Computer Vision

Computer Science Tripos Part II

Dr Christopher Town

12. Face detection, face recognition, and facial interpretation.

UNIVERSITY OF CAMBRIDGE

Some challenges of face detection and recognition:

- Faces are surfaces on **3D objects** (heads) and appearance depends on **viewpoint**, **orientation**, and **illuminant**.
- Facial surfaces have **relief**, and so parts can **occlude** other parts and **shadows** and **shading** depends on this and on the **illuminant**
- Faces have **variable specularity** (dry skin may be Lambertian, oily or sweaty skin may be specular).
- Faces are somewhat elastic and deformable due to changes in **facial expression** (social computation)
- Faces can have **partial occlusions** (e.g. glasses, cosmetics, cigarettes, moustaches, eyebrows)
- Facial appearance changes with **age**, and genetically identical twins have very similar appearance (bad for biometrics)

Open questions:

- What is the best ***representation*** to use for faces? Should this be 3D or 2D?
- How can ***invariances*** to size (hence distance), location, pose, and angle of view be achieved?
- What are the ***generic*** (i.e. universal) properties of all faces useful for ***detection***?
- What are the ***particular*** features useful for ***recognition***?
- How can we handle the variability of facial appearance due to intrinsic and extrinsic causes?

Louis Léopold Boilly, **Reunion de Têtes Diverses**

Gore-Clinton or Clinton-Clinton?

P.Sinha

Dr Chris Town

Cheney-Bush or Bush-Bush?

P.Sinha

Dr Chris Town

Sinha *et al.*: Face Recognition by Humans: Nineteen Results Researchers Should Know About

Fig. 6. *Even drastic compressions of faces do not render them unrecognizable. Here, celebrity faces have been compressed to 25% of their original width. Yet, recognition performance with this set is the same as that obtained with the original faces.*

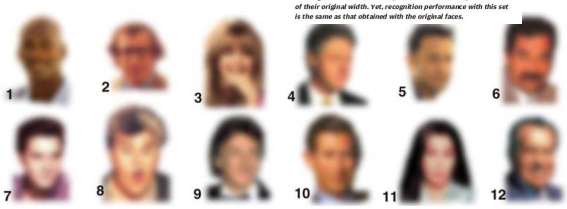


Fig. 1. *Unlike current machine-based systems, human observers are able to handle significant degradations in face images. For instance, subjects are able to recognize more than half of all familiar faces shown to them at the resolution depicted here. Individuals shown in order are: Michael Jordan, Woody Allen, Goldie Hawn, Bill Clinton, Tom Hanks, Saddam Hussein, Elvis Presley, Jay Leno, Dustin Hoffman, Prince Charles, Cher, and Richard Nixon.*

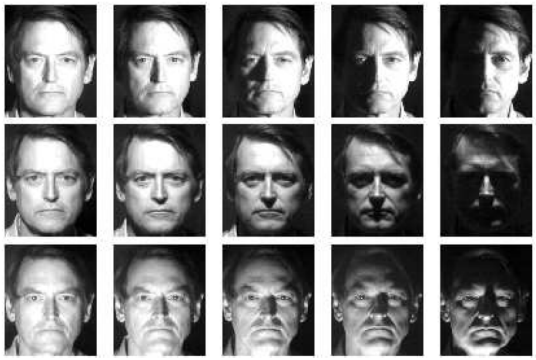
Dr Chris Town

Thompson, P. (1980). "Margaret Thatcher: a new illusion." *Perception* **9**:483-484

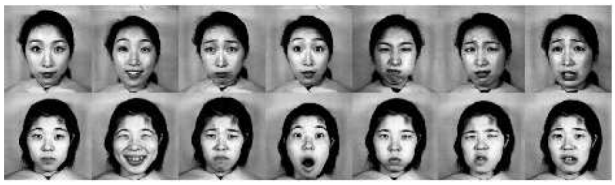
Dr Chris Town

Thompson, P. (1980). "Margaret Thatcher: a new illusion." *Perception* **9**:483-484

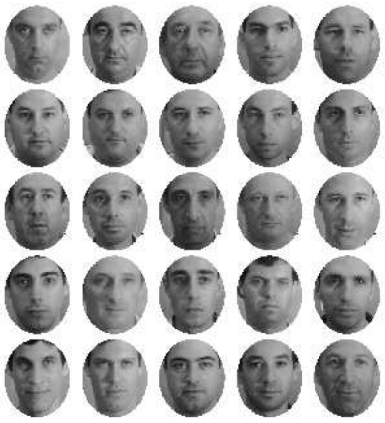
Dr Chris Town



Dr Chris Town



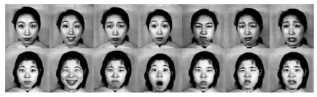
Dr Chris Town



Dr Chris Town

Dr Chris Town

Task	Within-Class Variability	Between-Class Variability
Face detection (classes: face / non-face)	bad	good
Face identification (classes: same/different faces)	bad	good
Facial expression interpretation (classes: same/different faces)	good	bad
Facial expression interpretation (classes: same/different expressions)	bad	good

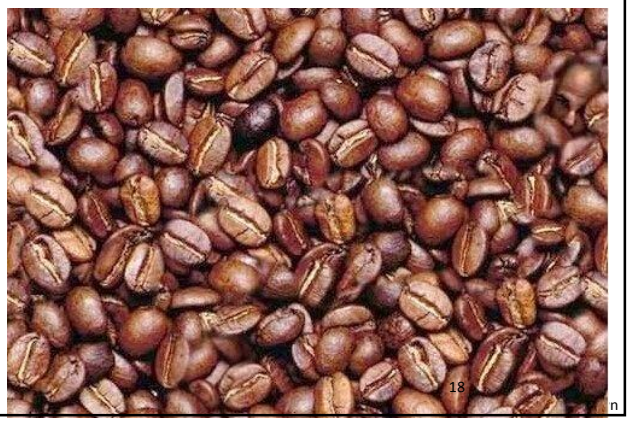


Dr Chris Town

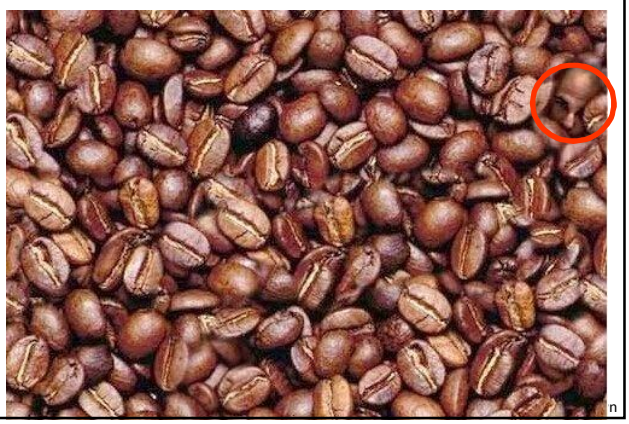
Can you find the faces in this image?

Dr Chris Town

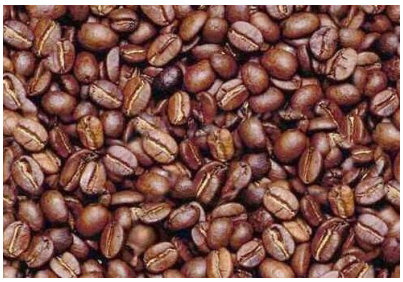
Can you find the faces in this image?



Can you find the faces in this image?



Face Detection



- The human visual system needs to apply serial attention to detect faces (context often helps to predict where to look)

Slide credit: David Lowe

Dr Chris Town

Sliding-Window Object Detection

- Template or classifier may need to be evaluated at many different positions and scales over the image.
- Starting with a detector size of 20x20 pixels and evaluating it at all possible offsets in a 400x400 image would require **(400-20+1)^2 = 145161** evaluations, just for a single scale of analysis!
- Clearly such a detector would have to be very efficient and have an extremely low false alarm rate to give reasonable performance.

Dr Chris Town

The Viola/Jones Face Detector

(2001)

- A widely used framework for object detection.
- Training is slow, but detection is very fast.

(some slides adapted from Paul Viola)

Dr Chris Town

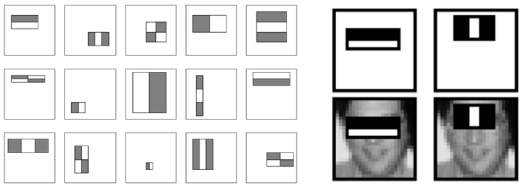
Classifier is Learned from Labeled Data

- Example Training Data
  - 5000 faces
  - 300 million non faces
    - 9400 non-face images
  - Faces are normalised
    - Scale, translation
- Many variations
  - Across individuals
  - Illumination
  - Pose (mainly in-plane orientation variation)

Dr Chris Town

Boosted Face Detection: Image Features

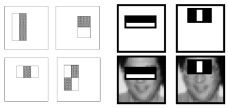
"Rectangle filters" similar to Haar



Dr Chris Town

Feature extraction

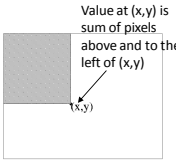
"Rectangular" filters



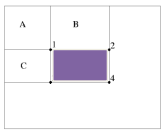
Feature output is difference between adjacent regions

Efficiently computable with integral image: any sum can be computed in constant time

Avoid scaling images → scale features directly for same cost



Integral image



$D = 1 + 4 - (2 + 3)$
$= A + (A + B + C + D) - (A + C + A + B)$
$= D$

Slide credit: K. Grauman

[Viola & Jones, CVPR 2001]

Dr Chris Town

AdaBoost

- Given a set of ***weak classifiers***

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j < p_j \theta_j \\ -1 & \text{otherwise} \end{cases}$$

that individually only have to be slightly better than random

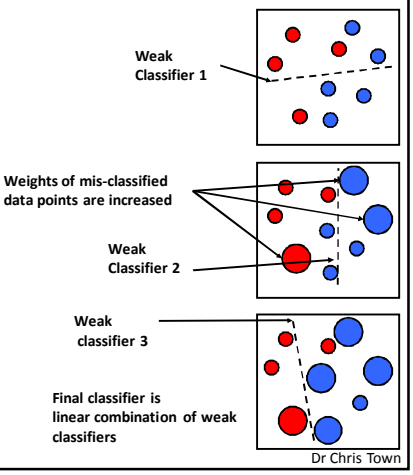
- We can iteratively combine classifiers to build a ***strong classifier***

$$h(x) = sign(\sum_j a_j h_j)$$

Dr Chris Town

AdaBoost

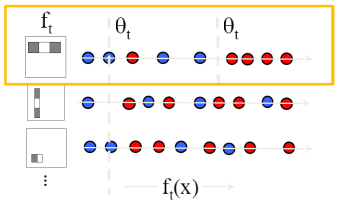
Freund & Shapire



Dr Chris Town

AdaBoost for Feature+Classifier Selection

- Want to select the single rectangle feature and threshold that best separates positive (faces) and negative (non-faces) training examples, in terms of *weighted* error.



Outputs of a possible rectangle feature on faces and non-faces.

Resulting weak classifier:

$$h_t(x) = \begin{cases} +1 & \text{if } f_t(x) > \theta_t \\ -1 & \text{otherwise} \end{cases}$$

For next round, reweight the examples according to errors, choose another filter/threshold combo.

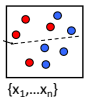
Slide credit: K. Grauman

Dr Chris Town

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \ldots, T$:
  1. Normalize the weights,
  $$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$
  so that w_t is a probability distribution.
  2. For each feature, j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t, $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
  3. Choose the classifier, h_t, with the lowest error ϵ_t.
  4. Update the weights:
  $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$
  where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is:
$$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$
where $\alpha_t = \log \frac{1}{\beta_t}$

AdaBoost Algorithm

Start with uniform weights on training examples



For T rounds

Evaluate *weighted* error for each feature, pick best.

Re-weight the examples:
incorrectly classified ⇒ more weight
Correctly classified ⇒ less weight

Final classifier is combination of the weak ones, weighted according to the error they had.

[Freund & Schapire 1995]

Dr Chris Town

Cascading Classifiers for fast detection

- For efficiency, apply less accurate but faster classifiers first to immediately discard windows that clearly appear to be negative; e.g.,
  - Build a chain of classifiers, choosing cheap ones with low false negative rates early in the chain

• Each stage is trained by adding features until the target detection and false positives rates are met (these rates are determined by testing the detector on a validation set).

• Stages are added until the overall target for false positive and detection rate is met.

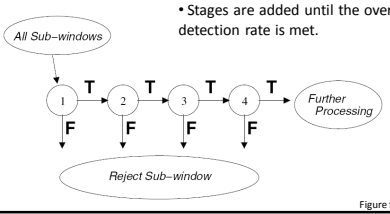


Figure from Viola & Jones CVPR 2001

Dr Chris Town

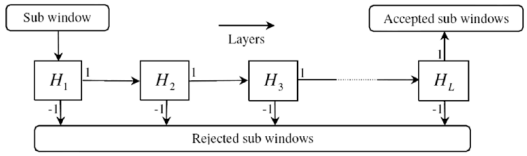
Overall detection rate D given detection rates d_i at each stage of the cascade:

$$D = \prod_{i=1}^{N} d_i$$

New image

Overall false positive rate F given false positive rates f_i at each stage of the cascade:

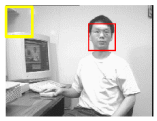
$$F = \prod_{i=1}^{N} f_i$$



Dr Chris Town

Overall detection rate D given detection rates d_i at each stage of the cascade:

$$D = \prod_{i=1}^{N} d_i$$



New image

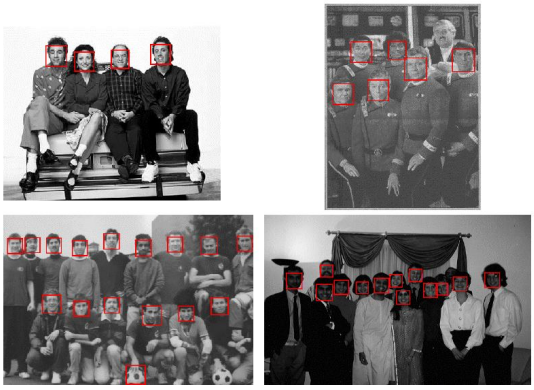
Overall false positive rate F given false positive rates f_i at each stage of the cascade:

$$F = \prod_{i=1}^{N} f_i$$

Since we may need to consider 10^5 sub-regions in a given image, we want F to be less than 10^{-5} in order to expect fewer than one false positive detection per image. To achieve $F = 10^{-5}$ for a 30 layer cascade, each f_i would have to be about 68% ($10^{-5/30} = 10^{-1/6}$), which looks rather easier than creating a single monolithic classifier with a false alarm rate below 0.00001! However, by the same argument, a decent detection rate of $D = 0.95$ would require d_i of $.95^{1/30}$ which is about 99.83%.

Dr Chris Town

Viola-Jones Face Detector: Results



Slide credit: Kristen Grauman

B. Leibe

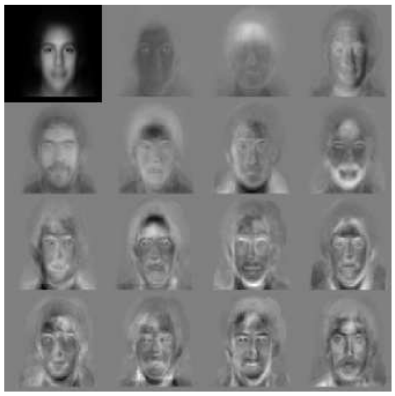
33

Dr Chris Town

Dr Chris Town

Dr Chris Town

Eigenfaces: PCA



Dr Chris Town

Principal Components Analysis (PCA)

also known as the *Karhunen-Loeve Transform*

$$\{x_i\}_{i=1}^{M} \quad x \in R^N \qquad M < N$$

Mean: $\mu = \frac{1}{M}\sum_{i=1}^{M} x_i$

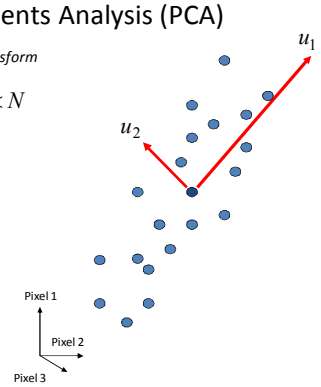
Sample covariance matrix:

$$C = \sum_{i=1}^{M}(x_i - \mu)(x_i - \mu)^T$$

Eigendecomposition:

$$C = ULU^T$$

where U represents the eigenvectors and L the eigenvalues



A. Torralba

Dr Chris Town

Computing eigenfaces by SVD

X =

svd(X,0) gives $X = U\,S\,V^T$

Covariance matrix

$XX^T = U\,S\,V^T\,V\,S\,U^T = U\,S^2\,U^T$

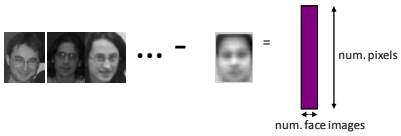
So the U's are the eigenvectors of the covariance matrix X

A. Torralba

Dr Chris Town

Computing eigenfaces by SVD

X =

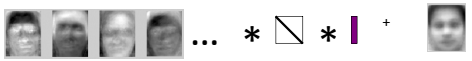


svd(X,0) gives $X = U\,S\,V^T$

Covariance matrix $XX^T = U\,S\,V^T\,V\,S\,U^T$
$= U\,S^2\,U^T$

Some new face image, x

x =



eigenfaces

S * v

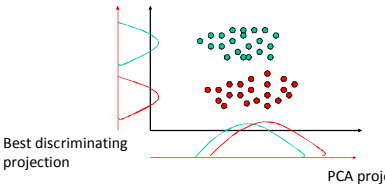
mean face

A. Torralba

Dr Chris Town

Projection vs discriminability

- PCA minimises projection error



- PCA is „unsupervised" no information on classes is used
- Discriminating information might be lost

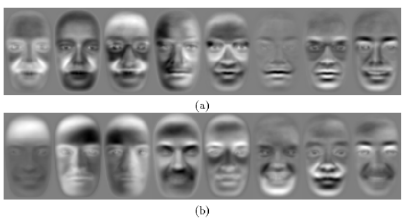
Slide credit: Ales Leonardis

Dr Chris Town

- Representation for a given face is very compact, Eigenbasis is pre-computed
- Eigenfaces are orthogonal basis vectors capturing the main modes of variation
- Set of faces from which Eigenbasis is computed can be adapted over time

Disadvantages:

- 2D appearance based, no invariances for pose or perspective
- Highly dependent on normalised scale and position
- Many Eigenfaces just capture variation in lighting, hairline, occlusions etc. rather than intrinsic appearance variation.



(a)

(b)

Figure 6: "Dual" Eigenfaces: (a) Intrapersonal, (b) Extrapersonal

Dr Chris Town

Wavelet representations of faces

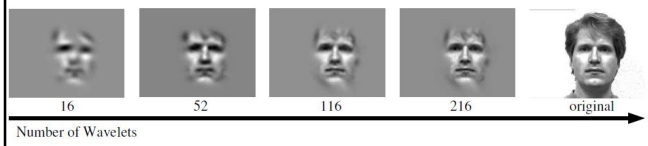
Illustration of Facial Feature Detection by Quadrature Filter Energy. Left panel: original image. Right panel (clockwise from top left): the real part after 2D Gabor wavelet convolution; the imaginary part; the modulus; and modulus superimposed on the original (faint) image, illustrating successful feature localisation by the **Quadrature Demodulator Network** in the previous Figure.

Dr Chris Town

Wavelet representations of faces

Because wavelets are ***localised***, they can track changes in facial expression in a local way.

This approach essentially treats a face as a kind of ***texture***, made up of various undulations in various positions, sizes, and orientations but without incorporating explicit models for the individual parts of faces.

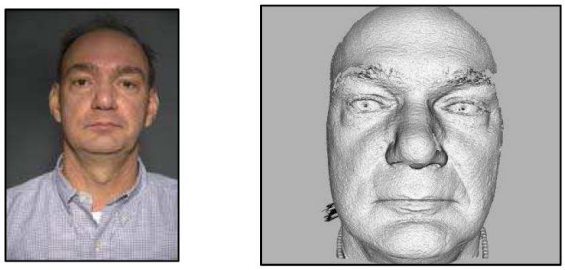


Dr Chris Town

3D Face Representation

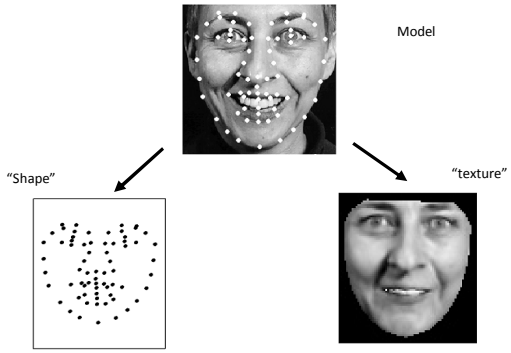
3D shape model: laser range-finding; stereo cameras; projection of structured light (grid patterns); or multi-view extrapolation

By projecting the texture (tone, colour, features, etc) onto the shape, one can generate models of the face in different poses.



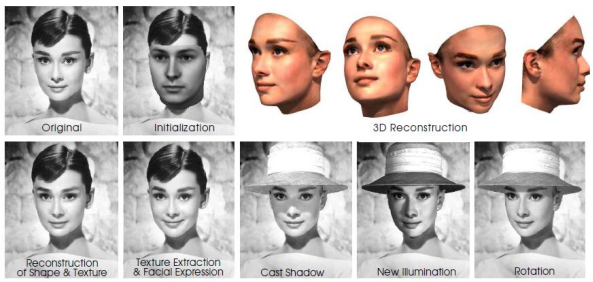
Dr Chris Town

A shape-texture face model



Slide: Dhruv Batra

Dr Chris Town

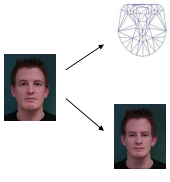


Movie2.avi

Dr Chris Town

The Morphable Face Model

•The actual structure of a face is captured in the shape vector $\mathbf{S} = (x_1, y_1, x_2, \ldots, y_n)^T$, containing the (x, y) coordinates of the n vertices of a face, and the appearance (texture) vector $\mathbf{T} = (R_1, G_1, B_1, R_2, \ldots, G_n, B_n)^T$, containing the color values of the mean-warped face image.



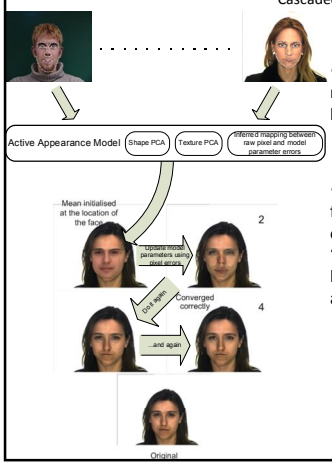
Shape S

Appearance T

Cootes, Edwards, and Taylor, "Active Appearance Models"

Dr Chris Town

Cascaded Classification Of Gender And Facial Expression Using Active Appearance Models



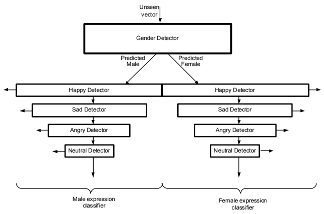
• ***The Data:*** Data used for building the model consisted of many images with 58 landmark points annotated.

• ***The Model:*** Parameters are derived from Principal Component Analysis on shape and texture vectors. The "Active" part of the model is the learned relationship between pixel and model errors.

Saatci, Y. and Town, C.P. **"Cascaded Classification of Gender and Facial Expression using Active Appearance Models"**, Proc. *International Conference on Automatic Face and Gesture Recognition*, 2006

Dr Chris Town

Architecture



Gender-specific Facial Expression Recognition

- Conclusion: Expression classification exhibits a degree of gender-dependency

Performance

Expression classification in females

	Happy	Sad	Angry	Neutral	Unclassified
Happy	97.1%	0%	0%	0%	2.94%
Sad	0%	80.7%	10.1%	5.32%	3.88%
Angry	0.98%	1.34%	77.7%	11.3%	8.73%
Neutral	0%	1.04%	3.44%	88.2%	7.32%

Expression classification in males

	Happy	Sad	Angry	Neutral	Unclassified
Happy	87.3%	0%	3.64%	7.28%	1.82%
Sad	0%	70.5%	16.1%	8.12%	5.40%
Angry	0.56%	5.02%	63.7%	13.3%	17.5%
Neutral	0%	2.30%	6.28%	73.3%	17.7%

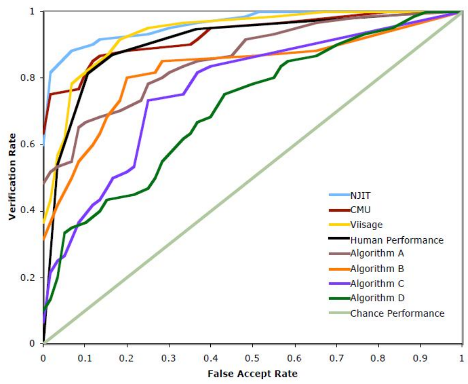
Expression classification (combined)

	Happy	Sad	Angry	Neutral	Unclassified
Happy	92.2%	0%	1.82%	3.64%	2.38%
Sad	0%	75.6%	13.1%	6.72%	4.64%
Angry	0.77%	3.18%	70.7%	12.3%	13.1%
Neutral	0%	1.67%	4.86%	81.0%	12.5%

Saatci, Y. and Town, C.P. **"Cascaded Classification of Gender and Facial Expression using Active Appearance Models"**, Proc. *International Conference on Automatic Face and Gesture Recognition*, 2006

Dr Chris Town

NIST 2007 Face Recognition Comparisons



Dr Chris Town

"There's more to vision than meets the eye"

Dr Chris Town