Recognizing faces - 1

Face Detection

Reference:
[Li-Jain-05] Stan Z. Li and Anil K. Jain, Handbook of Face Recognition, Springer, 2005
[Li-Jain-11] Stan Z. Li and Anil K. Jain, Handbook of Face Recognition, Second Edition, Springer, 2011

[Gonzalez-Woods-08] Rafael C. Gonzalez and Richard E. Woods, Digital Image Processing, Third Edition, Pearson Education, 2008.

[Viola-Jones-04] Paul Viola and Michael J. Jones, Robust Real-Time Face Detection, International Journal of Computer Vision, 57(2), 137-154, 2004

[Rowley-Baluja-Kanade-98] Henry A. Rowley, Shumeet Baluja and Takeo Kanade, Neural Network-Based Face Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(1), 23-38, 1998

[Rowley-Baluja-Kanade-98cvpr] Henry A. Rowley, Shumeet Baluja and Takeo Kanade, Rotation Invariant Neural Network-based Face Detection, Computer Vision and Pattern Recognition, pages 38-44, 1998

[Hsu-Mottaleb-Jain-02] Rein-Lien Hsu, Mohamed Abdel-Mottaleb and Anil K. Jain, Face Detection in Color Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(5), 696-706, 2002

[AZ-AV-09] Lecture notes written by Andrew Zisserman and Andrea Vedaldi at the University of Oxford for a computer vision course.

[Yale] The Yale Face Database B: http://cvc.yale.edu/projects/yalefacesB/yalefacesB.html

More related useful information can be found at the Face Recognition Homepage: http://www.face-rec.org/general-info/

Introduction

- Face detection has many applications.
- Face detection is the first step in automated face recognition.
- Its reliability has a major influence on the performance and usability of the entire face recognition system.
- Face detection can be viewed as classifying the pattern in the sub-window as either face or nonface.

Is it a face or nonface?

[Yale]

Is it a face or nonface?

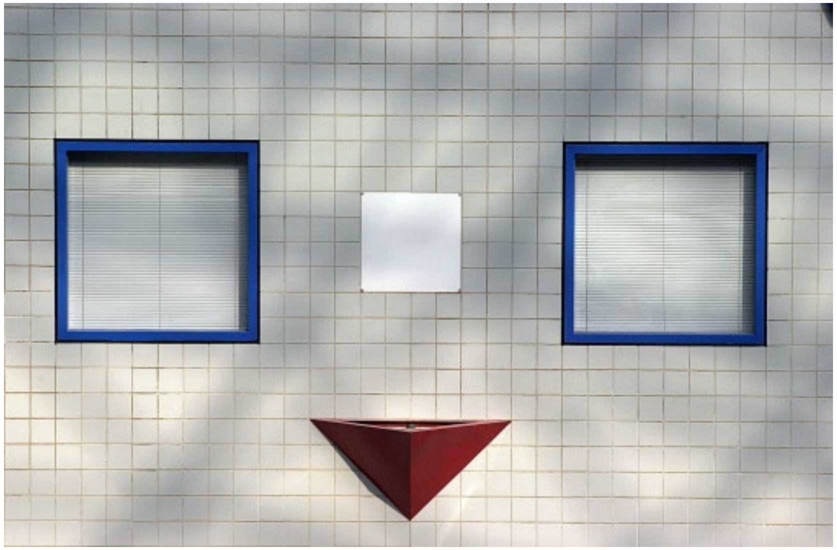
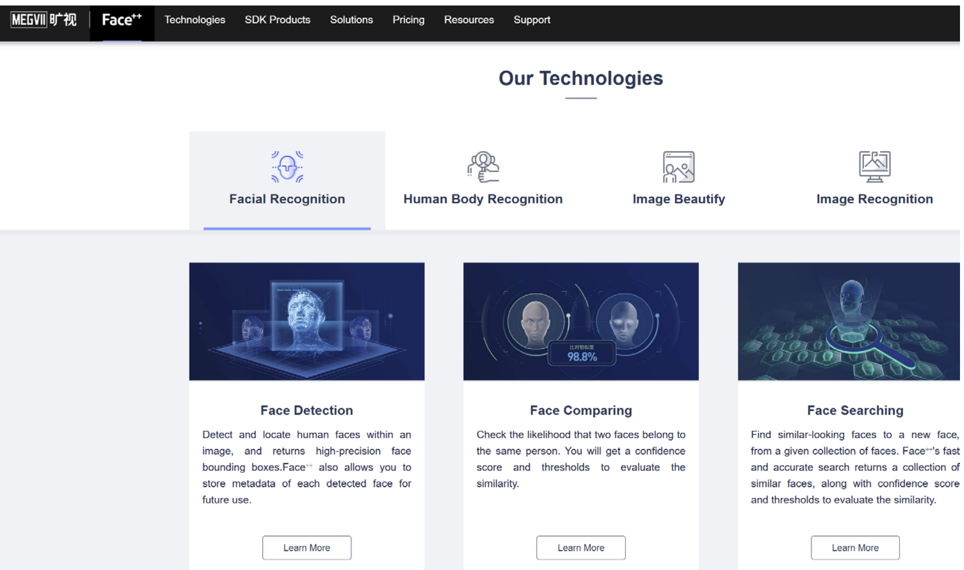


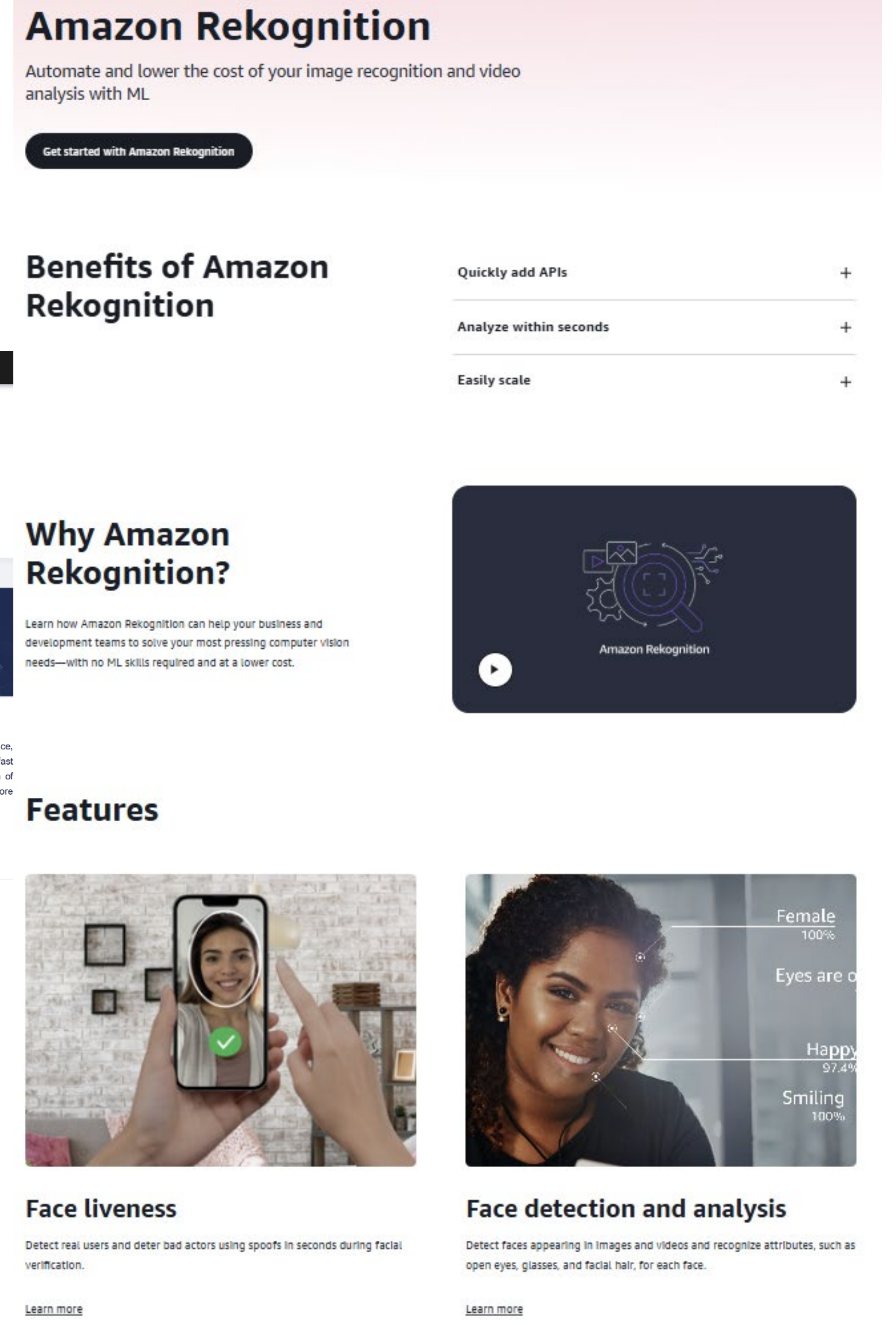
Image Source: http://ptc.ust.hk/

[Li-Jain-05]

Face detection

- Examples



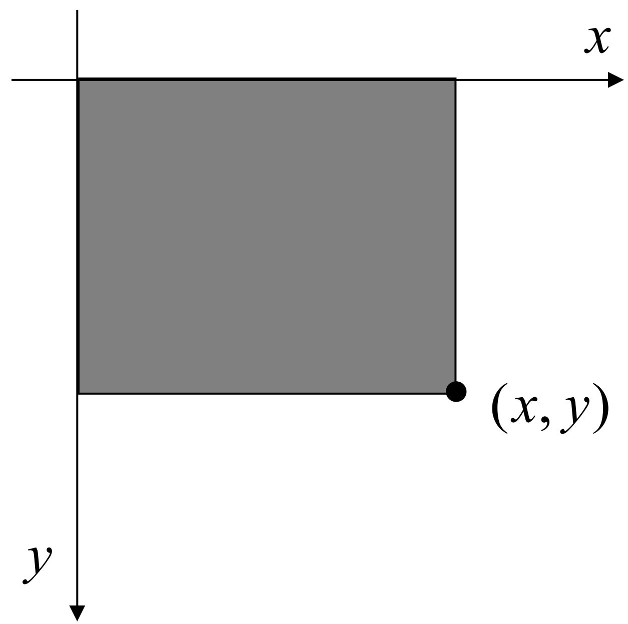


Integral image

- Face detection based on integral image and AdaBoost learning
- The integral image (http://en.wikipedia.org/wiki/Summed_area_table) $ii(x, y)$ at location (x, y) contains the sum of the pixel intensity values above and to the left of the location (x, y), inclusive.
- The ii is defined as

$$ii(x, y) = \sum_{x' \le x, y' \le y} i(x', y')$$

where $ii(x, y)$ is the integral image and $i(x, y)$ is the original input image.



[Viola-Jones-04]

Integral image

- Using the following pair of recurrences:

$$s(x,y) = s(x,y-1) + i(x,y)$$

$$ii(x,y) = ii(x-1,y) + s(x,y)$$

where $s(x, y)$ is the cumulative row sum, $s(x, -1) = 0$, and $ii(-1, y) = 0$, the integral image can be computed in one pass over the original image.

- Using the integral image, any rectangular sum can be computed in four array references.

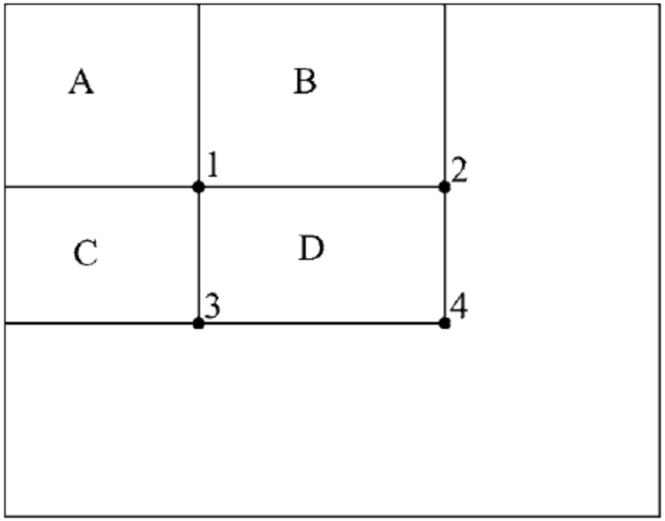
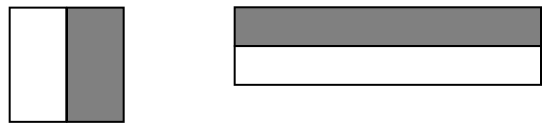


Figure 3. The sum of the pixels within rectangle D can be computed with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is $A + B$, at location 3 is $A + C$, and at location 4 is $A + B + C + D$. The sum within D can be computed as $4 + 1 - (2 + 3)$.

[Viola-Jones-04]

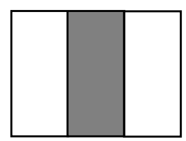
Rectangle features

- The features for face detection are Haar-like functions.
- There are three kinds of features.
- [1] Two-rectangle feature: The difference between the sum of the pixels within two rectangular regions.



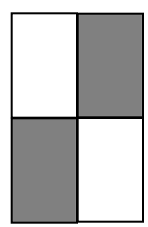
The sum of the pixels which lie within the white rectangle is subtracted from the sum of pixels in the grey rectangle.

- [2] Three-rectangle feature: The feature is the sum within two outside rectangles subtracted from the sum in a center rectangle.



The sum of the pixels which lie within the white rectangles is subtracted from the sum of pixels in the grey rectangle.

- [3] Four-rectangle feature: The difference between diagonal pairs of rectangles.



The sum of the pixels which lie within the white rectangles is subtracted from the sum of pixels in the grey rectangles.

More info about Haar-like features: https://docs.opencv.org/4.3.0/d5/d54/group__objdetect.html

[Viola-Jones-04]

Rectangle features

- The rectangle features are sensitive to the presence of edges, bars/lines, and other simple image structures in different scales and at different locations.

- Given that the base resolution of the detector is 24 x 24 pixels, the exhaustive set of rectangle features is quite large, $\underbrace{\text{160,000}}_{\text{per } 24 \times 24 \text{ pixels}}$.

- The computations of rectangle features are extremely efficiently.

- Given a feature set and a training set of positive and negative images, a classification function must be learned to classify a pattern into either face or non-face.

[Viola-Jones-04]

Learning algorithm

- In this work, the classifier is designed based on the assumption that a very small number of features can be combined to form an effective classifier.

- The AdaBoost (http://en.wikipedia.org/wiki/AdaBoost) learning algorithm is used to boost the classification performance of a simple learning algorithm. The simple learning algorithm is applied to all rectangle features.

- It does this by combining a collection of weak classification functions (weak classifiers with relatively high classification error) to form a stronger classifier. The final strong classifier takes the form of a weighted combination of weak classifiers followed by a threshold.

[Viola-Jones-04]

Adaboost

- Weak classifier h_t

$$h_t(\vec{x}) = \begin{cases} 1 & \text{if } \vec{x} \text{ represents a face image} \\ -1 & \text{otherwise} \end{cases}$$

- Adaboost is an algorithm for constructing a strong classifier from a linear combination of selected weak classifiers.

$$\sum_{t=1}^{T} \alpha_t h_t(\vec{x}) \quad \text{where } \alpha_t \text{ is weight}$$

- Strong classifier H

$$H(\vec{x}) = \mathrm{sgn}\left(\sum_{t=1}^{T} \alpha_t h_t(\vec{x})\right) \quad \text{where } \mathrm{sgn}(x) \text{ is sign function}$$

$$\mathrm{sgn}(x) = \begin{cases} -1 & \text{if } x \le 0 \\ 1 & \text{if } x > 0 \end{cases}$$

[AZ-AV-09, Viola-Jones-04]

Adaboost algorithm 1

- Given example images

$$\vec{x}_i, i = 1, \ldots, N \quad \text{where } N = \text{total number of images}$$

and classifications

$$\left(\vec{x}_i, y_i\right) \text{ where } \; y_i \in \{-1, 1\}$$

- For each weak classifier, h_j
    - The weak classifier is trained such that the classification error

$$e_j = \sum_{i=1}^{N} w_i \left(\frac{1 - h_j\left(\vec{x}_i\right) y_i}{2} \right) \quad \text{where } w_i \text{ is weight for each image}$$

$$\sum_{i=1}^{N} w_i = 1$$

is minimized.

[AZ-AV-09, Viola-Jones-04]

Adaboost algorithm 1

Frontal upright faces

Figure 8. Example of frontal upright face images used for training.

[AZ-AV-09, Viola-Jones-04]

Adaboost algorithm 2

- Start with equal weights on each image $\vec{x}_i$
- For $t = 1, \ldots, T$
  - Normalize all weights $w_i = \dfrac{w_i}{\sum_{j=1}^{N} w_j}$ such that $\sum_{i=1}^{N} w_i = 1$
  - Select the weak classifier h_k with minimum error

一般小于0.5

$$e_k = \sum_{i=1}^{N} w_i \left(\frac{1 - h_k(\vec{x}_i)\, y_i}{2} \right) \quad \text{where } 0 \le e_k \le 1$$

  - Set weight α_t for the selected weak classifier

$$\alpha_t = \frac{1}{2} \ln \left(\frac{1 - e_k}{e_k} \right)$$

ek越小的weak feature 在组合feature中的权重越大.

  - Reweight the examples (boosting)

$$w_i = w_i e^{\left(-\alpha_t y_i h_k(\vec{x}_i)\right)}$$

对于错误的图片提高比重，对于正确的图片降低比重、

[AZ-AV-09, Viola-Jones-04]

Adaboost algorithm 2

- The final strong classifier is

$$H(\vec{x}) = \operatorname{sgn}\left(\sum_{t=1}^{T} \alpha_t h_t(\vec{x})\right)$$

- In summary, it combines a number of weak classifiers, defined by *T*, to form a final strong classifier. It can be combined with HOG features or CNNs.

- Values of *T* can be 200 for N = 10^8 images and 180,000 filters.

- Given the above strong classifier, a new image can be classified as either face or non-face.

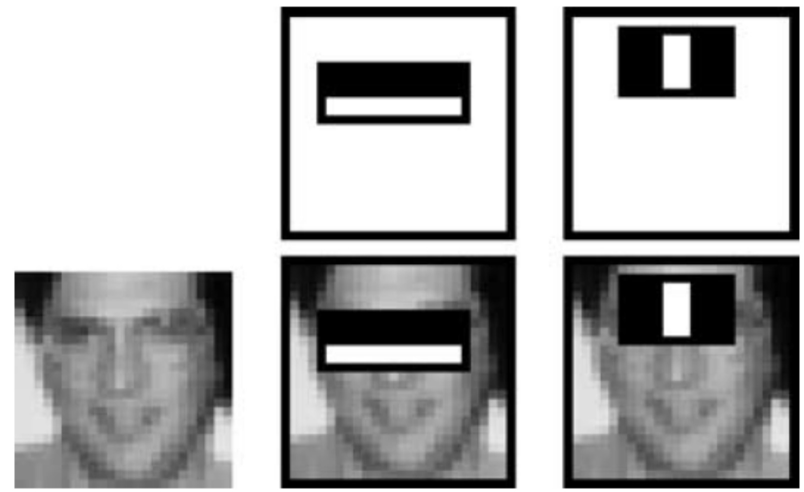


Figure 5. The first and second features selected by AdaBoost. The two features are shown in the top row and then overlayed on a typical training face in the bottom row. The first feature measures the difference in intensity between the region of the eyes and a region across the upper cheeks. The feature capitalizes on the observation that the eye region is often darker than the cheeks. The second feature compares the intensities in the eye regions to the intensity across the bridge of the nose.

[AZ-AV-09, Viola-Jones-04]

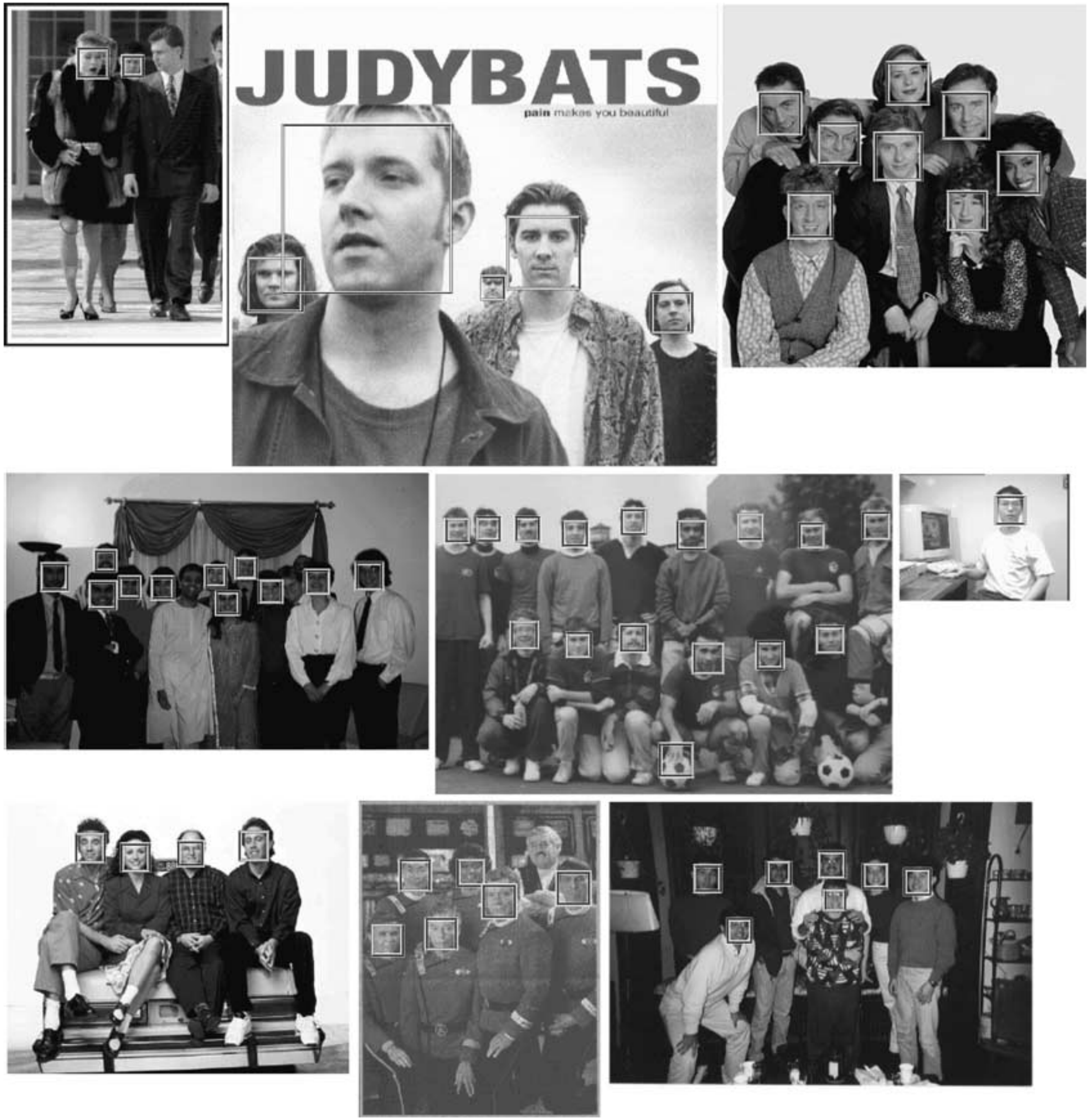


Figure 10. Output of our face detector on a number of test images from the MIT + CMU test set.

[Viola-Jones-04]

Post-processing

- A single face in an image may be detected several times at close locations or on multiple scales.
- A detection is confirmed if the number of multiple detections is greater than a given value; and given the confirmation, multiple detections are merged into a consistent one. It help eliminate many false detections, e.g., detections on the cloth in the image below.

over lap

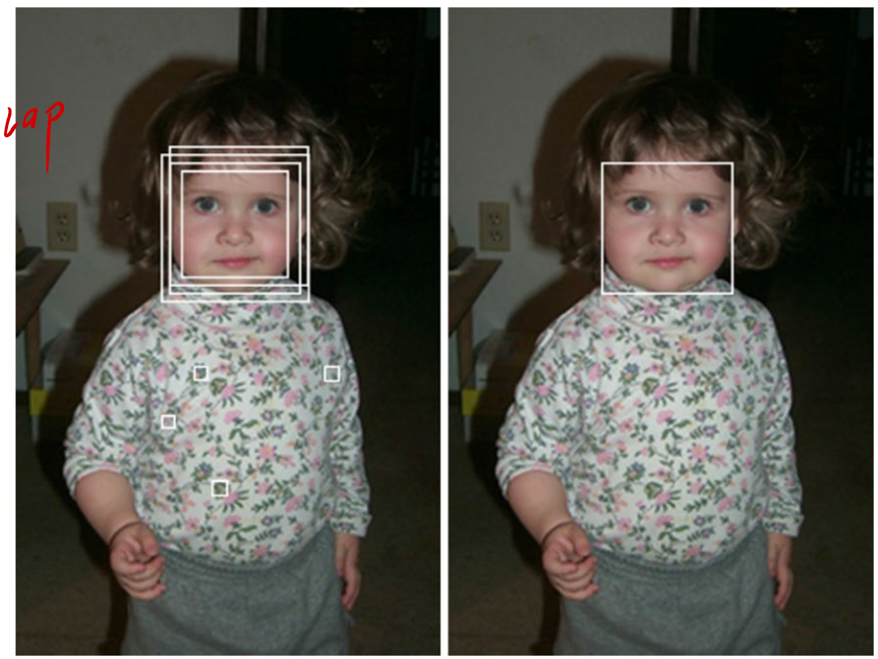


Fig. 11.13 Merging multiple detections

[Li-Jain-11, Chapter 11]

Face Recognition

Reference:

[Li-Jain-05] Stan Z. Li and Anil K. Jain, Handbook of Face Recognition, Springer, 2005

[Collins-07] Lecture notes written by Professor Robert Collins at The Pennsylvania State University, USA for a computer vision course.

[Belhumeur-Hespanha-Kriegman-97] Peter N. Belhumeur, Joao P. Hespanha, David J. Kriegman, Eigenfaces vs. Fisherfaces: Recognition using class specific linear project, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), 711-720, 1997

[Li-Chu-Liao-Zhang-07] Stan Z. Li, RuFeng Chu, ShengCai Liao and Lun Zhang, Illumination Invariant Face Recognition Using Near-Infrared Images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(4), 1-13, 2007

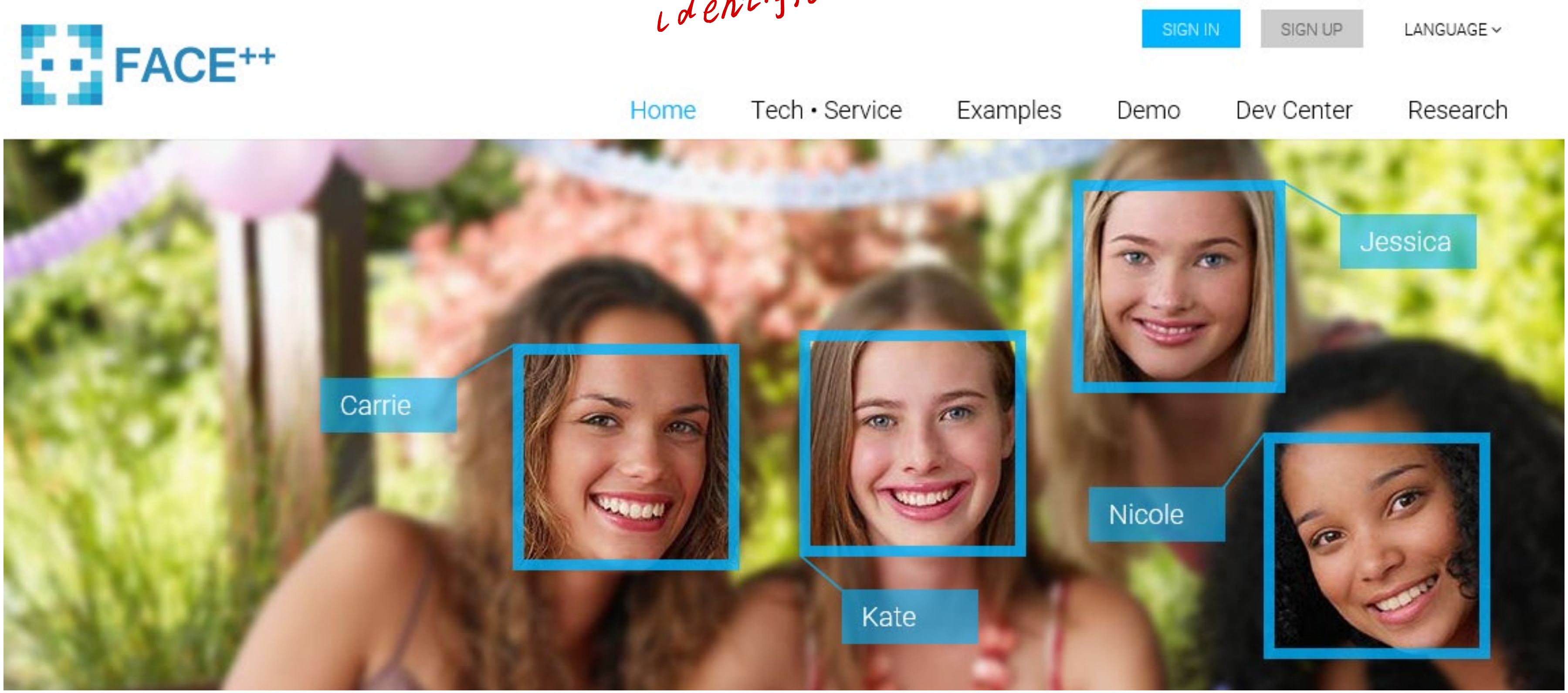
[Essex] The Essex Face Database: http://cswww.essex.ac.uk/mv/allfaces/index.html

The Extended Yale Face Database B: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/ExtYaleB.html

More related useful information can be found at the Face Recognition Homepage : http://www.face-rec.org/general-info/

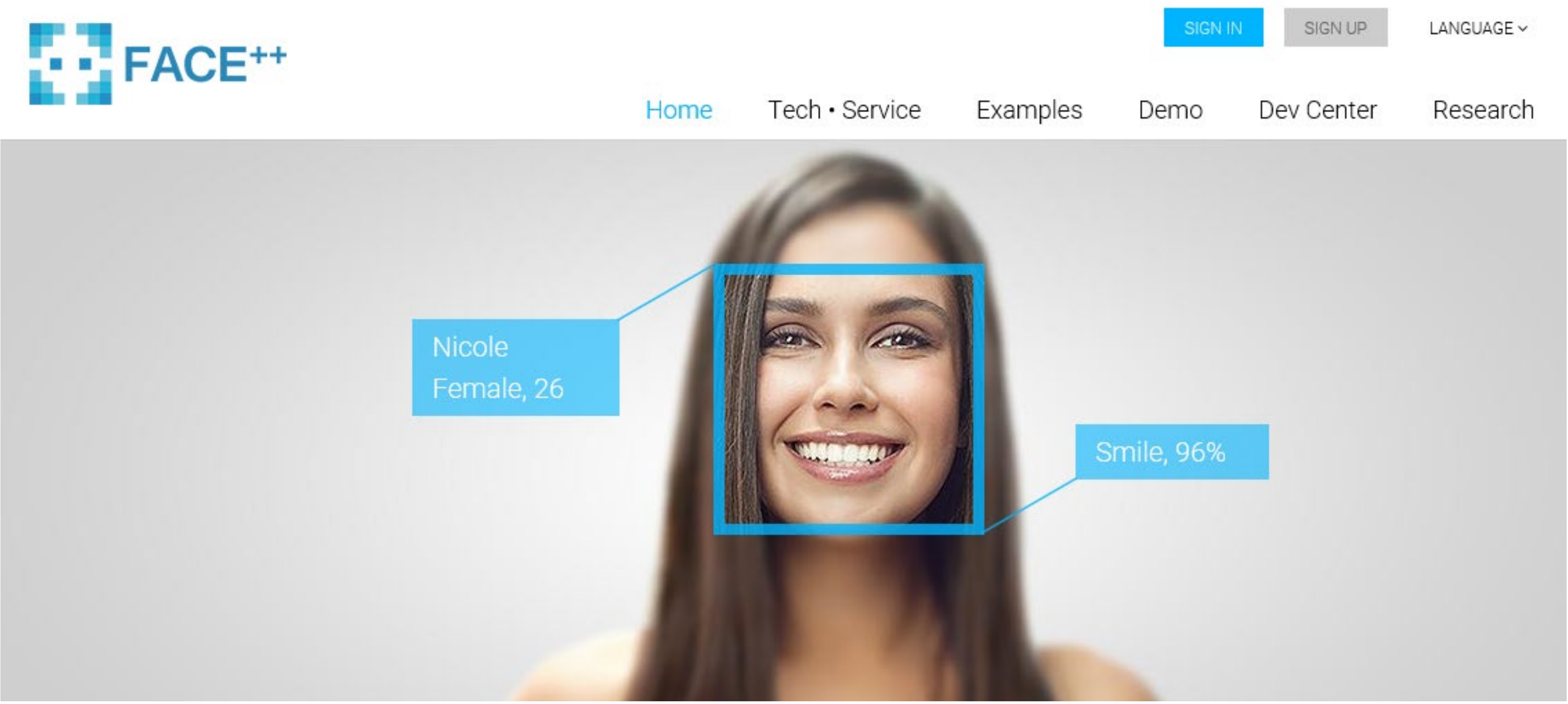
Face Recognition

identification



http://www.faceplusplus.com/

Face Recognition



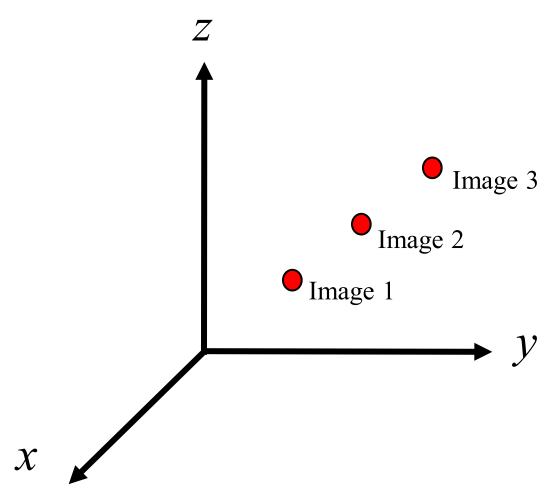
http://www.faceplusplus.com/

Introduction

- Images of faces often belong to a manifold of intrinsically low dimension.

- For example, if there are three 3x1 images (see below), then each image has three intensity values. If each intensity value is viewed as a coordinate in a 3D space, then each image can be viewed as a point in a 3D space.

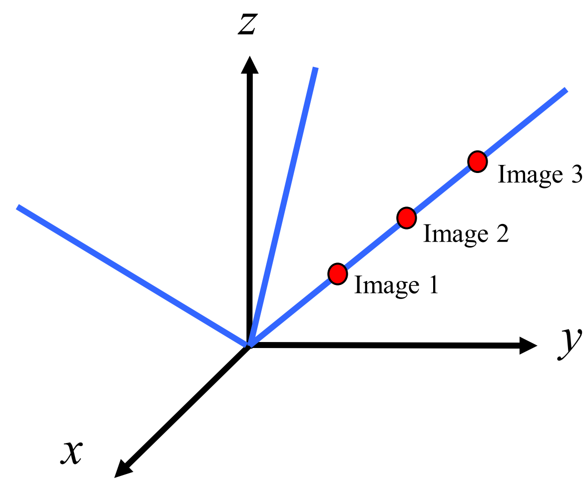
	Image 1	Image 2	Image 3
x	1	2	3
y	2	4	6
z	3	6	9



[Li-Jain-05, Collins-07]

?

	Image 1	Image 2	Image 3
x	1	2	3
y	2	4	6
z	3	6	9



- Consider a new coordinate system in which one of the axes passes through these points. Then, each point can be represented by one value only. The coordinate values in the other axes become zero or very small.

- To represent these points effectively, the number of dimensions can be reduced from three to one. It is the concept of dimensionality reduction (http://en.wikipedia.org/wiki/Dimension_reduction).

[Li-Jain-05, Collins-07]

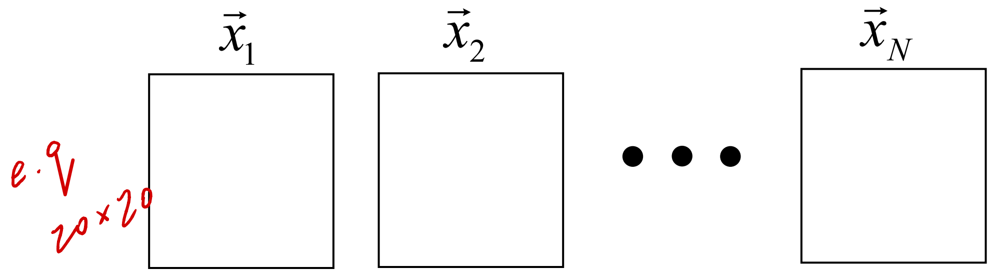
- Similarly, for images with size $m \times n$ and represented by 1D vectors, the number of dimensions can be reduced from $m \times n$ to a significantly low number.

- Principal component analysis (http://en.wikipedia.org/wiki/Principal_component_analysis) (PCA) is a method for performing dimensionality reduction of high-dimensional face images.

[Li-Jain-05, Collins-07]

Eigenfaces

- Let us consider a set of N sample images (image vectors after flattening) with $m \times n$ dimensions



- Each image is represented by a 1D vector with dimensions (mxn) x 1, e.g.
- The mean image vector is given by

$$\vec{x}_1 = \begin{bmatrix} x_{1,1} \\ x_{1,2} \\ \vdots \\ x_{1,mn} \end{bmatrix}$$

mean

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,mn} \end{bmatrix}$$

[Belhumeur-Hespanha-Kriegman-97, Collins-07]

Eigenfaces

- The scatter (covariance) matrix is given by

$N \times 400$

400×400

$$\vec{S} = \left[\vec{x}_1 - \overline{x} \quad \vec{x}_2 - \overline{x} \quad \cdots \quad \vec{x}_N - \overline{x}\right] \begin{bmatrix} \left(\vec{x}_1 - \overline{x}\right)^T \\ \left(\vec{x}_2 - \overline{x}\right)^T \\ \vdots \\ \left(\vec{x}_N - \overline{x}\right)^T \end{bmatrix}$$

$400 \times N$

- The corresponding/derived t eigenvectors with non-zero eigenvalues λ_i are ($t < m \times n$)

max 400

along this vector, images spread most

$$\vec{e}_1 \quad \vec{e}_2 \quad \cdots \quad \vec{e}_t$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_t$

- Then, a face image is represented by a combination of the selected eigenvectors.

$$\vec{x}_j \approx \overline{x} + \sum_{i=1}^{t} g_{ji} \vec{e}_i$$

where $g_{ji} = \left(\vec{x}_j - \overline{x}\right) \cdot \vec{e}_i$

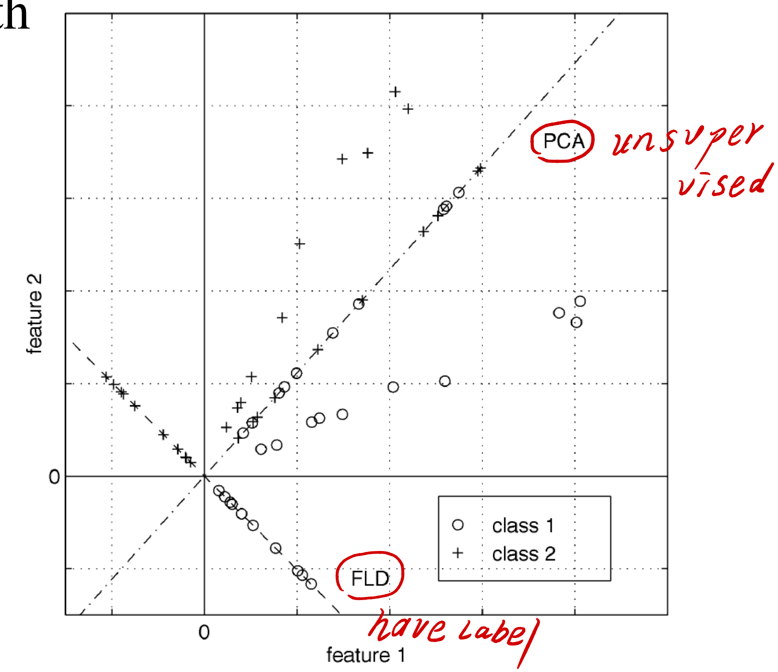


Fig. 2. A comparison of principal component analysis (PCA) and Fisher's linear discriminant (FLD) for a two class problem where data for each class lies near a linear subspace.

[Belhumeur-Hespanha-Kriegman-97, Collins-07]

Eigenfaces

- Since the eigenvectors have the same dimension as the image vectors, the eigenvectors are referred as Eigenfaces (http://en.wikipedia.org/wiki/Eigenface).
- The value of t is usually much smaller than the value of mn. Therefore, the number of dimensions can be reduced significantly.
- For example, there are 20 face images. Some of the images can be found at http://cswww.essex.ac.uk/mv/allfaces/index.html [Essex]. The mean image and their first 3 eigenfaces are shown below.

$\bar{x}$	$\vec{e}_1$	$\vec{e}_2$	$\vec{e}_3$
Mean image	1st Eigenface	2nd Eigenface	3rd Eigenface

…

[Belhumeur-Hespanha-Kriegman-97, Collins-07, Essex]

Eigenfaces

- Given the average image and the eigenfaces, an image can be represented by a linear combinations of average image vector and eigenfaces.

$$\vec{x} \approx \overline{x} + \sum_{i=1}^{t} g_i \vec{e}_i$$

Mean image $\overline{x}$ 1st Eigenface $\vec{e}_1$ 2nd Eigenface $\vec{e}_2$ 3rd Eigenface $\vec{e}_3$

$\vec{x} \approx$  $+g_1$  $+g_2$  $+g_3$ 

$$+ \cdots + g_t \vec{e}_t$$

- The values of g's become the identity of the above image for face recognition.

[Belhumeur-Hespanha-Kriegman-97, Collins-07, Essex]

Recognition using the Eigenfaces

- Process the image database
  - Pre-process the images, e.g. histogram equalization
  - Run PCA and compute eigenfaces
  - Calculate t coefficients for each image, $g_1, g_2, \ldots, g_t$
- Given a new image to be recognized $\vec{x}$
  - Calculate t coefficients for the new image

$$\left(g_1, g_2, \cdots, g_t\right)$$

- Detect if the new image is a face

$$\left\| \vec{x} - \left(\overline{x} + g_1 \vec{e}_1 + g_2 \vec{e}_2 \cdots + g_t \vec{e}_t \right) \right\| < \text{Threshold}$$

- If it is a face, find the closest labeled face based on the nearest neighbor in the t-dimensional space.

[Belhumeur-Hespanha-Kriegman-97, Collins-07, Essex]

More Challenging Face Recognition Problem

(a)

(b)

Acquiring Linear Subspaces for Face Recognition under Variable Lighting, K. Lee, J. Ho and D. Kriegman, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 5, May 2005, pages 1-15.

Near-infrared images for face recognition

- Most current face recognition systems are based on face images captured in the visible light spectrum.

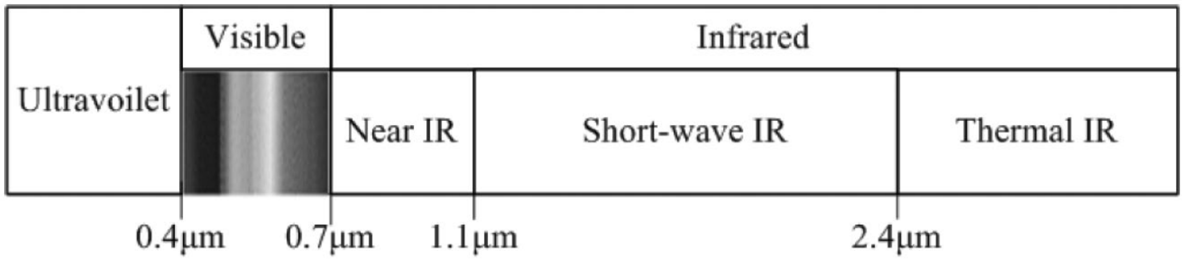


Fig. 1. Radiation spectrum ranges.

- The near infrared (NIR) imaging system is able to produce face images of good condition regardless of the visible lights in the environment. NIR images give a consistent appearance of the face under different environmental lighting variations.

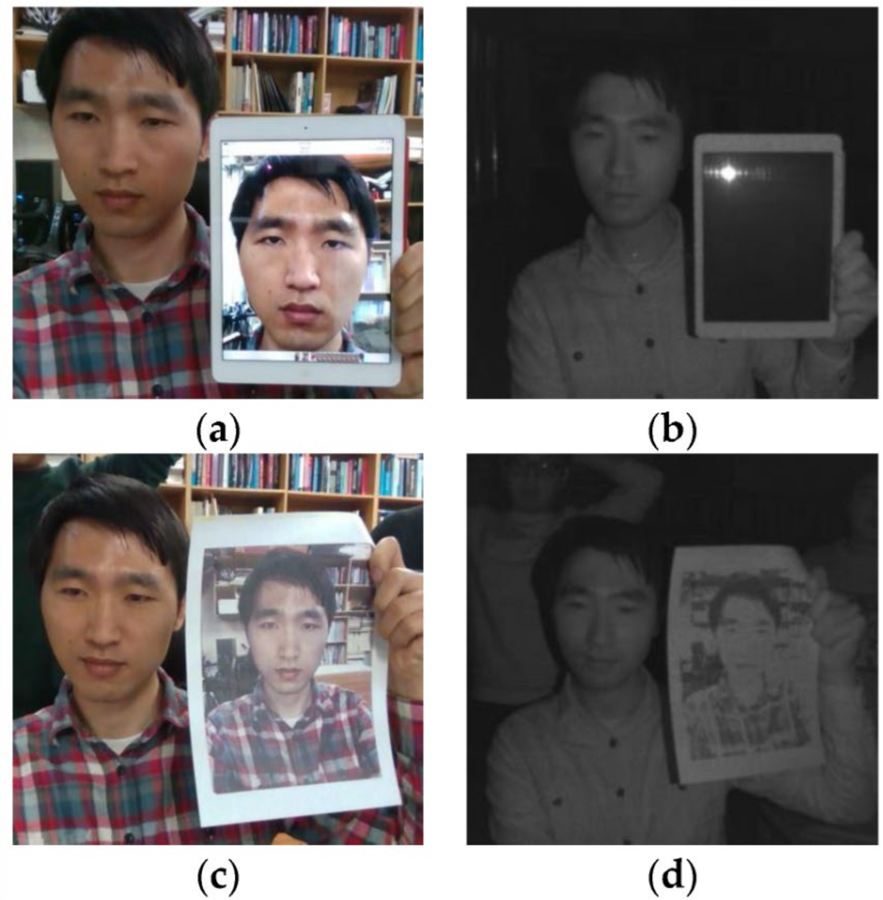
Fig. 3: NIR (top) and VIS (bottom) images of three subjects (columns) from the CASIA NIR-VIS 2.0 dataset.
https://mahmoods01.github.io/files/esorics23-nir-attacks.pdf

[Li-Chu-Liao-Zhang-07]

Near-infrared images for face recognition

- An NIR face recognition system effectively prevents spoofing attacks —unauthorized attempts to bypass the system with fake faces.
- The most common type of spoofing attack is the use of photos, either printed on paper or displayed on a digital device.
- An NIR face recognition system can easily block such attempts in the image acquisition stage because fake faces in the NIR spectrum are different from real faces.

Figure 1. Spoofing attacks on RGB and near-infrared (NIR) images: (**a**) RGB image holding a face image in RGB on a digital device; (**b**) NIR image of (a); (**c**) RGB image holding a printed face image in RGB; (**d**) NIR image of (**c**). All images were taken by the Intel RealSense SR 300.

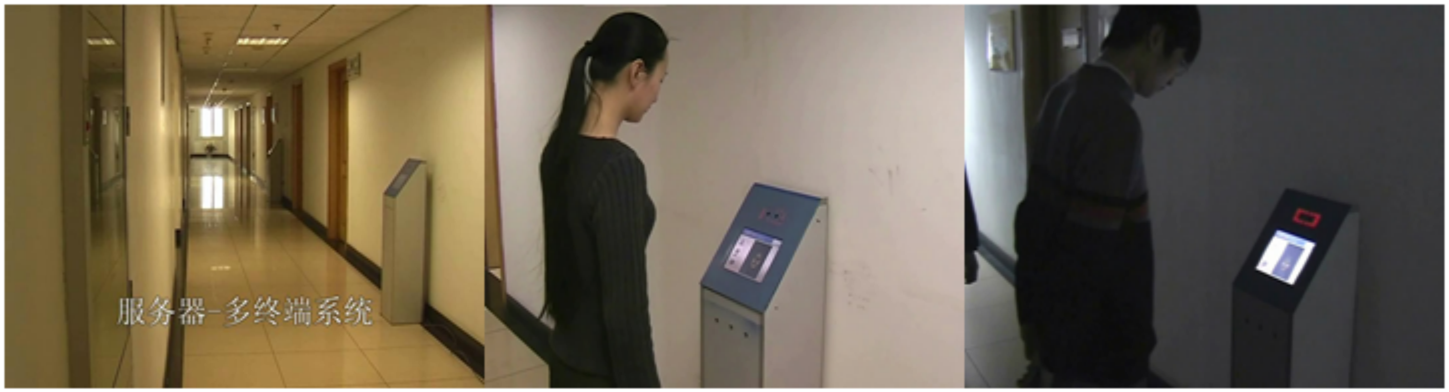


https://www.mdpi.com/2073-8994/11/10/1234

[Li-Chu-Liao-Zhang-07]

Near-infrared images for face recognition

AuthenMetric-F1: Near Infrared Based Face Recognition System



AuthenMetric Client Systems in Operation, working under Normal Lighting and in Darkness

video1, video2 AuthenMetric F1 is a highly reliable face recognition system. It consists of innovative hardware and software. The cutting-edge face recognition technology overcomes a major obstacle that limits performance of existing commercial systems, i.e. that of lighting variation. AuthenMetric F1 not only achieves higher recognition accuracy than any existing face and fingerprint recognition systems but is also fast, and can provide liveliness detection against fakes. In the testing scenarios in this demo, every enrolled participant spoke his name to the camera, and then turned to the system; the system identified (1 out of about 1000) his face and pronounced his name through a loudspeaker (the same as spoken by the participant). An un-enrolled participant (female) was rejected by the system. She was enrolled then and identified successfully by the system. The final part of the demo shows accurate and fast identification in a dark lighting condition. AuthenMetric on ARM9

http://www.cbsr.ia.ac.cn/demos/index-en.htm

[Li-Chu-Liao-Zhang-07]

Near-infrared images for face recognition

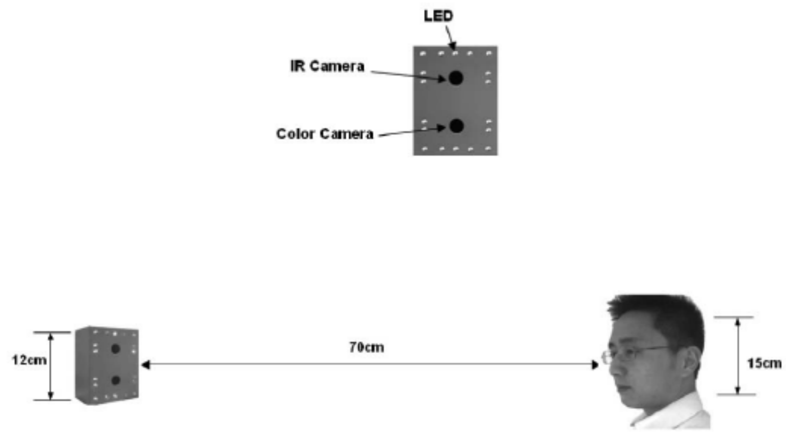


Fig. 2. Active NIR imaging system (upper) and its geometric relationship with the face (lower).

Fig. 3. Color images taken by a color camera versus NIR images taken by the present NIR imaging system. While unfavorable lighting is obvious in the color face images, it is almost unseen in the NIR face images.

Illumination Invariant Face Recognition Using Near-Infrared Images,
S. Li, R. Chu, S. Liao and L. Zhang,
IEEE Transactions on Pattern Analysis and Machine Intelligence,
Vol. 29, No. 4, April 2007, pages 627-639.