Computational pipeline for face analysis

Prof. Yosi Keller

Faculty of Engineering, Bar-Ilan University

Outline

1. Applications of face analysis
2. Why is face analysis (relatively) easy?
3. Face analysis pipeline
4. What are we doing?
5. Future applications

Applications of face analysis

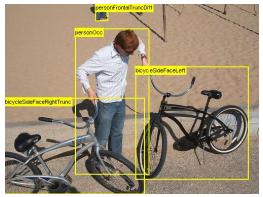
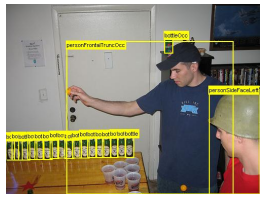
- Face recognition
- Gender estimation
- Ethnicity estimation
- Age estimation
- “Spacial” face similarity
  - Kin estimation
  - High/Low resolution face estimation
- Facial expressions

Why is face analysis (relatively) easy?

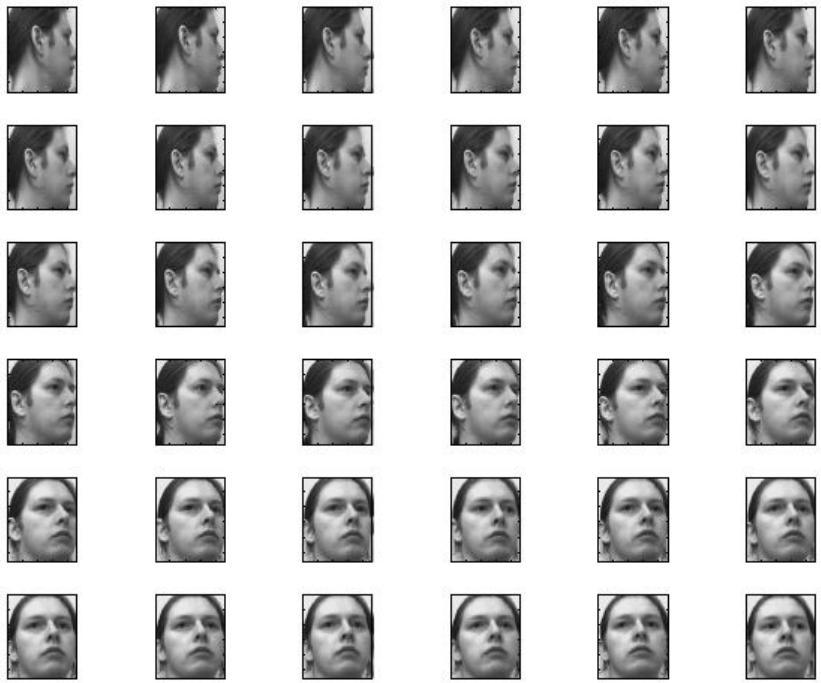
The most difficult attribute of visual recognition is the visual variability
Best detection results VOC 2007:

- Girshick et al. 2014: mAP of 53.3%
- Oxford VGG 16-layer network mAP of 66.0%

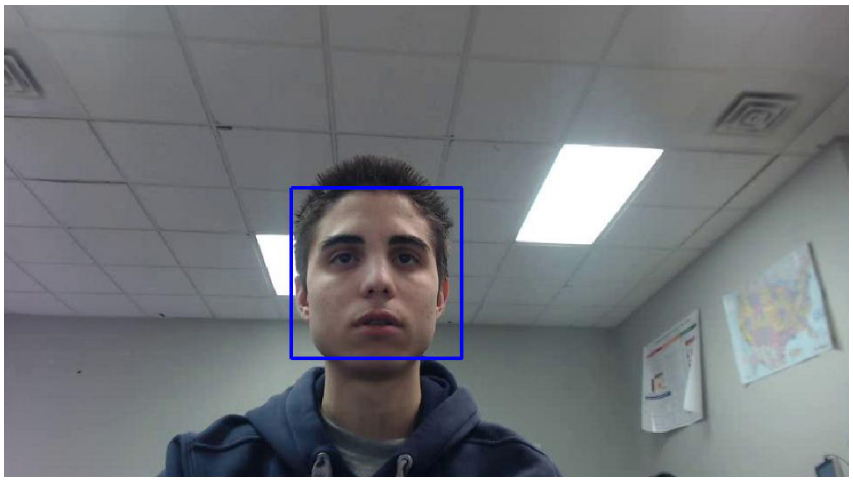


The variability of faces is mostly geometrical and parametric



Face detection



Classical work of Viola–Jones 2001:

- Fast and accurate
- Available in OpenCV and Matlab
- Can be improved using extra features
- Coarse estimate - “one size fits all”

Face alignment

”Supervised Descent Method and its Applications to Face Alignment”, Xiong and De la Torre, 2013
Very fast and accurate:

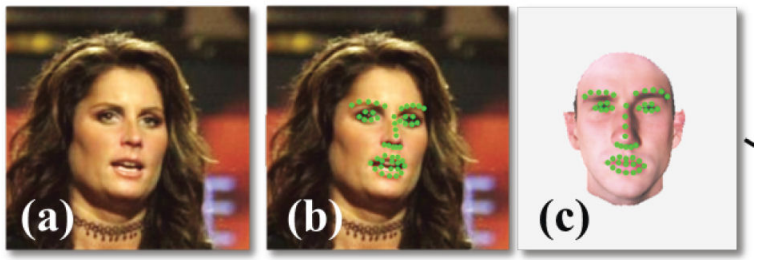
1. Compute (SIFT/Hog/LBP) descriptors at 49 points: $\mathbf{x}$
2. Regression: $\Delta = \mathbf{A} \cdot \mathbf{x}$

- The regression matrix is learnt.
- Runs at 60 FPS
- Extensions: regression tree and binary features
  - “Face Alignment at 3000 FPS via Regressing Local Binary Features”, Ren et al., 2014
  - “One Millisecond Face Alignment with an Ensemble of Regression Trees”,Kazemi et al., 2014

Our research: improve the convergence range
Show video

Pose estimation and 3D alignment

- Each point has a “name”
- Classical pose estimation: Perspective-n-Point problem (PnP)
  Posit



- Given the pose, the face image is rendered on the 3D model
- The 3D model can be rotated
- Missing parts added using symmetry
- Can be used for analysis

Face features

- Dense LBP, HOG, SIFT
- Face specific features: Weber Local Descriptor (WLD)
- Deep learning
  - Representation learning - requires a large training set
  - Task specific refinement

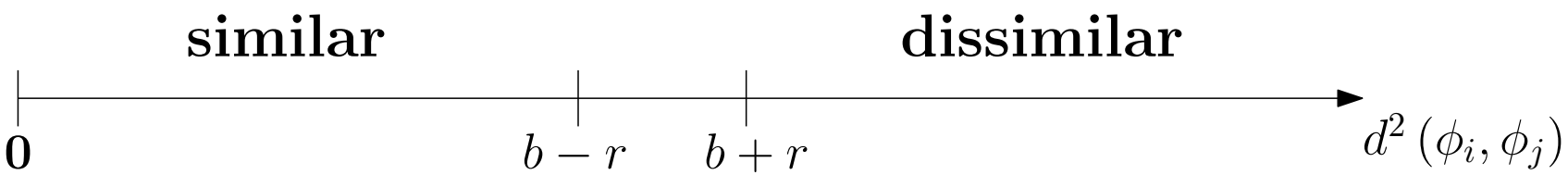
Distance learning I

Fisher Vector Faces in the Wild, Simonyan et al. 2013

- Can be applied directly
- The last layer in a deep network

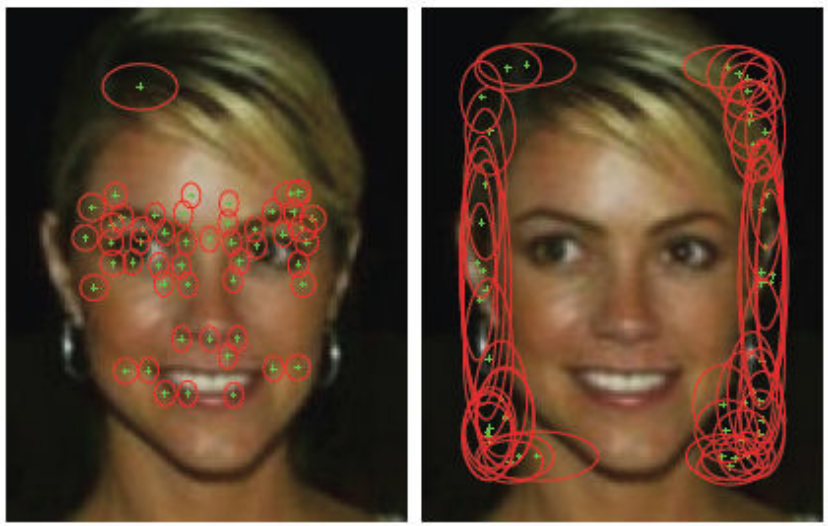
Learn a linear projection using margin maximization

$$d^2_{\mathbf{W}}\left(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j\right) = \left\|\mathbf{W}\boldsymbol{\phi}_i - \mathbf{W}\boldsymbol{\phi}_j\right\|_2^2$$



Distance learning II

Activations for recognition (from *Fisher Vector Faces in the Wild*)

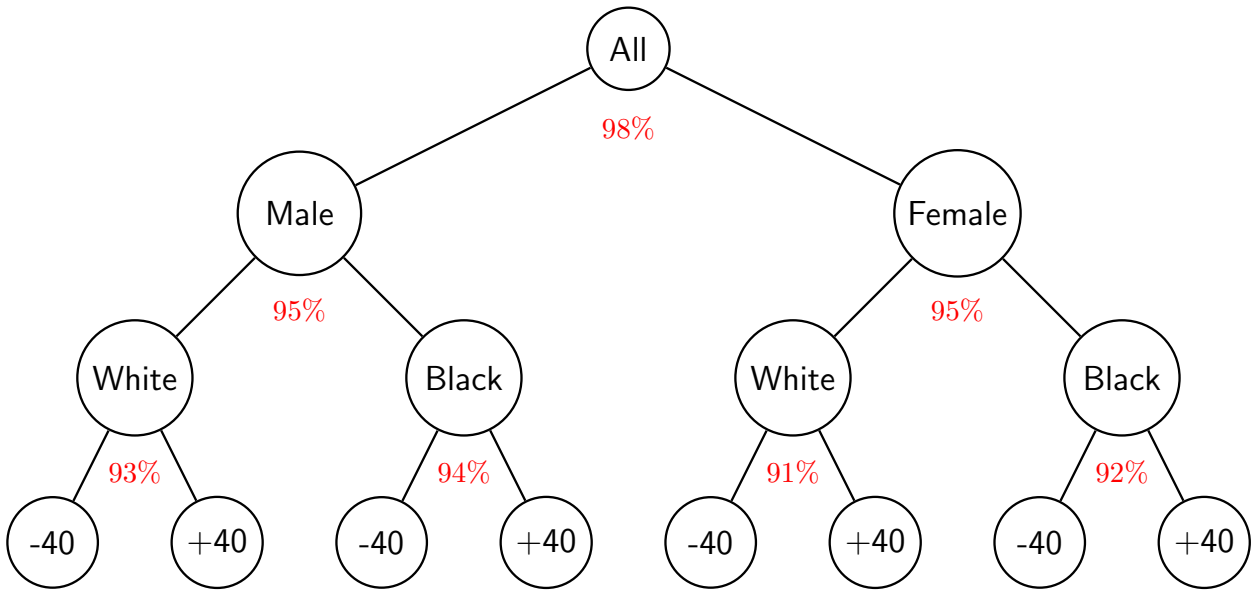


Application	**Same**	**Not same**	**Learning**	**Accuracy**
Verification	$name_i = name_j$	$name_i \neq name_j$	overfit	99%
Age estimation	$\lvert age_i - age_j \rvert \leq T$	$\lvert age_i - age_j \rvert > T$	generalization	3 years
Gender	$gender_i = gender_j$	$gender_i \neq gender_j$	generalization	98%
Ethnicity	$ethnicity_i = ethnicity_j$	$ethnicity_i \neq ethnicity_j$	generalization	98%

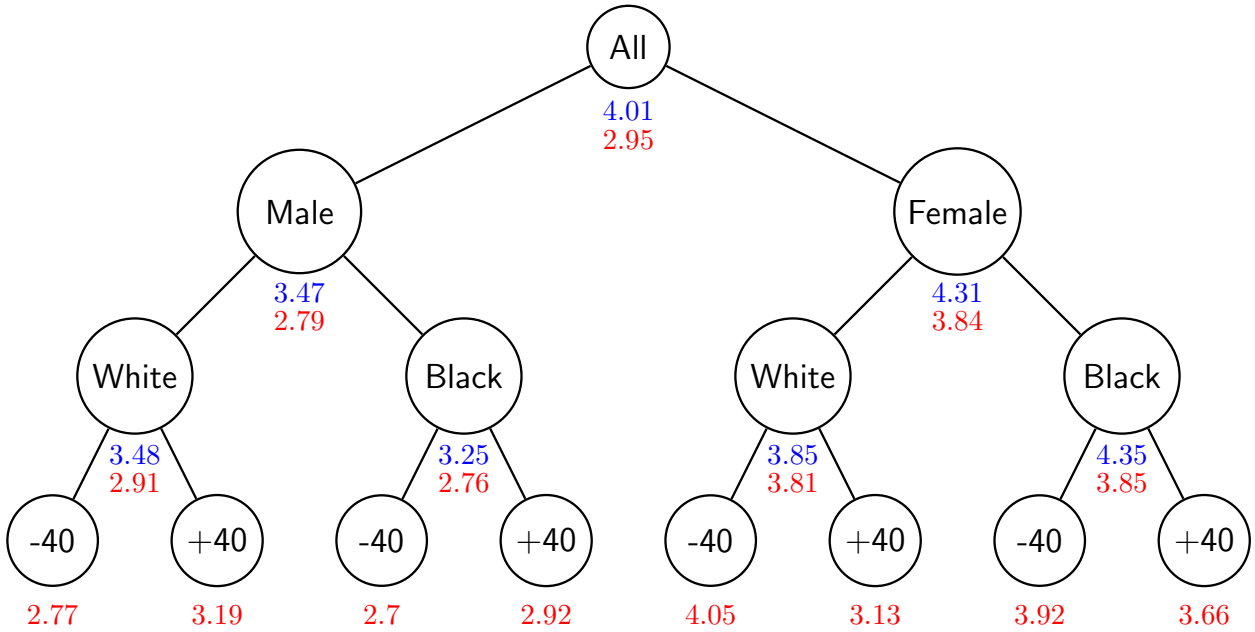
Face hierarchies I

- It is difficult to learn a classifier/regressor on a heterogeneous dataset
- Common solution - split data to folders and learn a set of classifiers/regressors
- Faces have a natural hierarchy based on: gender, ethnicity, age

Example of age estimation (Mor,Keller2014)



Face hierarchies II



Algorithm	MAE
KNN	6.26
PLS	4.56
KPLS (Guo 2011)	4.18

“Spacial” face similarity: non-identical twins I

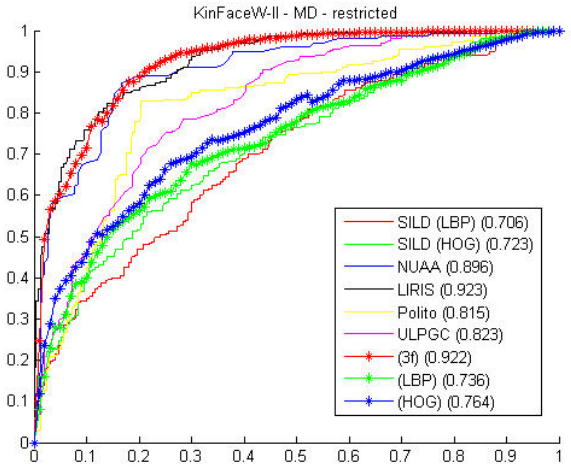
So far we considered

$$d^2_{\mathbf{W}}\left(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j\right) = \left\|\mathbf{W}\boldsymbol{\phi}_i - \mathbf{W}\boldsymbol{\phi}_j\right\|_2^2$$

But, there are other problems

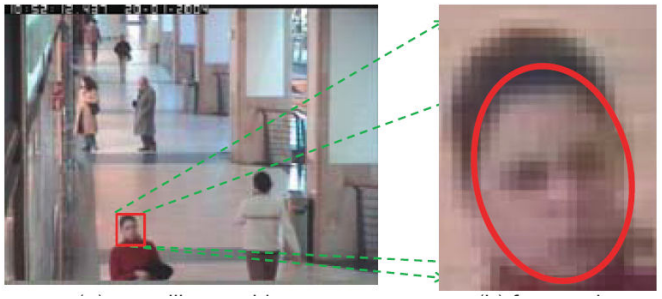
$$d^2_{\mathbf{W}}\left(\boldsymbol{\phi}_i, \boldsymbol{\phi}_j\right) = \left\|\mathbf{W}_o\boldsymbol{\phi}_i^o - \mathbf{W}_y\boldsymbol{\phi}_y^y\right\|$$

“Spacial” face similarity: non-identical twins II



Mahpod,Keller2015,FG2015

Similar problem



Yuen,Zou2012

But, we can use **multiple** low-resolution images

Future

- New technology - many potential applications
- Preprocessing infrastructure is important:alignment, frontalizations, features, etc.
- Some face analysis requires multiple images
- Temporal analysis: RNN, HMM
- Psychological analysis (“poker face”)