advantage of RIK is that there is a close form solution for 3D shapes:

$$k(\hat{\mathbf{s}}_j,\hat{\mathbf{s}}_k)=\exp\left(-\frac{\|\mathbf{z}_j^1-\mathbf{z}_k^1\exp^{i\theta_{jk}}\|^2+\|\mathbf{z}_j^2-\mathbf{z}_k^2\exp^{i\theta_{jk}}\|^2}{2\sigma^2}\right)$$

$$\|\mathbf{z}_j^l-\mathbf{z}_k^l\exp^{i\theta_{jk}^2}\|^2=\mathbf{z}_j^{l*}\mathbf{z}_j^l+\mathbf{z}_k^{l*}\mathbf{z}_k^l-2\|\mathbf{z}_j^{l*}\mathbf{z}_k^l\|$$

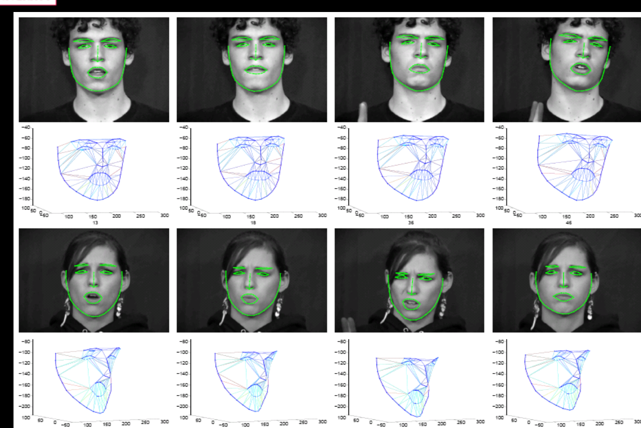
- The covariance matrix is computed in this kernel space: $C=\frac{1}{n}\sum_{i=1}^{n}\phi(s_i)\phi(s_i)^T$
- And the components:

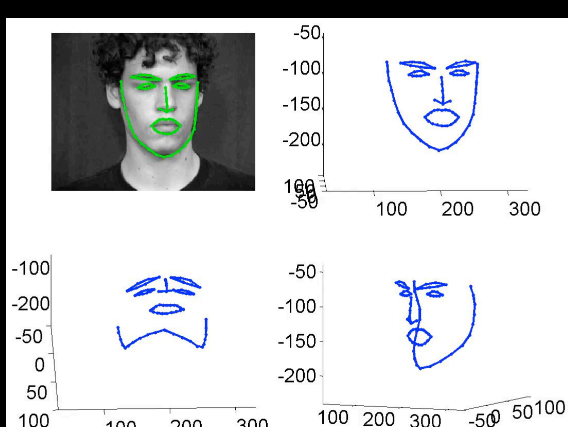
$$\langle\mathbf{v}^m,\phi(s_{new})\rangle=\sum_{i=1}^{n}\alpha_i^m k(s_i,s_{new})$$



Hamsici & Martinez, ICCV, 2009.

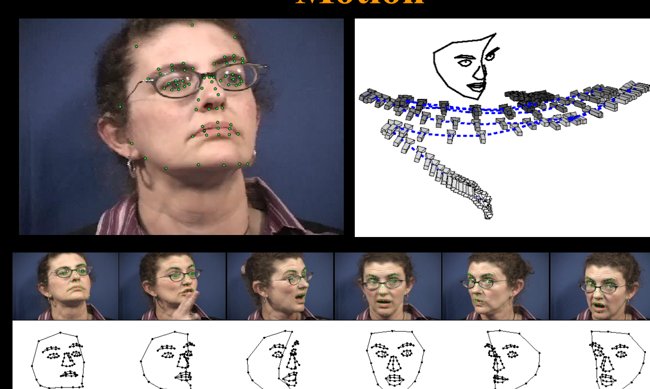
Results: ASL nonmanuals





Hamsici & Martinez, ICCV, 2009.

Non-Rigid Structure from Motion



Gotardo & Martinez, PAMI, 2010; Gotardo & Martinez, CVPR 2011.

Standard Approach in Rigid SFM

Decomponse $\mathbf{W}$ into a product of a camera (motion) factor $\mathbf{M}$ and a 3D shape factor $\mathbf{S}$:

$$\underbrace{\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ y_{11} & y_{12} & \dots & y_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T1} & x_{T2} & \dots & x_{Tn} \\ y_{T1} & y_{T2} & \dots & y_{Tn} \end{bmatrix}}_{\mathbf{W}} = \underbrace{\begin{bmatrix} \boxed{\widehat{\mathbf{R}}_1 \quad \widehat{\mathbf{t}}_1} \\ \widehat{\mathbf{R}}_2 \quad \widehat{\mathbf{t}}_2 \\ \vdots \quad\quad \vdots \\ \widehat{\mathbf{R}}_T \quad \widehat{\mathbf{t}}_T \end{bmatrix}}_{\mathbf{M}} \underbrace{\begin{bmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \\ Z_1 & Z_2 & \dots & Z_n \\ 1 & 1 & \dots & 1 \end{bmatrix}}_{\mathbf{S}}$$

$$\begin{bmatrix} \widehat{\mathbf{R}}_1 & \widehat{\mathbf{t}}_1 \end{bmatrix} \rightarrow \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \end{bmatrix}$$

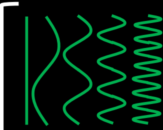
Rank constraint: solve for rank-4 $\mathbf{M}$ and $\mathbf{S}$ using SVD

$$\mathbf{W} = \mathbf{U\Sigma V}^T = \underbrace{(\mathbf{U\Sigma}^{\frac{1}{2}})}_{\mathbf{M}}\underbrace{(\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{V}^T)}_{\mathbf{S}}$$

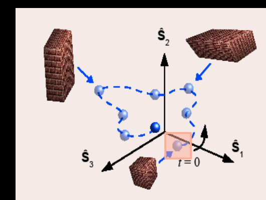
Shape Trajectory Approach (STA)

$$\mathbf{W} = \begin{bmatrix} \widehat{\mathbf{R}}_1 & & & \\ & \widehat{\mathbf{R}}_2 & & \\ & & \ddots & \\ & & & \widehat{\mathbf{R}}_T \end{bmatrix} \left(\underbrace{\begin{bmatrix} \boxed{c_{1,1} \quad c_{1,2} \quad \dots \quad c_{1,K}} \\ c_{2,1} \quad c_{2,2} \quad \dots \quad c_{2,K} \\ \vdots \quad\quad \vdots \quad\quad \ddots \quad\quad \vdots \\ c_{T,1} \quad c_{T,2} \quad \dots \quad c_{T,K} \end{bmatrix}}_{\mathbf{C}} \otimes \mathbf{I}_3 \right) \underbrace{\begin{bmatrix} \widehat{\mathbf{S}}_1 \\ \widehat{\mathbf{S}}_2 \\ \vdots \\ \widehat{\mathbf{S}}_K \end{bmatrix}}_{\mathbf{S}}$$

In our model,

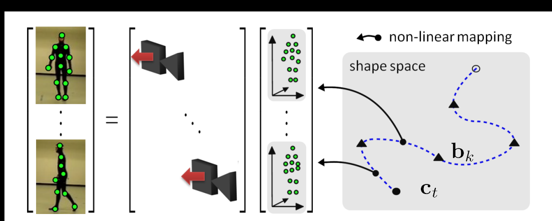
$\mathbf{C} = \Big[$  $\dots \Big] \mathbf{X}$

DCT basis vectors



Gotardo & Martinez, PAMI, 2010; Gotardo & Martinez, CVPR 2011.

Kernel Non-rigid SFM



- Consider, W=MS, $\mathbf{S} = \mathbf{M}^\dagger\mathbf{W}$, where $\mathbf{M}^\dagger = \mathbf{M}^T(\mathbf{MM}^T)^\dagger$ Then:

$$\widehat{\mathbf{W}} = \mathbf{M}\underbrace{\mathbf{M}^T(\mathbf{MM}^T)^\dagger\mathbf{W}}_{\mathbf{S}}, \quad \mathbf{MM}^T = \mathbf{D}(\underbrace{\mathbf{CC}^T}_{\mathbf{K}} \otimes \mathbf{I}_3)\mathbf{D}^T$$

$$\mathbf{c}_t^T\mathbf{c}_{t'} \quad \Rightarrow \quad \kappa(\mathbf{c}_t, \mathbf{c}_{t'}) = e^{-\gamma\|\mathbf{c}_t - \mathbf{c}_{t'}\|_2^2}$$

Gotardo & Martinez, ICCV 2011.

KNSFM with RIK or aSFM

We propose a new **aSFM** kernel:

$$\kappa(\mathbf{w}_t, \mathbf{w}_{t'}) = \exp\left(\frac{-r_{t,t'}^2}{\sigma^2}\right), \quad r_{t,t'} = \left\| \begin{bmatrix} \mathbf{w}_t \\ \mathbf{w}_{t'} \end{bmatrix} - \begin{bmatrix} \mathbf{M}_t \\ \mathbf{M}_{t'} \end{bmatrix} \mathbf{S}_a \right\|_F$$

Spatial smoothness controlled by scale parameter

And the **2D RIK** is:

$$\mathbf{z}_t = \frac{(\mathbf{w}_t)^T}{\|\mathbf{w}_t\|_F} \begin{bmatrix} 1 \\ \sqrt{-1} \end{bmatrix} \in \mathbb{C}^n \quad \Rightarrow \quad \kappa(\mathbf{z}_t, \mathbf{z}_{t'}) = \exp\left(\frac{-1 + |\mathbf{z}_t^*\mathbf{z}_{t'}|}{\sigma^2}\right)$$

New, customized RIKs can even take advantage of <u>object appearance</u> when correlated with 3D shape.

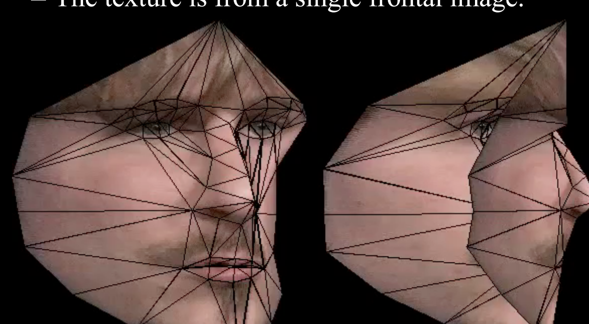
Hamsici, Gotardo & Martinez, ECCV 2012.

Reconstructed 3D Shape

Gotardo & Martinez, ICCV 2011; Hamsici, Gotardo & Martinez, ECCV 2012.

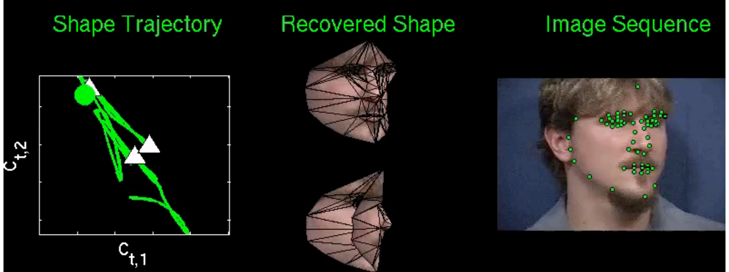
Reconstructd 3D Shape

- Results after triangulation and texture mapping
  - The texture is from a single frontal image.



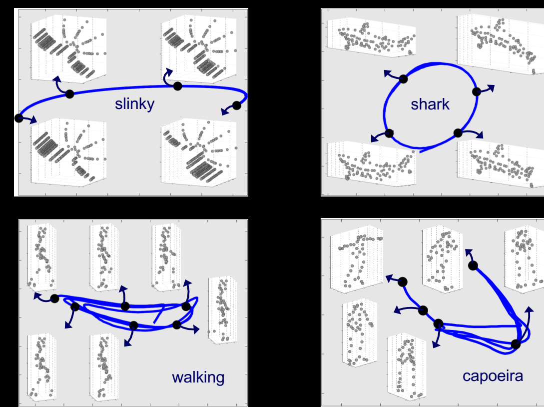
Gotardo & Martinez, ICCV 2011; Hamsici, Gotardo & Martinez, ECCV 2012.

KNSFM compactness



Gotardo & Martinez, ICCV 2011; Hamsici, Gotardo & Martinez, ECCV 2012.

KNSFM compactness

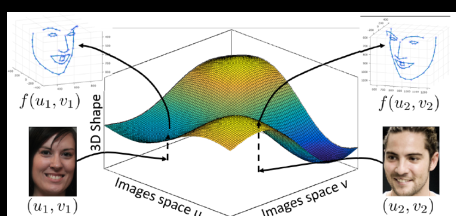


Gotardo & Martinez, ICCV 2011; Hamsici, Gotardo & Martinez, ECCV 2012.

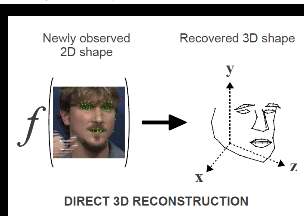
2D to 3D mapping

Advantage: Learned mapping is a by-product of the NRSFM solution

$$\mathbf{c}_\tau^T = f(\mathbf{w}_\tau)$$

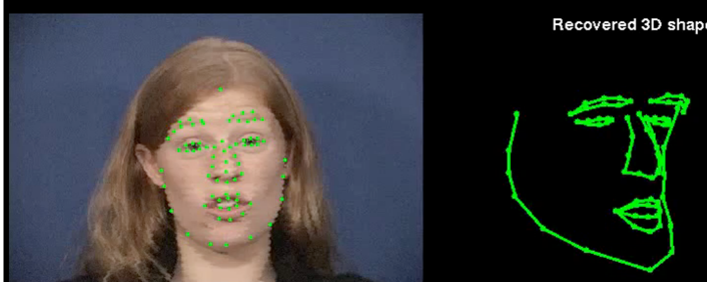


NRSFM becomes a "training" stage in which we learn the mapping f and the basis shapes in factor $\mathbf{S}$.



Hamsici, Gotardo & Martinez, ECCV 2012.

Examples of 2D to 3D mapping



Hamsici, Gotardo & Martinez, ECCV 2012.

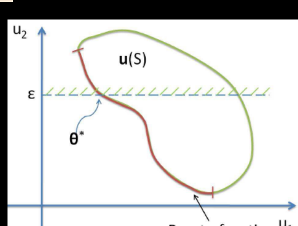
Kernel Methods in Regression

- We generally have two or more criteria to minimize, e.g., model fit (E_f) and model complexity (E_c) of a kernel map K(.).

$$\underset{\theta}{\text{minimize}} \quad u_1(\theta), u_2(\theta), \ldots, u_k(\theta)$$
$$\text{subject to} \quad \theta \in S,$$

where $u_i(\theta)$ are the objective functions, S in R^p.

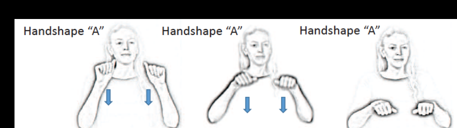
Pareto-optimality: A solution θ^* with $u_i(\theta^*) > u_i(\theta)$, for any other θ, i=1,…,k.

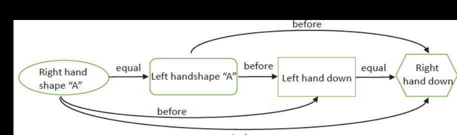


You, Benitez-Quiroz & Martinez, IEEE TNNLS 2014.

Behavior Analysis Labeled Graphs

Behavior is defined using high-level concepts/attributes, e.g., "left hand" "moves down," "right hand" "same."



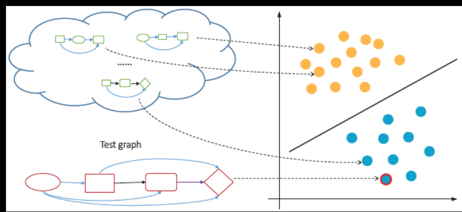


Definition: A directed labeled graph is $G_k=(V_k, E_k, L, f_k)$, V_k the nodes, E_k the edges, L a set of labels, and f_k a function that assigns labels to nodes and edges,

Zhao & Martinez, IEEE PAMI, 2016.

Labeled Graphs

Goal: find a feature representation of labeled graphs. $\mathbf{x}_i$ specifies the numbers of times a path P occurs in G_i.



$$K(G_i, G_j) = \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j = \sum_{a \equiv \text{paths of lenght } z} \sum_{b \equiv \forall \text{path}} x_{iab} x_{jab} \,.$$

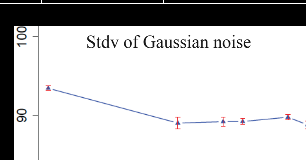
Zhao & Martinez, IEEE PAMI, 2016.

Labeled Graphs

ASL databases – concept classification.

Method	Accuracy	Time/sample
This approach	91.60%	1.62 s
OSCM	70%	9.43 s
TGAK	88%	14.35 s
Path kernel	89.69	--
RW kernel	89.70%	2.27

Robustness to Gaussian noise.



Zhao & Martinez, IEEE PAMI, 2016.

Acknowledgments

National Institutes of Health

National Eye Institute

National Institute on Deafness and Other Communication Disorders

Google Faculty Research Awards

http://cbcsl.ece.ohio-state.edu/

HFSP

NSF

Thank you!