Facial Expression Recognition using a Dynamic Model and Motion Energy

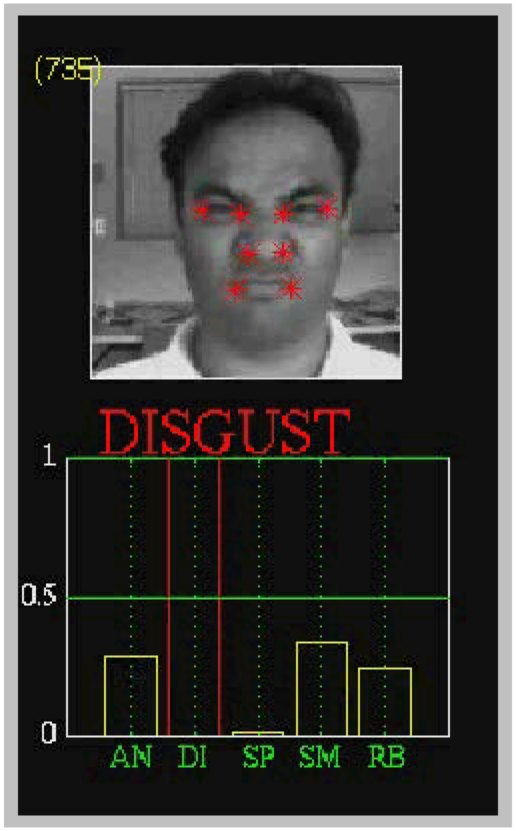
Irfan Essa,

Alex Pentland

(a review by Paul Fitzpatrick for 6.892)

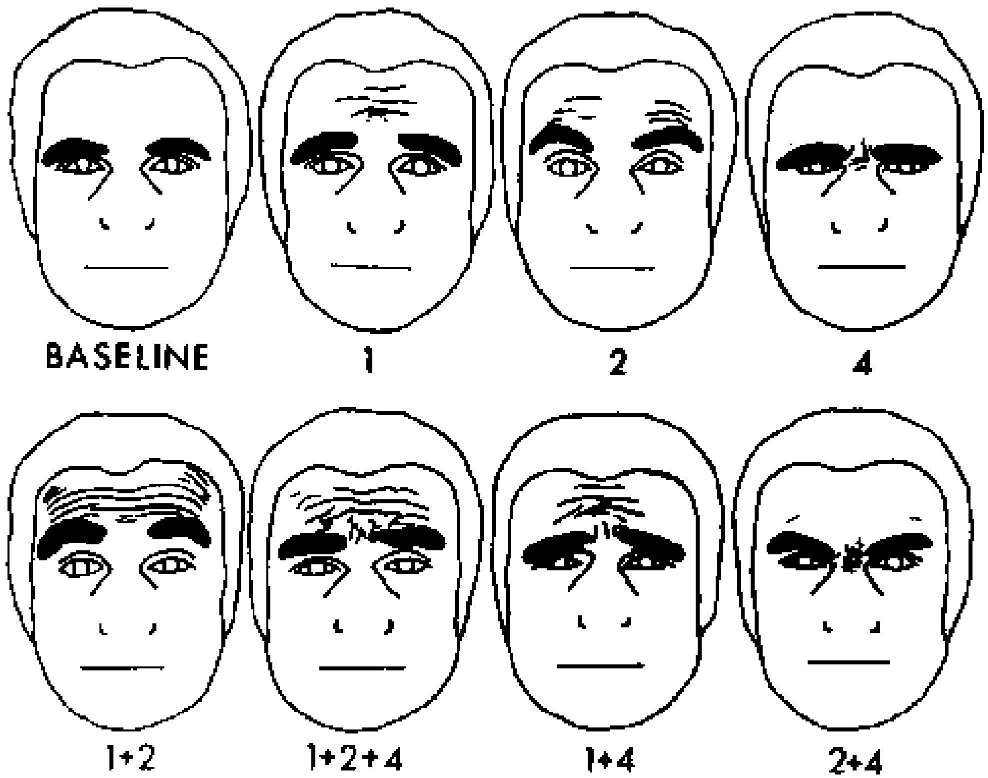
Overview

- Want to categorize facial motion
- Existing coding schemes not suitable
  - Oriented towards static expressions
  - Designed for human use
- Build better coding scheme
  - More detailed, sensitive to dynamics
- Categorize using templates constructed from examples of expression changes
  - Facial muscle actuation templates
  - Motion energy templates



Motivation

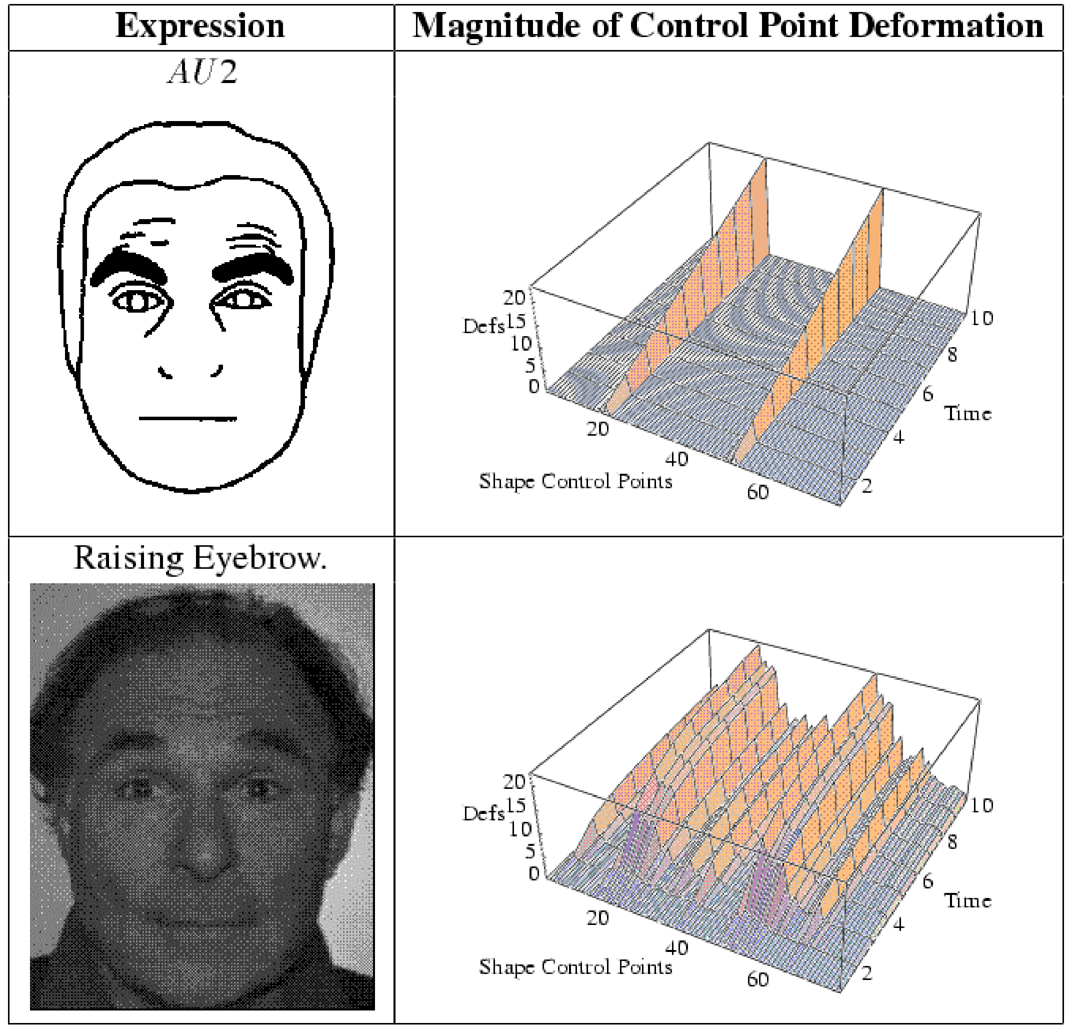
Facial Action Coding System



- FACS allows psychologists code expression from static facial "mug-shots"
- Facial configuration = combination of "action units"

Motivation
Problems with action units

- Spatially localized
  - Real expressions are rarely local
- Poor time coding
  - Either no temporal coding, or heuristic
  - Co-articulation effects not represented



Motivation

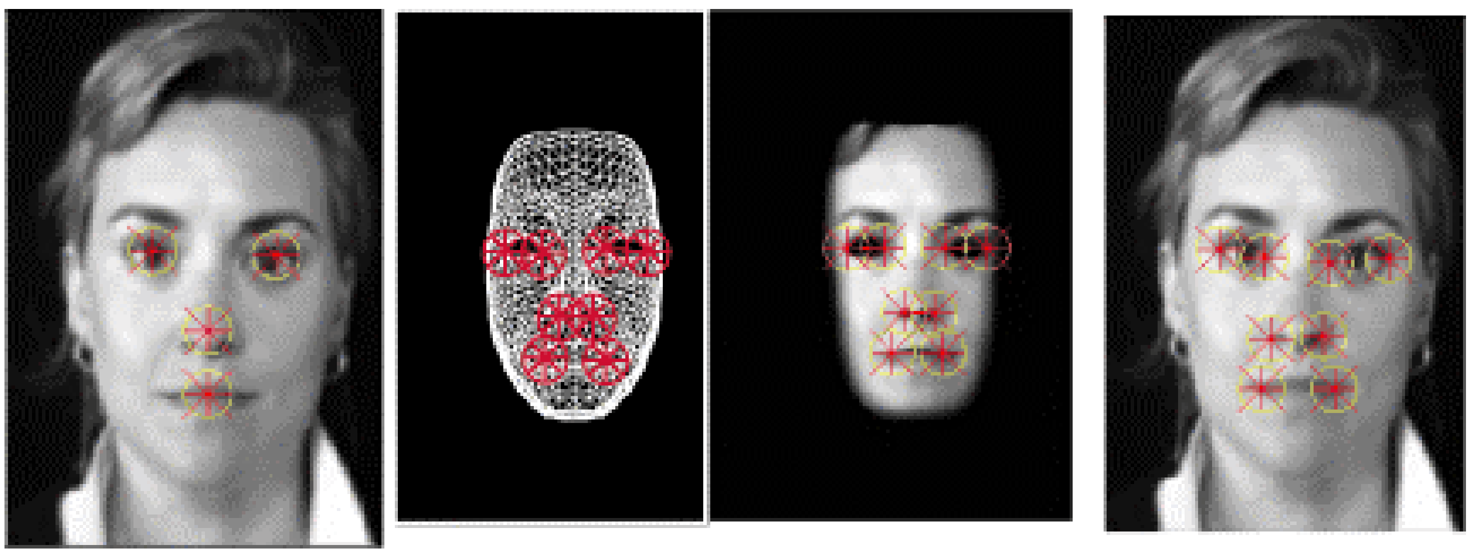
Solution: add detail

- Represent time course of all muscle activations during expression
- For recognition, match against templates derived from example activation histories
- To estimate muscle activation:
  - Register image of face with canonical mesh
  - Through mesh, locate muscle attachments on face
  - Estimate muscle activation from optic flow
  - Apply muscle activation to face model to generate "corrected" motion field, also used for recognition

Modeling

Registering image with mesh

- Find eyes, nose, mouth
- Warp on to generic face mesh
- Use mesh to pick out further features on face

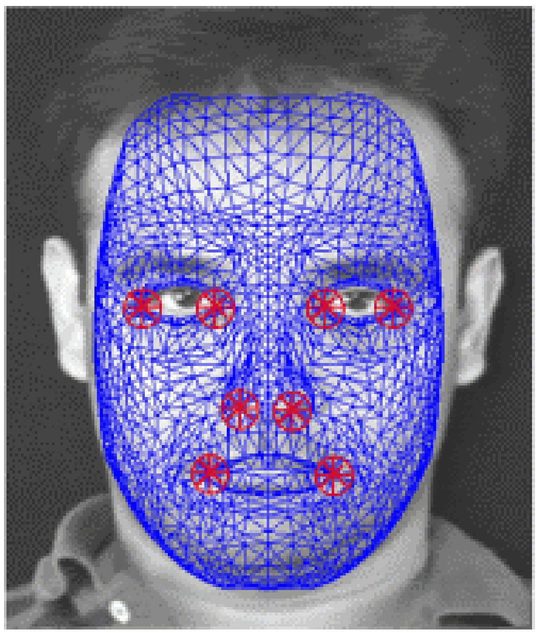


Eyes, Nose & Mouth Located (Turk *et. al* 91, Pentland & Moghaddam 94,95) | Facial Model | Warped to Generic Model | Mesh Points Extracted

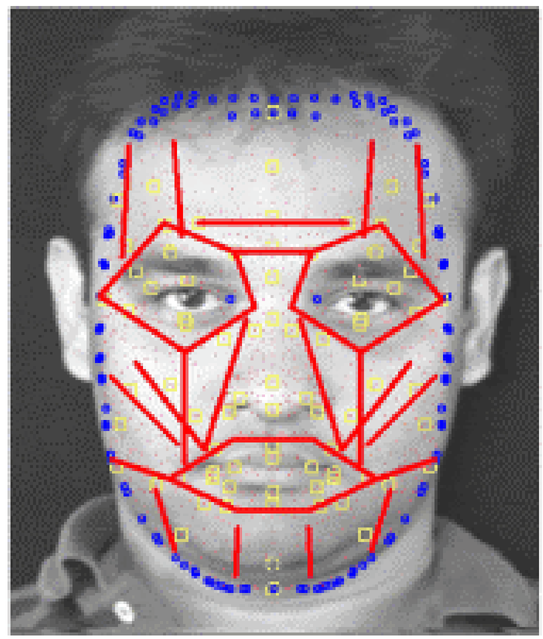
Modeling

Registering mesh with muscles

- Once face is registered with mesh, can relate to muscle attachments
- 36 muscles modeled; 80 face regions



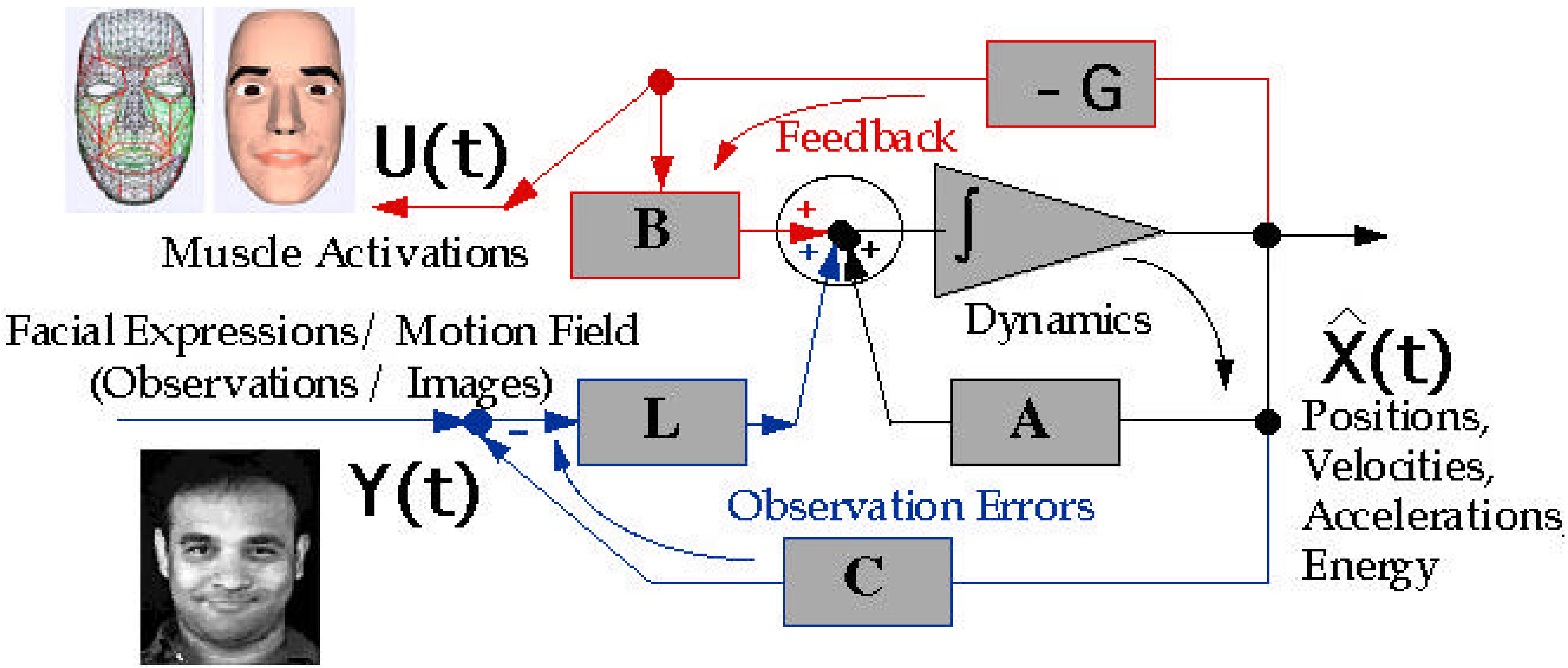
Mesh



Muscles

Modeling
Parameterize face motion

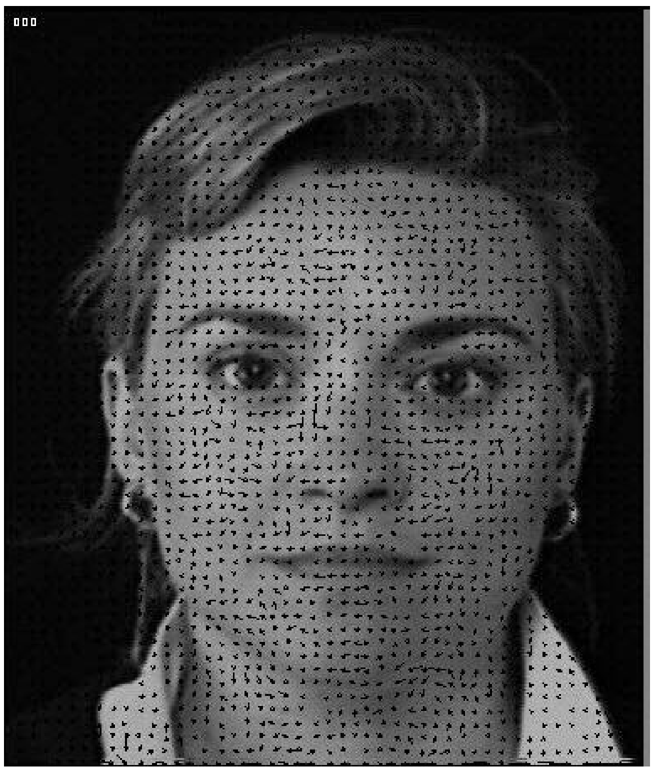
- Use continuous time Kalman filter to estimate:
  - Shape parameters: mesh positions, velocities, etc.
  - Control parameters: time course of muscle activation

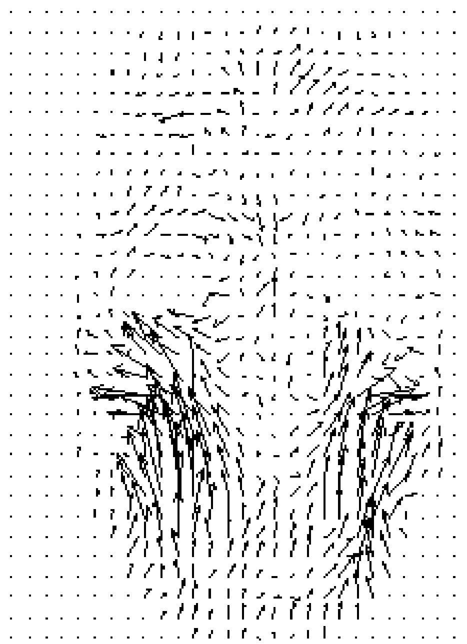


Modeling

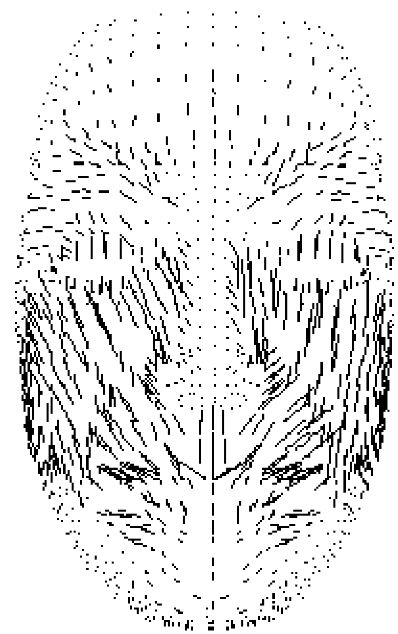
Driven by optic flow

- Computed using coarse to fine methods
- Use flow to estimate muscle actuation
- Then use muscle actuation to generate flow on model





Flow

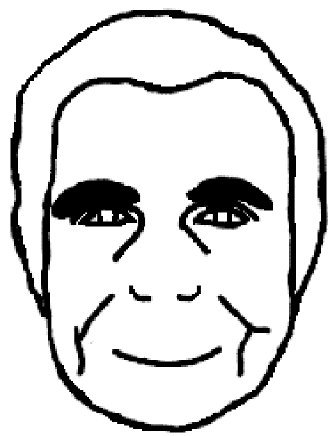
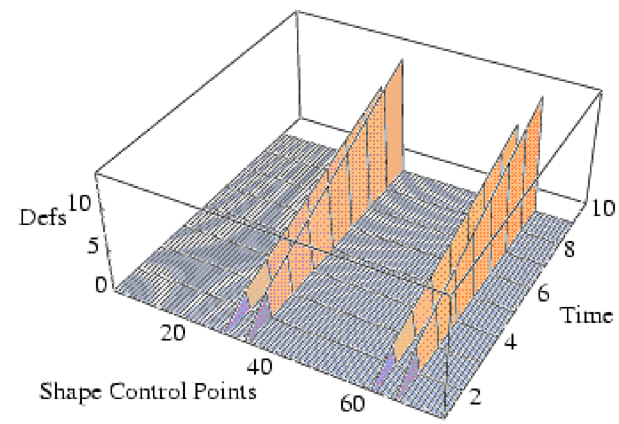
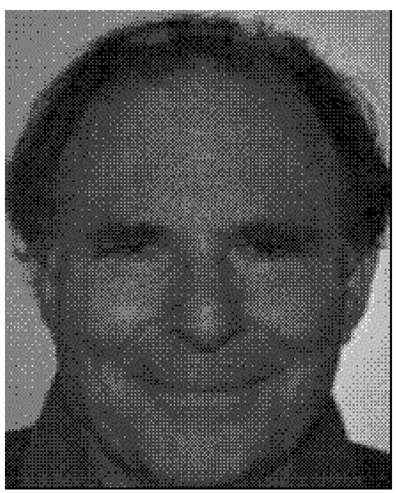
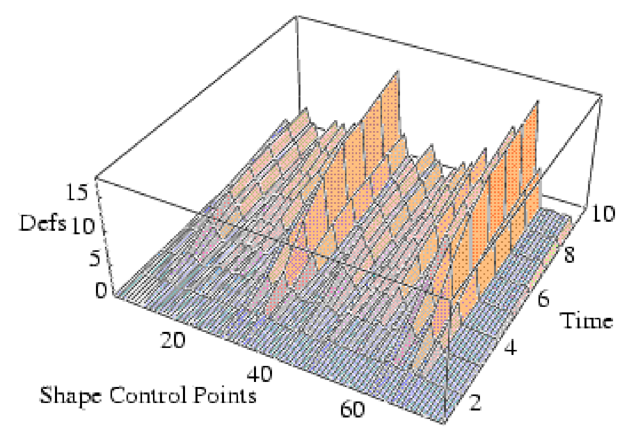


Motion on the Model

Analysis

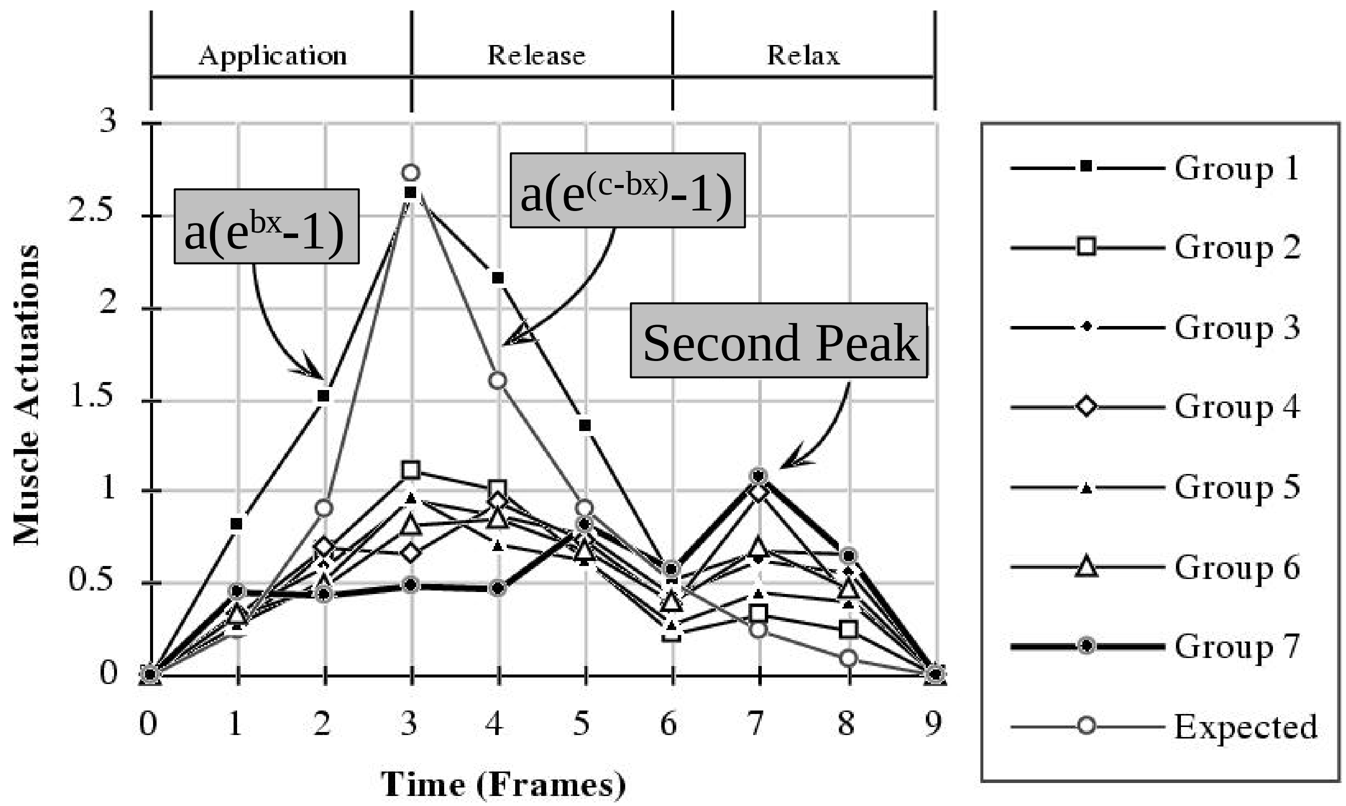
Spatial patterning

- Can capture simultaneous motion across the entire face
- Can represent the detailed time course of muscle activation
- Both are important for typical expressions

Expression	Magnitude of Control Point Deformation
AU 12 	
Smile. 	

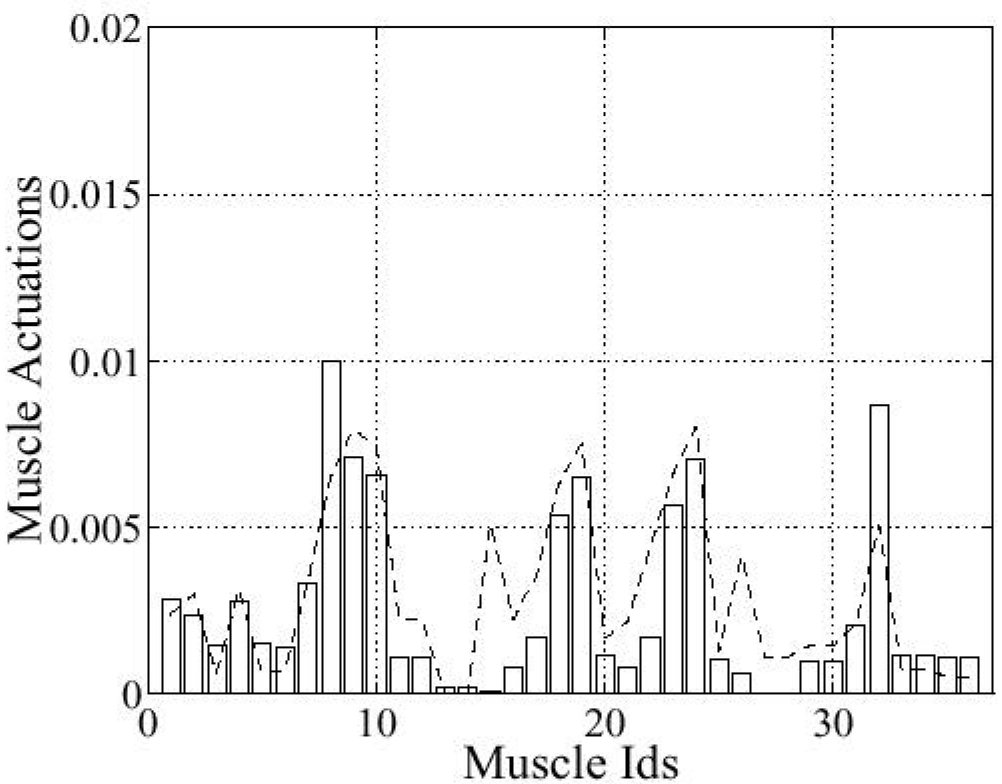
Analysis
Temporal patterning

- Application/release/relax structure – not a ramp
- Co-articulation effects present



Recognition
Peak muscle actuation templates

- Normalize time period of expression
- For each muscle, measure peak value over application and release
- Use result as template for recognition
  - Normalizes out time course, doesn't actually use it for recognition?

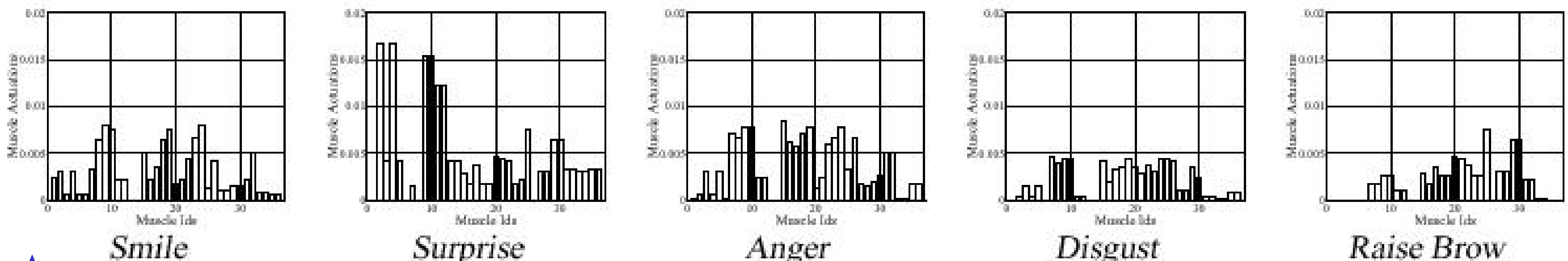


Peak muscle actuations during smile
(dotted line is template)

Recognition

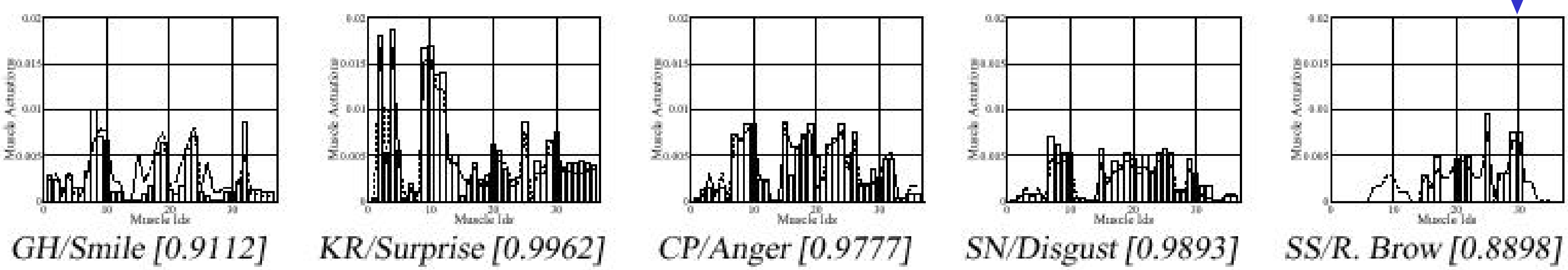
Peak muscle actuation templates

- Randomly pick two subjects making expression, combine to form template
- Match against template using normalized dot product



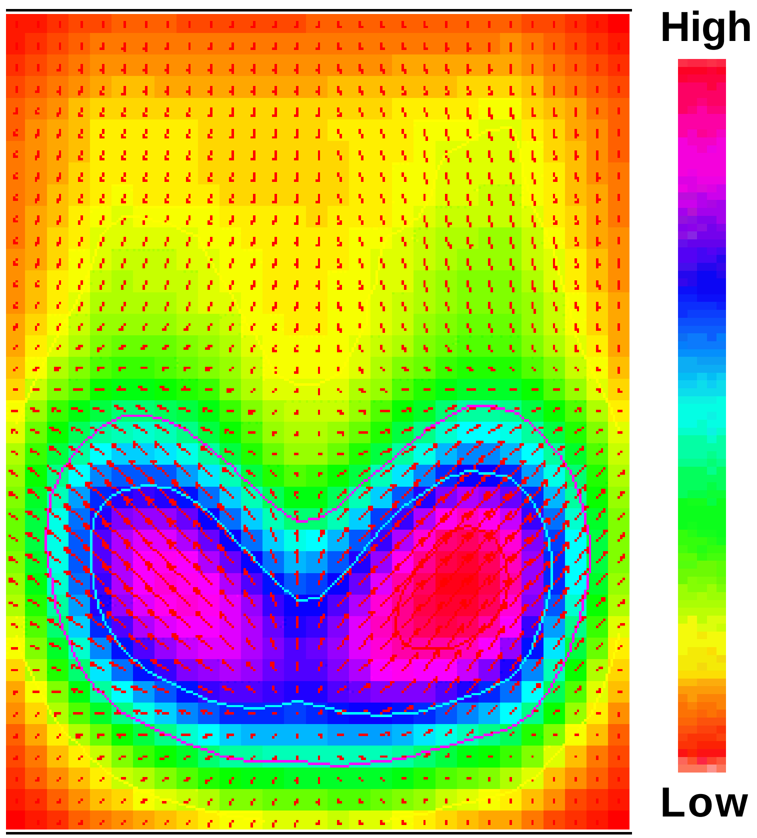
Templates

Peak muscle actuations for 5 subjects



Recognition
Motion energy templates

- Use motion field on face model, not on original image
- Build template representing how much movement there is at each location on the face
  - Again, summarizes over time course, rather than representing it in detail
  - But does represent some temporal properties

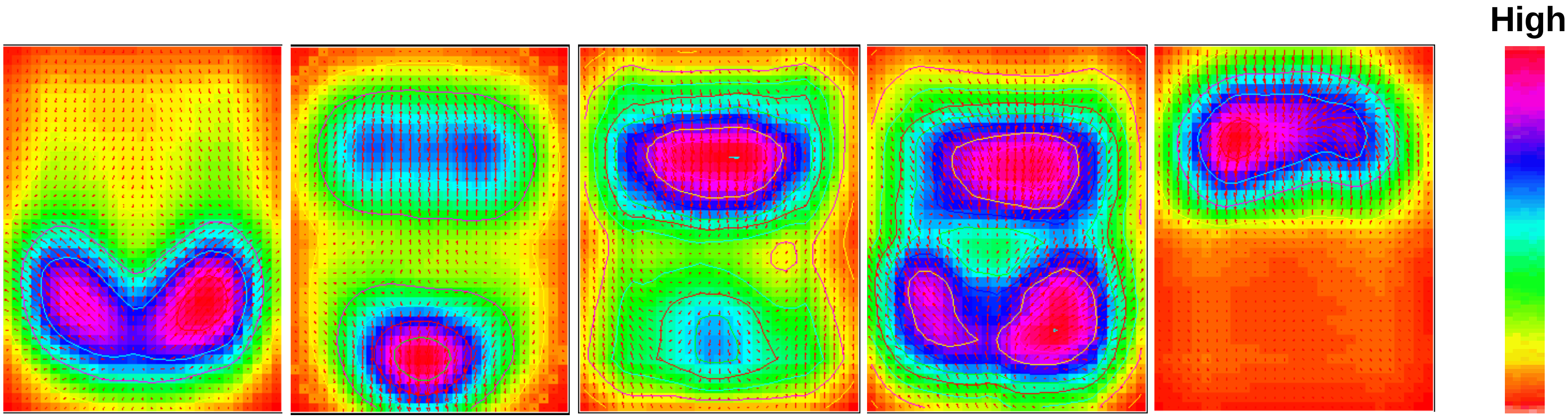


Motion energy template for smile

Recognition

Motion energy templates

- Randomly pick two subjects making expression, combine to form template
- Match against template using Euclidean distance



Smile *Surprise* *Anger* *Disgust* *Raise brow*

Data acquisition

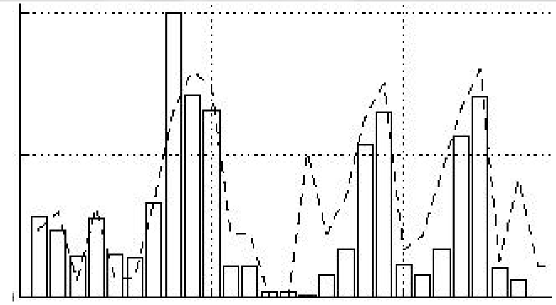
- Video sequences of 20 subjects making 5 expressions
  - smile, surprise, anger, disgust, raise brow
- Omitted hard-to-evoke expressions of sadness, fear
- Test set: 52 sequences across 8 subjects

Results

Data acquisition

Results
Using peak muscle actuation

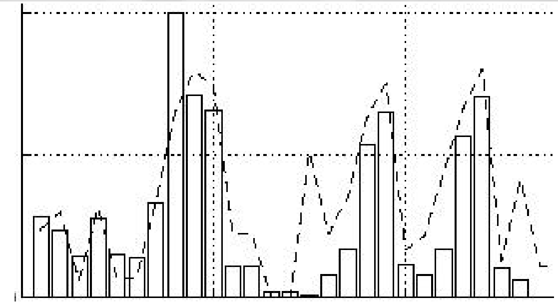
- Comparison of peak muscle actuation against templates across entire database
- 1.0 indicates complete similarity

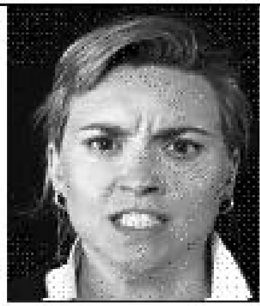
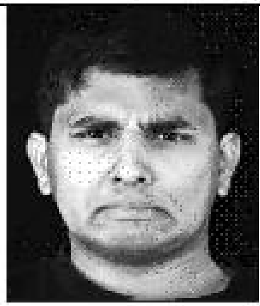
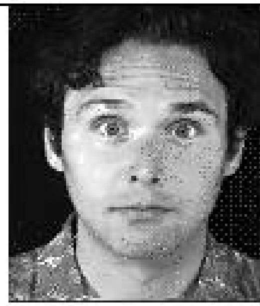


Expressions	Smile	Surprise	Anger	Disgust	Raise Brow
Template					
Smile	**0.97** $\pm$ **0.03**	0.63 $\pm$ 0.04	0.95 $\pm$ 0.01	0.86 $\pm$ 0.04	0.59 $\pm$ 0.16
Surprise	0.58 $\pm$ 0.03	**0.99** $\pm$ **0.01**	0.59 $\pm$ 0.04	0.57 $\pm$ 0.05	0.56 $\pm$ 0.09
Anger	0.90 $\pm$ 0.05	0.55 $\pm$ 0.05	**0.97** $\pm$ **0.02**	0.91 $\pm$ 0.01	0.65 $\pm$ 0.14
Disgust	0.82 $\pm$ 0.06	0.57 $\pm$ 0.05	0.92 $\pm$ 0.03	**0.95** $\pm$ **0.03**	0.78 $\pm$ 0.10
Raise Brow	0.58 $\pm$ 0.05	0.57 $\pm$ 0.07	0.70 $\pm$ 0.05	0.78 $\pm$ 0.06	**0.96** $\pm$ **0.04**

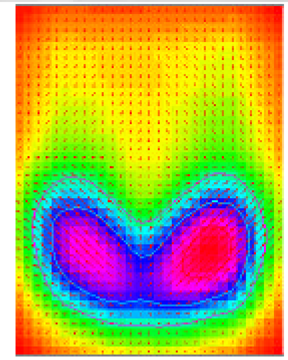
Results
Using peak muscle actuation

- Actual results for classification
- One misclassification over 51 sequences



Expressions	Smile	Surprise	Anger	Disgust	Raise Brow
Template					
Smile	**12**	0	1	0	0
Surprise	0	**10**	0	0	0
Anger	0	0	**9**	0	0
Disgust	0	0	0	**10**	0
Raise Brow	0	0	0	0	**10**
Success	100%	100%	90%	100%	100%

Results
Using motion energy templates

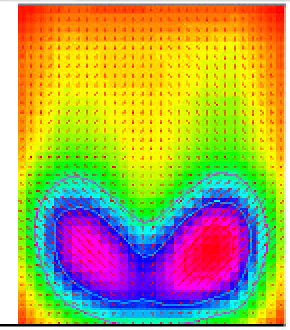


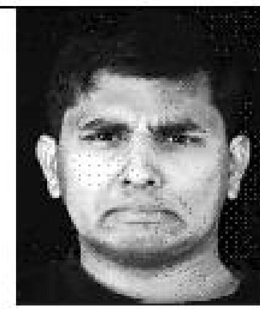
- Comparison of motion energy against templates across entire database
- Low scores indicate greater similarity

Expressions	Smile	Surprise	Anger	Disgust	Raise Brow
Template					
Smile	**94.1±34.7**	266.2 ± 52.3	$\boxed{234.5 \pm 62.7}$	153.7 ± 59.7	306.6 ± 15.3
Surprise	230.9 ± 8.7	**123.6±70.7**	160.5 ± 38.3	173.5 ± 14.2	233.4 ± 14.1
Anger	$\boxed{225.7 \pm 16.5}$	199.2 ± 76.0	**98.3±46.3**	$\boxed{160.1 \pm 29.1}$	147.0 ± 15.5
Disgust	149.0 ± 22.7	198.1 ± 54.0	$\boxed{140.3 \pm 43.7}$	**99.3±23.4**	224.3 ± 16.2
Raise Brow	339.9 ± 32.9	321.6 ± 96.4	208.9 ± 33.0	293.2 ± 26.8	**106.8±27.0**

Results
Using motion energy templates

- Actual results for classification
- One misclassification over 49 sequences



Expressions	Smile	Surprise	Anger	Disgust	Raise Brow
Template					
Smile	**12**	0	0	0	0
Surprise	0	**10**	0	0	0
Anger	0	0	**9**	0	0
Disgust	0	0	1	**10**	0
Raise Brow	0	0	0	0	**8**
Success	100%	100%	90%	100%	100%

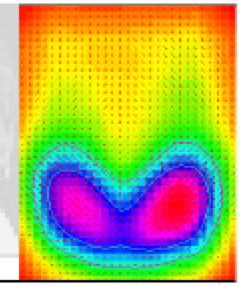
Comments

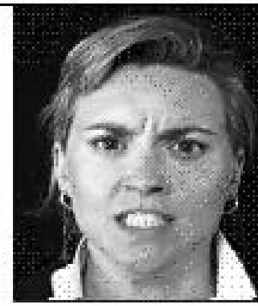
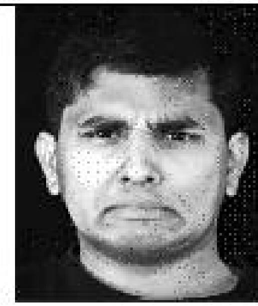
Small test set

- Test set is a little small to judge performance
- Simple simulation of the motion energy classifier using their tables of means and std. deviations shows:
  - Large variation in results for their sample size
  - Results are worse than test data would suggest
  - Example: anger classification for large sample size has accuracy of 67%, as opposed to 90%
- Simulation based on false Gaussian, uncorrelated assumption (and means, deviations derived from small data set!)

Comments

Naïve simulated results



Expressions	Smile	Surprise	Anger	Disgust	Raise Brow
Template					
Smile	90.7%	1.4%	2.0%	19.4%	0.0%
Surprise	0.0%	64.8%	9.0%	0.1%	0.0%
Anger	0.0%	18.2%	67.1%	3.8%	9.9%
Disgust	9.3%	13.1%	21.4%	76.7%	0.0%
Raise brow	0.0%	2.4%	0.5%	0.0%	90.1%

Overall success rate: 78% (versus 98%)

Comments
Motion estimation vs. categorization

- The authors' formulation allows detailed prior knowledge of the physics of the face to be brought to bear on motion estimation
- The categorization component of the paper seems a little primitive in comparison
- The template-matching the authors use is:
  - Sensitive to irrelevant variation (facial asymmetry, intensity of action)
  - Does not fully use the time course data they have been so careful to collect

Conclusion

Video, gratuitous image of Trevor