Department of Electrical & Computer Engineering

How the Face and Hands Talk

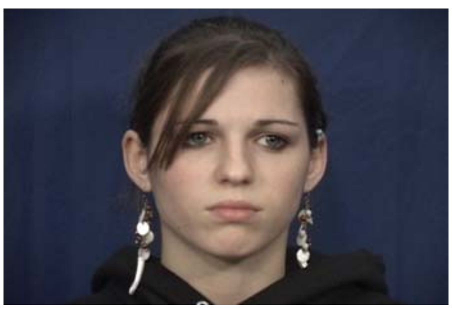
Adventures in the modeling of sign languages

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

Hands: manuals

Martinez et al., ICMI, 2002

Face: nonmanuals



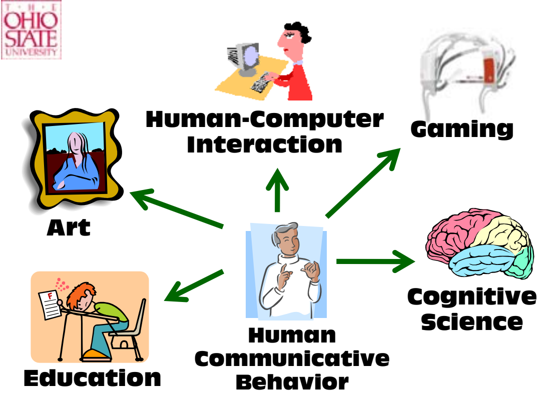
Nonmanuals are part of the **grammar**.

Martinez & Wilbur, 2008

American Sign Language

WH questions

Yes/no questions



The Hands (manuals)

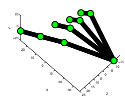
- Manual signs are constructed using three building blocks: 1) handshape, 2) motion, and 3) place of articulation.

search drink family class

Handshape

- We need to recover the 3D shape of the hand from a sequence of 2D views.
- The major problems are noise (i.e., imprecise detection of knuckles) and occlusions.

$$\mathbf{P}=\{\mathbf{p}_1,\ldots,\mathbf{p}_n\},\ \mathbf{p}_i=(x_i,y_i,z_i)^T.$$

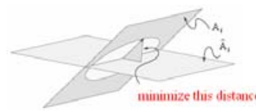


Ding & Martinez, AVSS, 2007

Structure from Motion

- Any 3-column of the data matrix defines a subspace, $\hat{A}_i$.
- A 3-column of the (ideal) noise-free data matrix defines another subspace, A_i.
- The perturbation between two matrices is related to that of the spanning spaces:

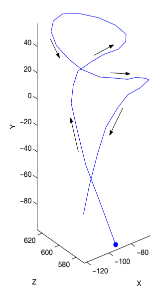
$$pb(Q_i,\hat{Q}_i)=\theta(A_i,\hat{A}_i) \equiv \theta(span(Q_i),span(\hat{Q}_i))$$



Jia & Martinez, PAMI, *to appear*

Motion

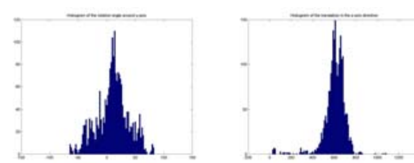
- To recover the 3D motion, we need to compute the 3D translation and rotation at each frame.
- Since we already have the 3D handshape, we can use any of the 3-point resection algorithms to determine the missing parameters.
- Unfortunately, such solutions are very sensitive to noise, which is a particular problem in this application.



Ding & Martinez, AVSS, 2007

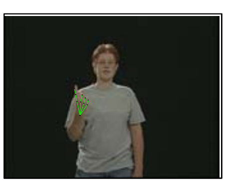
Motion

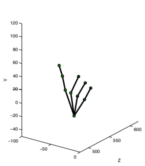
- To resolve this issue, we create a 3x3 window around each fiducial (e.g. knuckle).
- Each position within this window corresponds to a potentially correct location of the fiducial. The median is our solution.

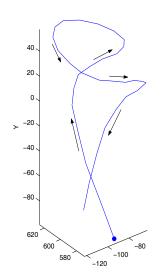


Ding & Martinez, AVSS, 2007

Handshape & Motion: 3D Reconstruction



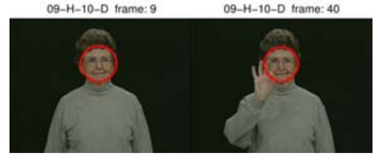




Ding & Martinez, AVSS, 2007

Place of Articulation

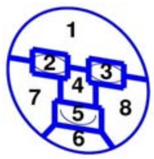
- We use the face as a center of reference.
- Use a face detector of your choice.
- Employ a Gaussian model to detect false positive/negatives. Model the center of the face and its radius to facilitate detection of outliers.



Ding & Martinez, 2008

Place of Articulation

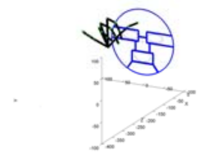
- Using pre-segmented face images, learn a model of the distinct relevant areas of the face.
- Areas in our model: forehead (1), eyes (2-3), nose (4), mouth (5), jaw (6), and cheeks (7-8).



Ding & Martinez, 2008

Modeling Results

Forever

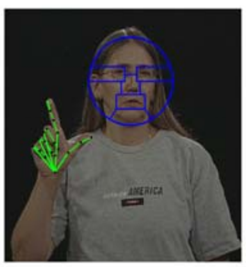


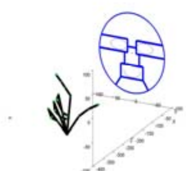
The closest hand point to the face defines the POA.

Ding & Martinez, 2008

Modeling Results

Laundry

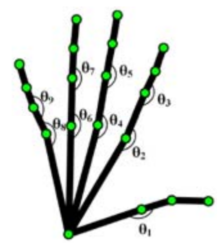




Ding & Martinez, 2008

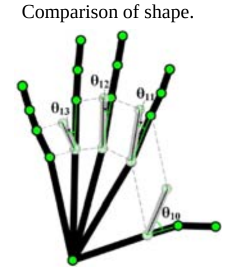
Recognition: Handshape

Comparison of angles.



$$E_{angle} = \|\Theta_1 - \Theta_2\|_2.$$

Comparison of shape.



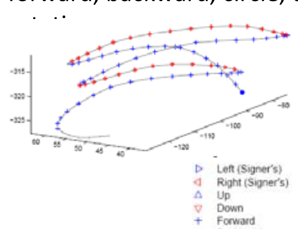
$$E_{shape} = \|\mathbf{T}_1\mathbf{P}_1 - \mathbf{P}_2\|_2 + \|\mathbf{P}_1 - \mathbf{T}_2\mathbf{P}_2\|_2$$

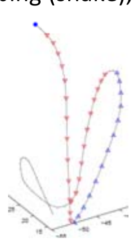
$$\min_{\mathbf{T}_i}\|\mathbf{T}_i\mathbf{P}_i - \mathbf{P}_{i+1}\|_2.$$

Ding & Martinez, 2008

Recognition: Motion

- We classify each segment in the recovered motion trajectories into one of the following classes: up, down, left, right, forward, backward, circle, tapping (shake),

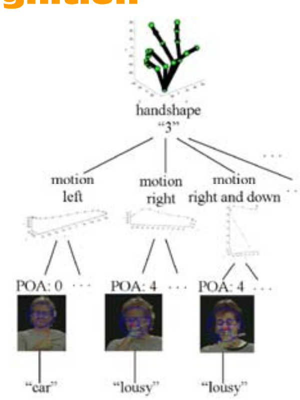




Ding & Martinez, 2008

Recognition

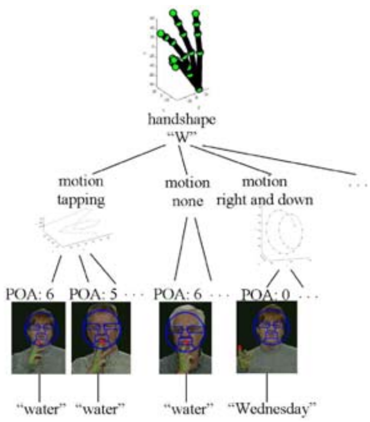
- Each handshape is combined with the motion class and the POA to describe concepts ("words").
- Each concept can be signed in multiple ways.
- We can train the system without the need to use sample video sequences.



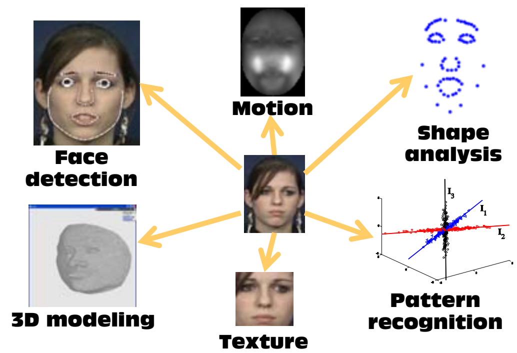
• Tested on 380 different seqs.

• 10 diff subjects.

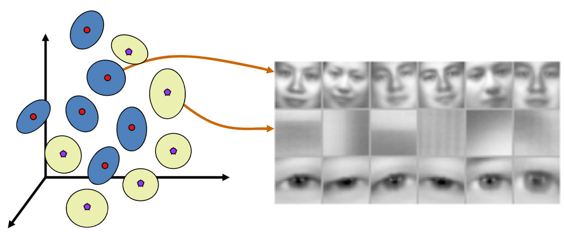
Recognition Rate: ~94%.



Face Modeling



Face and Facial Feature Detection



FRGC	False Negatives	False Positives	Detection Rate
Face Detector	133/16,028 (0.8%)	2/16,028 (~0%)	99%
Eye Detector	226/15,893 (1.7%)	0/15,893 (0%)	98%

Hamsici & Martinez, FRGC, 2005

SDA: Nonlinearly separable class pdf

Between-subclass scatter matrix:

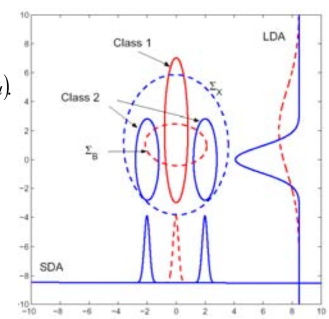
$$\Sigma_B = \sum_{i=1}^{C}\sum_{j=1}^{H_i} p_{ij}\left(\mu_{ij}-\mu\right)^T\left(\mu_{ij}-\mu\right).$$

Basis vectors:

$$\Sigma_B \mathbf{V} = \Sigma_X \mathbf{V}\Lambda.$$

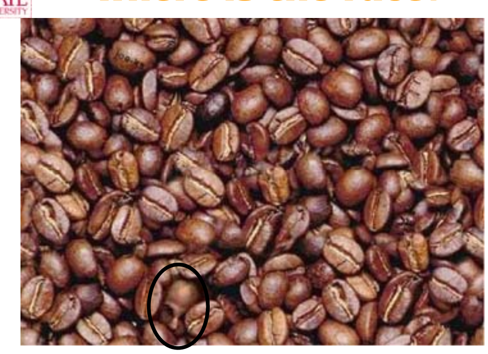
How many subclasses (H):

Minimize the conflict, *K*.



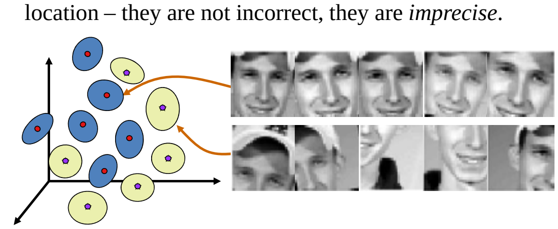
Zhu & Martinez, PAMI, 2006

Where is the face?



Features VS context

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

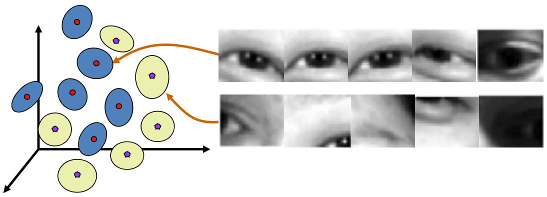


Key idea: Use context information to train where **not** to detect faces and facial features.

Ding & Martinez, CVPR, 2008

Features VS context

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

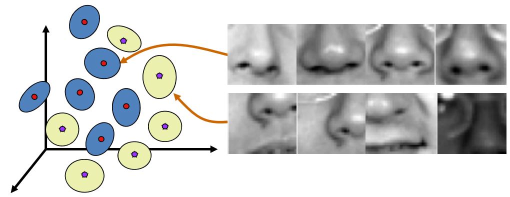


Key idea: Use context information to train where **not** to detect faces and facial features.

Ding & Martinez, CVPR, 2008

Features VS context

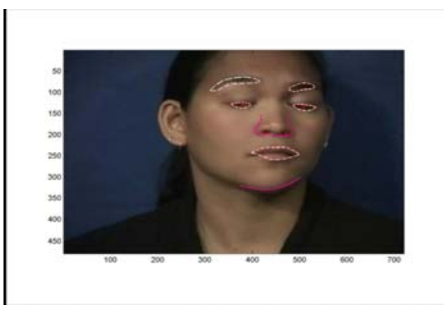
Observation: Most detections are near the correct location – they are not incorrect, they are *imprecise*.



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

Ding & Martinez, CVPR, 2008

Precise Detailed Detection



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

Ding & Martinez, CVPR, 2008

Detection in Stills

Ding & Martinez, CVPR, 2008

Shape Description

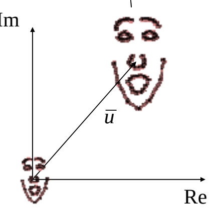
- We now have an outline of the 2D shape of the major internal features of the face.
- We need to find a mechanism to represent the face and recover the 3D information – i.e. the 3D shape.
- We need a representation that is invariant to translation, rotation and scale.
- Structure from motion
- Complex symmetry.

Spherical Distributions

$$u = [x_1 + iy_1, x_2 + iy_2, \dots, x_p + iy_p]^T \in C^p,$$

$$u_N = u - \bar{u} \in C^{p-1}$$

$$z = \frac{u_N}{\|u_N\|} \in CS^{p-2}$$



Complex Bingham

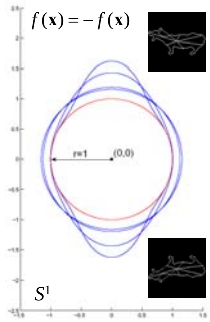
Hamsici & Martinez, Journal of Machine Learning Research, 2007.

Shape Description

Complex Bingham pdf represent ellipsoidally (and antipodally) symmetric distributions. Its pdf is:

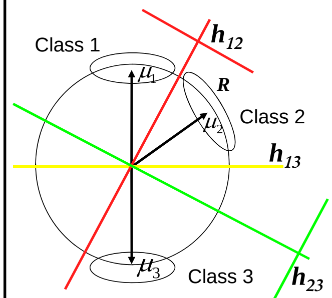
$f(\mathbf{z} \mid \mathbf{A}) = c_B(\mathbf{A}) \exp\{\mathbf{z}^* \mathbf{A}\mathbf{z}\}$,

where **A** is a matrix defining the variances in S^{p-1}. The eigenvectors of **A** can be easily estimated from samples, *but the corresponding eigenvalues cannot*.



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.

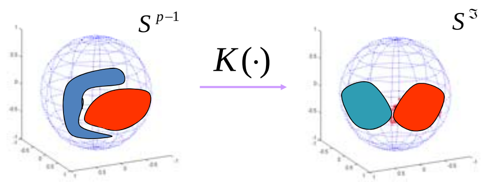
Theorems

- Two complex Bingham, $B_1(\boldsymbol{A})$ & $B_2(\boldsymbol{R}^T\boldsymbol{A}\boldsymbol{R})$, are spherical-homoscedastic if $\boldsymbol{R}$ in $SO(p)$ defines a planar rotation in the subspace spanned by any two eigenvectors of $\boldsymbol{A}$.
- If we estimate two spherical-homoscedastic complex Bingham distributions using two Gaussian pdf, then the Bayes classifier obtained with these Gaussians is the same as the Bayes classifier calculated on the original spherical pdf.

Hamsici & Martinez, JMLR & ICCV, 2007.

Kernel Spherical-Homoscedastic Shapes

- We can now 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 algorithms

Hamsici & Martinez, ICCV, 2007.

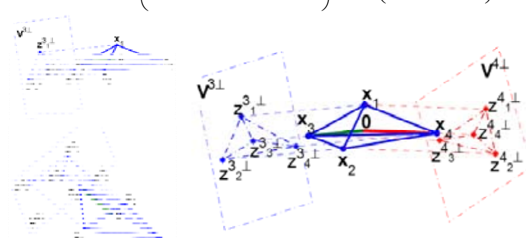
SH Shapes

- For shape representation, analysis, and classification, we can represent faces as variations from a mean face.
- The mean face is represented by the first eigenvector.
- Consecutive eigenvectors represent facial changes.

Hamsici & Martinez, ICCV, 2007.

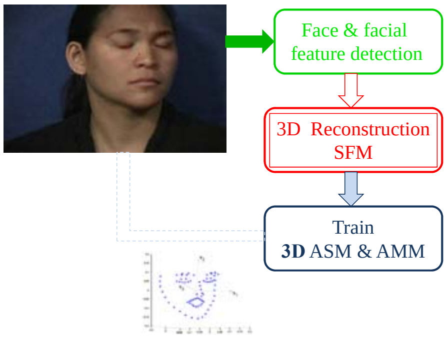
Rotation Invariant Kernels

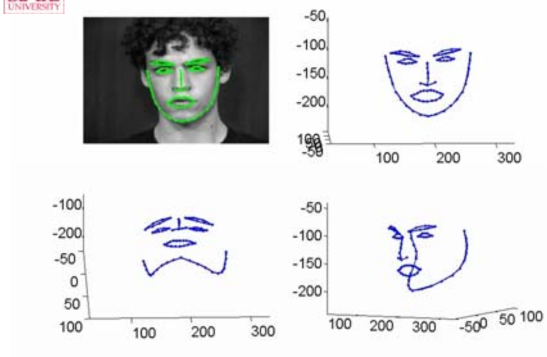
$$k(z_j, z_k) = \exp\left(-\frac{\left\|z_j - z_k \exp(i\theta_{z_j z_k})\right\|^2}{2\sigma^2}\right) = \exp\left(-\frac{2 - 2\left\|z_j^* z_k\right\|}{2\sigma^2}\right)$$

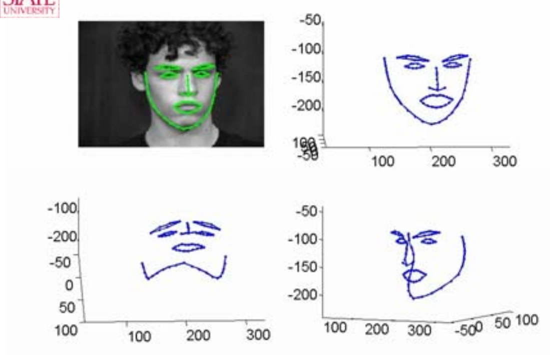


Hamsici & Martinez, 2008.

Face Modeling







Some References

- L. Ding and A.M. Martinez, "**Precise Detailed Detection of Faces and Facial Features,**" IEEE Computer Vision and Pattern Recognition (CVPR), 2008.
- H. Jia and A.M. Martinez, "**Low-Rank Matrix Fitting Based on Subspace Perturbation Analysis with Applications to Structure from Motion,**" IEEE Transactions on Pattern Analysis and Machine Intelligence, *accepted.*
- O.C. Hamsici and A.M. Martinez, "**Spherical-Homoscedastic Distributions: The equivalency of spherical and Normal distributions in classification,**" Journal of Machine Learning Research, 8(Jul):1583-1623, 2007.
- M. Zhu and A.M. Martinez, "**Subclass Discriminant Analysis,**" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 8, pp. 1274-1286, 2006.
- O.C. Hamsici and A.M. Martinez, "**Spherical-homoscedastic Shapes,**" IEEE International Conference on Computer Vision (ICCV), 2007.
- L. Ding and A.M. Martinez, "**Recovering the Linguistic Components of the Manual Signs in American Sign Language,**" IEEE Conference on Advanced Video and Signal-based Surveillance (AVSS), 2007.
- A.M. Martinez, "**Matching Expression Variant Faces,**" Vision Research, Vol. 43, Issue 9, pp. 1047-1060, 2003.

Thank you

Joint work:

- Liya Ding.
- Onur Hamsici.
- Manli Zhu.
- Hongjun Jia.
- Don Neth.

Acknowledgement: National Institutes of Health
National Science Foundation
Center for Cognitive Science at OSU