KARLSRUHE INSTITUTE OF TECHNOLOGY

Biometric Systems for Person Identification

Manuel Lang

August 17, 2017

CONTENTS

1 Introduction

1.1 Background

- Bio: life, metric: measurement
- Authentification/identification
  - What I know (passwords, PIN)
  - What I have (ID-cards, smart-cards)
  - What I am (biometrics)
- Applications: banking, IT security, healthcare, law and order, time and attenfance, welfare, consumer products
- Physical vs behavioral biometrics
  - Fingerprints
  - Iris
  - Hand geometry
  - Face
  - Vein pattern
  - Retinal scanning
  - Ear shape
  - Voice
  - Signature
  - Typing pattern
  - Gait recognition (how you walk)
  - Heart rate analysis

1.2 Performance measurement

- Usability metrics
  - Failure to enroll (FTE): Failure of a biometric system to form a proper enrollment reference for an end user. Common failures include end users who are not properly trained to provide their biometrics, the sensor not capturing information correctly, or captured sensor data of insufficient quality to develop a template.
  - Failure to Acquire (FTA): This failure occurs if the feature extraction (including all preceding operations) was not successful during a recognition attempt. Reasons may be inability to capture, insufficient sample quality (e.g., too noisy sample data), or insufficient number of features (e.g., too few minutiae). The probability of a Failure to Acquire event is called Failure to Acquire Rate (FTA). FTA can be adjusted by increasing or decreasing quality thresholds. Generally, a high quality threshold need not correspond to a better over-all recognition performance!
- Performance Metrics
  - False Acceptance Rate (FAR) aka False Match Rate (FMR): accepting user who is not registered, mistaking one registered user for another

  - Fals Rejection Rate (FRR) aka False Non-Match Rate (FNMR): rejeting registered user
  - High FRR reduces usability
  - High FAR reduces security

2 Pattern Recognition

2.1 Descriptors

2.1.1 Gabor Wavelets

- Gabor filters: 2-D sine waves modulated by a Gaussian envelope.
- Gabor Wavelet Transform: $G_{mn}(x,y) = \sum_{s=0}^{S} \sum_{t=0}^{T} I(x-s, y-t) g_{m,n}(s,t)$

2.1.2 Edge Histogram

- Captures the spatial distribution of different types of edges

2.1.3 SIFT

- divide area around keypoint into 4x4 subregrions
- build orientation histogram with 8 bins for each subregion
- normalize resulting 128D (4x4x8) vector to unit length
- properties
  - scale-invariant
  - rotation-invariant
  - robust to illumination change
  - robust to noise
  - robust to minor changes in view-point

2.1.4 SURF

- fast approximation of SIFT idea
- equivalent quality for object identification

2.1.5 Local Binary Pattern Histogram

- divide image into cells
- compare each pixel to each of its neighbors
- where the center pixel's value is greater than the neighbor's value, write "1", otherwise, write "0"
- compute histogram over the cell
- use the histogram over the cell for classification

2.2 Keypoint Detectors

2.2.1 Hessian & Harris

2.2.2 Laplacian, DoG

2.2.3 Harris-/Hessian-Laplace)

2.3 Dimensionality Reduction

2.3.1 PCA

- leave out dimensions and minimize error made
- covariance matrix
  - given n sets samples $v_1, ..., v_n$
  - with means $v_i^{'}$ and $v_i = (a_1, ..., a_d)$ (d = dimensionality)
  - a multi-dimensional covariance estimator is defined as $cov(V)_{i,j} = \frac{1}{n-1} v_i - v_i^{'}$
- goal of PCA: make covariance matrix as diagonal as possible
- The covariance matrix is diagonalized by an orthogonal matrix of its eigenvectors.
- The higher the eigenvalue the more variance is captured along the dimension.
- change of basis
  - given the new basis vectors $p_1, ..., p_n$ we can transform data samples x_i in the following manner $PX = \begin{bmatrix} p_1 \\ \vdots \\ p_m \end{bmatrix} \begin{bmatrix} x_1 \dots x_n \end{bmatrix}$
  - i. e. we are projecting x_i onto the new basis vectors $y_i = \begin{bmatrix} p_1 \cdot x_i \\ \vdots \\ p_m \cdot x_i \end{bmatrix}$
- assumptions

  - basis (principle components) is orthogonal
  - change of basis is a linear operation (for non-linear problems: kernel PCA)
  - mean and variance are sufficient statistics
  - large variances have important dynamics

2.3.2 LDA

2.4 Classification

2.4.1 Bayesian Classification

- predict class ω_i of given feature vector x with Bayes rule
- $P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}$ with $p(x) = \sum_i p(x|\omega_i)P(\omega_i)$

2.4.2 Gaussian Classification

- assumption: $p(x|\omega_i) \approx N(\mu,\Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]$
- problem: if the assumption(s) do not hold, the model does not represent reality well
- estimation of μ and Σ with Maximum Likelihood (ML)
  - use paramaters that best explain the data (highest likelihood): $l(\mu,\Sigma) = p(data|\mu,\Sigma) = p(x_0,x_1,...,x_n|\mu,\Sigma) = p(x_0|\mu,\Sigma)\cdot p(x_1|\mu,\Sigma)\cdot ...\cdot p(x_n|\mu,\Sigma)$
  - $log(l(\mu,\Sigma)) = log(p(x_0|\mu,\Sigma)) + ... + log(p(x_n|\mu,\Sigma))$
  - maximimize $log(l(\mu,\Sigma))$ over μ and Σ

2.4.3 Gaussian Mixture Models (GMMs)

- approximate true density function using a weighted sum of several Gaussians
- $p(x) = \sum_i w_i \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)]$ with $\sum_i w_i = 1$
- any density can be approximated but might need many Gaussians
- need to estimate parameters of the Gaussians as well as the weights (use Expectation Maximization (EM) Algorithm)

2.4.4 Expectation Maximization (EM)

- we don't exactly know to which Gaussian each data point belogs, but we can estimate
- EM Algorithm
  - initialize parameters of GMM randomly
  - repeat until convergence

    * Expectation: compute the probability p_{ij} that data point i belongs to Gaussian j, take the value of each Gaussian at point i and normalize so they sum up to one
    * Maximization: compute new GMM parameters using soft assignments p_{ij}, maximum likelihood with data weighted according to p_{ij}

2.4.5 PARAMETRIC VS. NON-PARAMETRIC

- Gaussian and GMMs are called parametric classifiers as they assume a specific form of probability distribution with some paramaters
- methods which do not assume a specific form of probability distribution are called non-parametric (e. g. parzen windows, k-nearest neighbors)
- parametric classifiers need less training data because less parameters need to be estimated, but only work well if model fits data
- non-parametric classifiers work well for all types of distributions, but need more data to correctly estimate the distribution

2.4.6 GENERATIVE VS. DISCRIMINATIVE

- methods that model $P(\omega_i)$ and $p(x|\omega_i)$ explicitly are called generative models: $p(x|\omega_i)$ allows to generate new samples of class ω_i
- discriminative models directly model $P(\omega_i|x)$ or just output a decision ω_i given an input pattern x
- discriminative models are often easier to train because they solve a simpler problem

2.4.7 LINEAR DISCRIMINANT FUNCTIONS

- separate two classes with a linear hyper plane: $y(x) = w^T x + w_0$
- decide y_1 if $y(x) > 0$ and y_2 if $y(x) < 0$ with w^T the normal vector of the hyper plane
- examples: Perceptron, Linear SVM

2.4.8 INSTANCE-BASED LEARNING

- learning means storing all training instances
- classification = assigning target function to a new instance
- referred to as lazy learning instance
- examples: Template-Matching, k-Nearest Neighbor

3 Iris Recognition

3.1 Iris Recognition System

- Acquisition of the eye -> Image
- Locate pupil and iris (pupil not relevant)
- demarcated zones (radial features in iris)
- polar representation
- gabor filters
- iris code

3.2 Daugman's integro-differential operator

- the main step is to find the pupillary (inner) and limbic (outer) boundaries
- Daugman's integro-differential operator was historically used
  - $max_{r,x_0,y_0} G_\sigma(r) * \int \frac{d}{d_r} \frac{I(x,y)}{2\pi r} d_s$
  - searches over the image domain (x,y) for the maximum in the blurred partial derivative with respect to increasing radius r, of the normalized contour integral of I(x,y) along a circular arc d_s of radius r and center (x_0, y_0)
  - chosoe (x_0, y_0, r) that maximizes Gaussian smoothing over r of difference with respect to r of average pixel intensity along (portions of) circle x_0, y_0, r
  - bounding box of the largest dark region might be a range to search for x_0, y_0 with r in a feasible range
  - non-circular pupillary and limbic boundaries have been handled by: fitting ellipses, fitting circles and adjusting using active contours, using balloon active countours
  - inner and outer boundaries are used to unwrap the iris region to a rectangle
  - rubber sheet model: all iris regions in all images are mapped to the same size rectangular version (each pixel is mapped into a polar pair)
  - iris code: texture filters can then be applied at a fixed grip on the rectangle to generate the same size iris code for any image
  - the rectangular image also has a binary mask that tells where the iris region was occluded/what iris code bits not to use
  - the amsk records where the natural iris texture is occluded by eyelids, specular reflections, eyelashes, strands of hair,...
  - encoding with 2-D gabor filtrs
  - measurement with fractional hamming distance (FHD) where 1.0 means all different, 0.0 all same and 0.5 random agreement

4 Palmprint Recognition

4.1 Palmprint

- palms of human hands contain unique pattern of ridges and valleys
- larger than finger: expected to be even more reliable
- more expensive scanners because of bigger area
- highly accurate biometric system could be combined using hands

4.2 Features

- geometric features: width, length and area of palm
- line features: length, position, depth and size of various lines
- point features or minutiae: ridges, ridge endings, bifurcation and dots, datum point: endpoints of principal lines

4.3 Algorithms

4.3.1 Online Images

- Histogram equalization
- Low-pass filtering

4.3.2 Offline Images

1. Estimate and midfy the orientation field
2. Remove noise in a grey-scale image (filter)
3. Binarization based on local threshold
4. Noise removal in binary image (morphological operators)

4.3.3 Minutiae-based matching

- most widely used technique
- location, direction and orientation of each point
- higher recognition accuracy
- does not take advantage of textural or visual features
- time consuming because of muntiae extraction

4.3.4 CORRELATION-BASED MATCHING(ALIGNMENT)

- line up the palm images and subtract them
- determine if the ridges in the two palm images correspond
- less tolerant to elastic, rotational and translational variances and noise within the image
- algorithm
  1. binarize the image using a threshold
  2. obtain the boundaries of the gaps
  3. compute the tangent of the two gaps
  4. line up (x1, y1) and (x2, y2) to get the y-axis of the palmprint coordinate system
  5. extract a sub-image of fixed size based on the coordinate system (gabor filter for feature extraction, hamming distance for matching)

4.3.5 RIDGE-BASED MATCHING

- ridge pattern landmark features such as sweat pores, spatial attributes and geometric characteristics of ridges and/or local texture analysis
- faster than minutiae
- overcomes difficulties associated with extracting minutiae from poor quality images
- lower distinctiveness than the minutiae
- algorithm
  1. line feature extraction
  2. line feature matching (two lines are considered the same if they have a small euclidean distance)

4.3.6 MINUTIAE BASED LATENT-FULL

- Features used: ridge orientation and ridge period features extracted around each minutiae
- clustering using k-means
- alignment using clustering
- matching using a circular spatial grid
- match propagation to see if nearby minutiae pairs also match (final latent to full match)

5 Fingerprint Recognition

5.1 Types of fingerprints

- 60-65% of population has loops
- 30-35% has whorls
- and 5% has arches

5.2 Feature extraction

- at the local level, there are different local ridge characteristics
- the two most prominent ridge characteristics, called minutiae, are ridge termination and ridge bifurcation
- at the very-fine level, intra-ridge details (sweat pores) can be detected very distinctive; however, very high-resolution images are required

Local ridge orientation and frequency

- orientation is computed based on gradient phase angles
- robust computation based on local averaging of gradient magnitudes
- the local frequency at a point [x,y] is the inverse of the number of ridges per unit length along a segment orthogonal to the orientation

Singularity and core detection

- the Poincaré index $P_{G,C}$ under a vector field G and an in G immersed curved C is defined as the total rotation of the vectors of G along C
- $P_{G,C}(i, j) = 360$ř for a whorl type singular region
- $P_{G,C}i, j = 180$ř for a loop type singular region
- $P_{G,C}i, j = -180$ř for a delta type singular region

5.3 Matching

- Problems: displacement, rotation, partial overlap, nonlinear disortion, changing skin condition, noise, feature extraction errors, etc.
- many ambiguous fingersprints, whose exclusive membership cannot be reliably stated even by human experts
- correlation-based matching: intensity based correlation between the fingerprint images are computed

  - cross-correlation between the two templates
  - has many problems and is not reliable

- minutae-based matching: minutiae are extracted from two fingerprints and stored as sets of points in the 2d plane. matching is done based on minutae pairings.
  - template representation T and input fingerprint I (feature vector of variable length)
  - each miniutia m is a triplet with location x,y and angle Θ
  - $T = m_1, m_2, ..., m_m, m_i = x_i, y_i, \Theta_i, i = 1..m$
  - $I = m'_1, m'_2, ..., m'_n, m'_j = x'_j, y'_j, \Theta'_j, j = 1..n$
  - m - number of minutiae in T and n number of minutiae in I
  - matching if spatial distance (sd) between I and T is smaller than a given tolerance r_0 and the direction difference dd is smaller than an angular tolerance Θ_0
  - mapping is needed: function that maps a minutia m'_j from I into m'_j''
  - RANSAC (RANdom SAmple Consenus)
    * Objective robust fit of model to data set S which contains outliers
    * algorithm
      1. Randomly select a sample of m minutiae points and instantiate the model from this subset
      2. Determine the set of points m_i which are within a distance threshold t of the model. The set mm is the consensus set of samples and defines the inliers of S.
      3. If the subset of S_i is greater than some threshold T, reestimate the model using all the points in S_i and terminate
      4. If the size of S_i is less than T, select a new subset and repeat the above.
      5. After N trials the largest consensus set S_i is selected and the model is reestimated using all the points in the subset S_i
    * computational complexity might be high -> use rough alignment: find core, find average ridge orientation on left and right side of core, rotate fingerprint around the core such that the different before left and right ridge orientations are minimum

- ridge feature-based matching: local orientation and frequency of ridges, ridge shape, texture, etc are used for matching
  - size of the fingerprint and shape of the external fingerprint shilhouette
  - number, type and position of singularities
  - shape features
  - global and local texture information
  - sweat pores
  - fractal features

6 Face Recognition

6.1 Background

6.2 Tasks

- Closed set recognition
- Open set recognition
- Verification
- Known/Unknown

6.3 Traditional Approaches

6.3.1 Appearance-based

- holistic, fiducial regions, statistical (i. e. they process the whole face as the input)
- local feature based (i. e. they process facial features, such as eyes, mouth, etc. seperately)

Eigenfaces

- A face image defines a point in the high dimensional image space
- different face images share a number of similarities with each other
- face images can be described by a relatively low dimensional subspace
- project the face images into an appropriately chosen subspace and perform classification by similarity computation (distance, angle)
- dimensionality reduction procedure used here is called Karhunen-Loéve transformation or principal component analysis
- PCA for face images
  - y: face image (1d)
  - face matrix $Y = [y_1, y_2, y_3, ..., y_K]$
  - mean face $m = (1/K) * \sum y$
  - covariance matrix $C = (Y - m)(Y - m)^T$
  - eigenvalues $D = U^T C U$ with eigenvectors U
  - representation coefficients $\Omega = U^T * (y - m)$
- training
  - acquire initial set of face images (training set): $Y = [y_1, y_2, y_3, ..., y_K]$

  - calculate the eigenfaces from the training set keeping only the M images corresponding to the highest eigenvalues: $U = (u_1, u_2, ..., u_M$
  - calculate representation of each known individual k in face space $\Omega_k = U^T * (y_k - m)$

- testing
  - project input new image y into face space: $\Omega = U^T * (y - m)$
  - find most likely candidate class k by distance computation $\epsilon_k = ||\Omega - \Omega_k||$ for all Ω_k

- principal components are called eigenfaces and they span the face space

- projections onto the face space
  - images can be reconstructed by their projections in face space: $Y_f = \sum_{i=1}^{M} \omega_i u_i$
  - appearance of faces in face-space does not change a lot
  - difference of mean-adjusted image (Y-m) and projection Y_f gives a measure of faceness

- extension: view-based eigenspaces

- problems and shortcomings
  - Eigenfaces do not distinguish between shape and appreance: Active Shape Models (ASM), Active Apperance Models (AAM)
  - PCA does not use class information: PCA projections are optimal for reconstruction from a low dimensional basis, they may not be optimal from a discrimination standpoint: "Much of the variation from one image to the next is due to illumination changes."

Linear Discriminant Analysis (LDA) - Fischerfaces

- preserves separability of classes

- maximizes ratio of projected between-classes to projected within-class scatter

- $W_{fld} = \arg\max_W \frac{|W^T S_B W|}{|W^T S_W W|}$

- Between-class scatter $S_B = \sum_{i=1}^{c} |x_i|(\mu_i - \mu)(\mu_i - \mu)^T$

- Within-class scatter $S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$

- c: number of classes, μ_i: mean of class X_i, $|X_i|$: number of samples of X_i

Local vs holistic

- Local variations on the facial appearance, i.e. due to different expression, occlusion and lighting, lead to modifications on the entire representation in the holistic approaches, while in local approaches only the corresponding local region is effected.
- Face images contain different statistical illumination - high frequency at the edges, i.e. eyebrows, low frequence at smooth regions, i.e. cheeks. Easier to represent the varying statistics linearly by using local representation.
- Local approaches facilitate the weighting of each local region in terms of their effect on face recognition.

Modular Eigenspaces

- does classification using fiducial regions (eyes, nose, -mouth is excluded in this study-) instead of using entire face
- face images are divided into N smaller subimages
- PCA is applied on each of these sub-images
- performed better than global PCA on large variations of illumination and expression

6.3.2 Local feature based Face Recognition

- to mitigate the effects of expression, illumination and occlusion variations by performing local analysis and by fusing the outputs of extracted local features at the feature or at the decision level
- some popular facial descriptions achieving good results: Local binary Pattern Histogram (LBP), Gabor Feature, Discrete Cosine Transform (DCT), SIFT, etc.

Facial feature vector with Gabor wavelet transformation (GWT)

- $O_{u,v}(x,y) = I(x,y) * \psi_{u,v}(x,y)$ with gabor kernel $\psi_{u,v}(x,y)$ and input image $I(x,y)$
- typically, multiple scales u and orientations v are used
- therefore $O_{u,v}(x,y)$ becomes a high dimensional vector
- techniques to reduce the dimension are applied (such as PCA)

Elastic Bunch Graphs (EBG)

- A “Jet” is a set of 40 complex Gabor wavelet coefficients obtained for one image point
- A graph consists of N facial landmark points (nodes)
- nodes are labelled with jets
- edges are labelled with distance vectors

- A “bunch graph” includes different jets for different poses and appearances (e.g. closed eye, open eye, ...)

Why LBP

- two same gradients may correspond to rather different local structures, thus ambiguous
- the concept of uniform LBP provides the possibility to effectively remove outliers

High dimensional dense local feature extraction

- computing features densly
- for example on overlapping patches in many scales in the image
- problem: very very high dimensionality
- solution: encode into a compact form (bag of visual word (BoVW) model, Fisher encoding)

Fisher vector encoding

- aggregates feature vectors into a compact representation
- fitting a parametric generative model e.g. Gaussian Mixture Model (GMM)
- encode derivative of the likelihood of Model with respect to its parameters
- capturing the average first and second order difference between dense features and each of GMM centers $\Phi_k^{(1)} = \frac{1}{N\sqrt{w_k}} \sum_{p=1}^{N} \alpha_p(k)(\frac{x_p-\mu_k}{\sigma_k})$, $\Phi_k^{(1)} = \frac{1}{N\sqrt{2w_k}} \sum_{p=1}^{N} \alpha_p(k)(\frac{(x_p-\mu_k)^2}{\sigma_k} - 1)$
- Fisher vector is obtained by stacking these difference vectors $\phi = [\Phi_1^{(1)}, \Phi_1^{(2)}, ..., \Phi_K^{(1)}, \Phi_K^{(2)}]$
- the fisher vector's dimensionality is 2*K*d, where d is the dimensionality of the underlying feature vectors
- dimensionality still high -> subspace learning (PCA is not a good idea)

Problem: Matching across face pose

- problem: different view-point / head orientation
- recognition results degrade, when images of different head orientation have to be matched
- three major directions to address the face recognition across pose problem: geometric pose normalization (image affine warps), 2D specific pose models, image rendering at pixel or feature level, 3D face model fitting

- alignment using just eye-positions is not sufficient
- idea: find several facial features (mesh)
- use complete mesh to normalize face

Active Appearance Models

- a texture and shape-based parametric model
- efficient fitting algorithm: inverse compositional (IC) algorithm
- independent shape and appearance model $s = (x_1, y_1, x_2, y_2, ..., x_v, y_v)^T = s_0 + \sum_{i=1}^{n} p_i s_i$, $A(x) = A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x), \forall x \in s_0$
- fitting goal: $\arg\min\_p, \lambda \sum_{x \in s_0} [A_0(x) + \sum_{i=1}^{m} \lambda_i A_i(x) - I(W(x,p))]^2$
- fitted modal can be used to warp image to frontal pose (e.g. using piecewise affine transformation of mesh triangles)

Face Recognition Based on Fitting a 3D Morphable Model

- A method for face recognition across variations in pose and illumination
- Simulates the process of image formation in 3D space
- Estimates 3D shape and texture of faces from single images by fitting a statistical morphable model of 3D faces to images
- Faces are represented by model parameters for 3D shape and texture
- The morphable face model is based on a vector space representation of faces that is constructed such that any combination of shape and texture vectors S_i and T_i describes a realistic human face: $S = \sum_{i=1}^{m} a_i S_i$, $T = \sum_{i=1}^{m} b_i T_i$
- face vectors
  - The definition of shape and texture vectors is based on a reference face I_0
  - The location of the vertices of the mesh in Cartesian coordinates is (x_k, y_k, z_k) with colors (R_k, G_k, B_k)
  - Reference shape and texture vecors are defined by: $S_0 = (x_1, y_1, z_1, x_2, ..., x_n, y_n, z_n)^T$ and $T_0 = (R_1, G_1, B_1, R_2, ..., R_n, G_n, B_n)^T$
  - To encode a novel scan I, the flow field from I_0 to I is computed
- PCA is performed on the set of shape and texture vectors separately
- Eigenvectors form an orthogonal basis: $S = \bar{s} + \sum_{i=1}^{m-1} \alpha_i \cdot s_i$, $T = \bar{t} + \sum_{i=1}^{m-1} \beta_i \cdot t_i$

6.5 Deep Neural Networks for face recognition

6.5.1 DeepFace - Facebook

- Learn a deep (7 layers) NN (20 million parameters) on 4 million identiy labeled face images directly on RGB pixels
- Alignment: use 6 fiducial points for 2D warp, then 67 points for 3D model, frontalize the face for input to NN
- Representation: output is fed in k-way softmax, that generates probability distribution over class labels, goal of training is to maximize the probability of the correct class

6.5.2 FaceNet - Google

- problem is that even a modest number of 5x5 convolutions can be prohibitively expensive on top of a convolutional layer with a large number of filters
- solution: reduce the dimensionality
- 1x1 convolutions used to compute reductions before the expensive 3x3 and 5x5 convolutions
- include the use of rectified linear activation (ReLU)
- map images to a compact Euclidean space where distances correspond to face similarity
- CNNs to optimize embedding
- Triplet-based loss function for training

7 Multimodal Biometrics

7.1 Why Multimodal Biometrics?

- Reliability of Single Biometric - error rates too high
- Biometric data is noisy
- Acquiring multiple biometrics is easy
- Failure to enrull rate (FTE) can be reduced
- More robust against spoofing attacks

7.2 Intramodal Combination

- Within the single biometric - combine multiple matchers (experts)

7.3 Multimodal Combination

- Among multiple biometrics - combine single/multiple matchers (experts)

7.4 Design of Multimodal Systems

- Choice and number of biometrics
  - Voice, face
  - Voice, lip Movement
  - Facial thermogram, face
  - Iris, face
  - Palmprint, hand geometry
  - Ear, voice
  - Fingerprint, face
  - Fingerprint, face, voice
  - Fingerprint, face, hand geometry
  - Fingerprint, voice, hand geometry
  - Fingerprint, hand geometry
- Level of fusion
  - Fusion prior to matching (sensor level, feature level)
    - When the feature vectors are homogeneous (e.g., multiple fingerprint impressions of a user's finger), a single resultant feature vector can be calculated as a weighted average of the individual feature vectors.
    - When the feature vectors are non-homogeneous (e.g., feature vectors of difference biometric modalities like face and iris), we can concatenate them to form a single feature vector. Feature vectors must be compatible.
  - Fusion after matching (matching mode level, decision level)
    - At the decision level: decision level fusion (majority voting and or rules), dynamic classifier selection (choose best performing classifier), rank level (fuse the top N returned matches)
    - Matching score level: score normalization required (min-max, Z-score, Tanh), classification approach, combination approach (multiple matchers generate single score each, scores are combined by sum, min, max, match weighting etc)
- Fusion methodology
- Cost vs performance
- Multimodal databases

7.5 Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics

- Fingerprint verification
  - Normalization: crop, normalize to 130*150, remove backgroup, histogram equalization
  - Eigenfacces and Eigenears: PCA computes the eigenvectors and eigenvalues, following the FERET approach, use the eigenvectors corresponding to the first 60 percent of the large eigenvalues and drop the first eigenvector as it represents illumination, another approach uses the fixed percent of total energy
  - Database: training set consits of 197 subjects, each of whom has both a face image and an ear image, this is a separate (gallery, probe) data set for three experiments: the day variation, the lighting variation and the pose variation
- We need to define a measure that indicates the confidence of the decision fusion criterion.
- The confidence of a given decision criterion may be characterized by its FAR.
- In order to estimate FAR, the impostor distribution needs to be computed.
- Decision Fusion
  - Each of the top n possible identities established by the face recognition module is verified by the fingerprint verification module: either rejects all the n possibilites or accepts only of them as the genuine identity.
  - It is usually specified that the FAR of the system should be less than a given value.
  - The goal of decision fusion, in essence, is to derive a decision criterion which satisfied the FAR specification

7.6 Integrating Faces and Fingerprints for Personal Identification

- Fingerprint verification
  - Alignment stage: transformations such as translation, rotation, and scaling between an input and a template in the databased are estimated, then the input minutiae are aligned with the template minutiae
  - Matching stage: both the input minutiae and the template minutiae are converted to strings in the polar coordinate system, and an elastic string matching algorithm is used to match the resulting strings

7.7 Crossmodal biometrics

7.7.1 What is crossmodal?

- different sensors

- different imaging modalities
- practical and largely unsolved problem
- examples: matching across different sensors - low resolution CCTV vs high resolution stored database images, matching across different underlying modalities - infrared captured images vs stored high resolution visible spectrum images

8 SOFTBIOMETRICS

- Human traits that can aid identification
- Behavioral trait, e.g. gait (the way you walk) and many more such as speech or signature
- Physical traits: age, gender, ethnicity
- Sematic traits (what you perceive): hair (long vs short), body weight, clothing, color, glasses, ...

8.1 GAIT BIOMETRICS

- can recognize people
- is available at a distance when other biometrics are obscured or at too low resolution
- by computer vision, it needs moving feature extraction
- many researchers worldwide, many datasets, many approaches
- modeling movement: pendular thigh motion model, coupled and forced oscillator, anatomically-guided skeleton
- silhouette descriptions: established statistical analysis, temporal symmetry, velocity moments, dynamics of area
- average silhouettes signature
  - background is taken from each frame and pixels thresholded resulting in a binary image
  - normalize silhouettes by height to account for distance
  - average silhouettes
  - resulting image is the signature
- recognition: generate average gait silhouette, compare with dabase, perform recognition

- use human labels
- are grounded in psychology
- use psychology in their generation
- analyse correlation between human vision and computer vision
- advantages of semantic descriptions (height, sex): no ageing, available at a distance/low resolution/poor quality, fit with human description/forensics, complement automatically-perceived measures, need for search mechanisms
- disadvantages: psychology/perception, need for labeling

Traits and Terms

- global features
  - features mentioned most often in witness statements
  - sex and age quite simple
  - ethnicity
    - notoriously unstable
    - there could be anywhere between 3 and 100 ethnic groups
    - we've chosen 3 main subgroups and 2 extra to match UK police force groupings
- body features
  - based on whole body description stability analysis by MacLeod et al: features showing consistency by different viewers looking at the same subjects
  - mostly comprised of 5 point qualitative measures (very thing -> very fat, very short -> very long)
  - most likely candidate for association with gait
- head features
  - mentioned cosistently by people even at long distances
  - prominent area of gaze
  - hair length and color inherently connected with style: many different hair styles, style avoided due to unfamiliarity of annotators

8.2.1 Age Estimation

- aging process is rather complex
- differs not only between different ethnic groups
- depends on people's genes, health condition, living style, etc.
- large usage: electronic customer relationship management (ECRM), security, control and surveilance monitoring, biometrics and entertainment to secure age liminations in daily life, protect minors from age-restricted content in the internet, in combination to improve other areas of the computer vision
- implemented using support vector regression on different principle components

8.2.2 Person Search and Retrieval State of the art

- semantic attribute inference (wearing backpack, female, longhair, etc.)
- person reidentification (direct low level embedding for full body description)
- efficient query and retrieval
- View Specific Pedestrian Attribute Inference (VeSPA)
  - deep convolutional neural network
  - underlying GoogleNet structure
  - idea: incorporate pose information in the model

9 Biometric System Attacks

9.1 Automated Biometric System Model

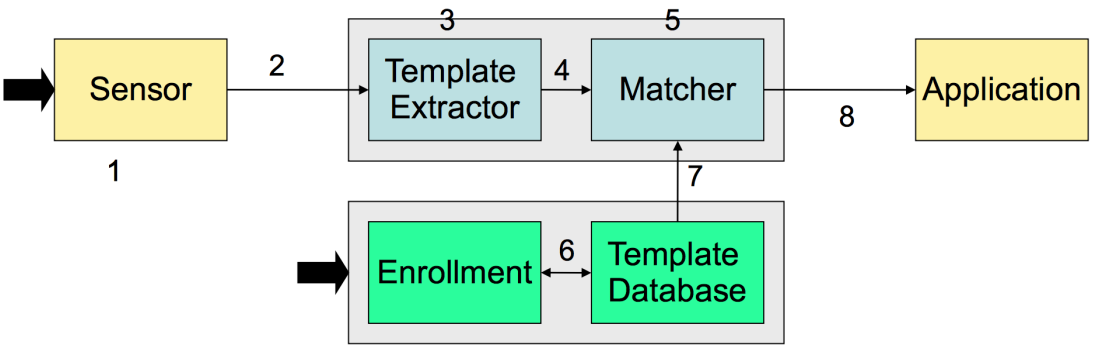

Figure 9.1: possible basic points of attack that plague biometric authentification systems

9.2 Attacking a Biometric System

9.2.1 Attacking Biometric Identifiers

- Attacking at sensor level
- Coercive attack
  - the true biometric is presented, but in an unauthorized manner
  - a genuine user is forced by an attacker to identify him or herself to an authentification system -> the system should detect coercion instances reliably without endangering lives (stress analysis, guards, video recording)
  - the correct biometric is presented after physical removal from the rightful owner -> the system should detect liveness (movement of iris, electrical activity, temperature, pulse in fingers)
- Impersonation attack (type 1 attacks)
  - an unauthorized individual changes his or her biometrics to appear like an authorized one: Voice and face are the most easily attacked, fake fingerprints or even fingers have been reported
  - changes one's apperance to cause a false negative error in screening systems: disguises or plastic surgeries
  - combination of multiple biometrics makes replications more difficult, specially when synchronization is analyzed (works well for the first case - no suggestion for the second)
  - type 1 attacks - fingerprints
    - make a legal use cooperate
    - using latent fingerprints
    - artificial fingerprints
    - gelatine print (gummy bear attack)
    - silicon print
    - and of course... chopping off ones finger
  - type 1 attacks - iris
    - porselain eye
    - photo of an eye
    - colored contact lens
    - print iris image onto colored contact lens
  - type 1 attacks - face
    - photo or video of person
    - mask

- Replay attack
  - a recording of true data that is presented to the sensor -> prompt random text to be rea, detect tri-dimensionality or require change of expression

9.2.2 FRONT-END ATTACKS (EXTRACTOR AND MATCHER LEVEL)

- replay attack (before template extractor)
  - a recording of true data is transmitted to extractor
  - easier than attacking the sensor
  - digital encryption and time-stamping can protect against these attacks
- electronic impersonation (before template extractor)
  - injection of an image created artificially from extracted features
  - example: an image of an artificial fingerprint created from minutia captured from a card
  - no defense suggested
- trojan horse (on template extractor)
  - extracted features are replaced (assuming the representation is know)
  - the extractor would produce a pre-selected feature set at some given time or under some condition
  - no defense suggested
- communication (between template extractor and matcher)
  - attacks during transmission to remote matcher
  - specially dangerous in remote matchers
  - no defense suggested
- trojan horse (on matcher)
  - match decision is manipulated
  - example: a hacker could replace the biometric library on a computer with a library that always declares a true match for a particular person
  - no defense suggested

9.2.3 CIRCUMVENTION (BETWEEN MATCHER AND APPLICATION)

- overriding the matcher's output
- collusion
  - some operators have super-user status, which allows them to bypass the authentification process

  - attackers can gain super-user status by stealing this status or agreement with operator

- covert acquisition
  - biometric stolen without the user knowledge, but just parametric data used (difference from impersonation)

- denial
  - an authentic user be denied by the system (false rejection)
  - not considered fraud because no unauthorized access was granted
  - but it disrupts the functioning of the system

9.2.4 BACK-END ATTACKS

- all seen so far (enrollment): enrollment has all the stages above
- communication attack (between template database and matcher): attacks during transmission between matcher and central or distributed database
- communication attack (between enrollment and template database): attacks during transmission from enrollment stage to central or distributed database
- viruses, trojans on application
- hacker's attack: modification or deletion of registers and gathering of information

9.2.5 OTHER ATTACKS

- password systems are vulnerable to brute force attacks
- the number of characters is proportional to the bit-strength of password
- biometrics: equivalent notion of bit-strength, called instrinsic error rate
- hill climbing
  - repeatedly submit biometric data to an algorithm with slight differences, and preserve modifications that result in an improved score
  - can be prevented by limiting the number of trials or giving out only yes/no matches
- swamping
  - similar to brute force attack, exploiting weakness in the algorithm to obtain a match for incorrect data
  - example: fingerprints - submit a print with hundreds of minutiae in the hope that least threshold of number of them will match the stored template
  - can be prevented by normalizing the number of minutiae

- piggy-back
  - an unauthorized user gains access through simultaneous entry with a legitimate user (coercion, tailgating)
- illegitimate enrollment
  - somehow an attacker is enrolled (collusion, forgery)

9.3 Biometric System Counter attacks

9.3.1 Combining Smartcards and Biometrics

- biometrics - reliable authentication
- smartcards - store biometrics and other data
- suggestion: valid enrolled biometrics + valid card
- benefits
  - authentication is done locally - cuts down on communication with database
  - the information never elaves the card - secure by design
  - attacks occur locally and are treated locally
  - keeps privacy

9.3.2 Challenge-Response Protocol

- dynamic authentication - prevents mainly replay attacks
- the system issues a challenge to the user, who must respond appropriately (prompted text - increases the difficulty of recorded biometrics use)
- it will demand more sophisticated attacks and block the casual ones
- extension, e.g. number projected in the retina, that must be typed

9.3.3 Cancellable Biometrics

- “Cancelable biometrics refers to the intentional and systematically repeatable distortion of biometric features in order to protect sensitive user-specific data.”
- once a biometric identifier is somehow compromised, the identifier is compromised forever
- privacy: a hacked system can give out user’s information (medical history and susceptibility)
- proscription: biometric information should not be used for any other purpose than its intended use

- concerns: not an extra bit of information should be collected, data integrity and data confidentially are two important issues, cross-matching: matching against law enforcement databases, biometric cannot change (issue a new credit care number, etc.)
- cancellable biometrics is a technique that alleviate some of these concerns: biometrics are distorted by some non-invertible transform, if one representation is compromised, another one can be generated
- signal domain distortions: distortion of the raw biometric signal (morphed fingerprint, split voice signal and scamble pieces)
- feature domain distortions: distortion of preprocessed biometric signal (fingerprint minutiae)
- signal compression: the signal temporarily loses its characteristics
- encryption: secure transmission (signal is restored after it)
- cancellable biometrics: signal loses definitely its characteristic, it's desirable that the distorted signal is impossible to be restored

9.3.4 ANTI SPOOFING

- Spoofing: "The process of defeating a biometric system through the introduction of fake biometric samples."
- Artificially created biometrics
  - lifted latent fingerprints
  - artificial fingers
  - image of a face or iris
  - high quality voice recordings
  - worst case - dismembered fingers
- Obfuscation - hiding your identity (spoofing - posing as another individual)
  - negative identification applications
  - may form new identity for positive identification
  - mutilation of fingerprint
  - texture-contact lens to hide iris pattern
  - theatre makeup/putty to change facial characteristics
- application-specific risk assessment
  - what is the role of biometrics in my application (is it needed)
  - does it improve upon former methods of identity management

  - what is the impact of spoofing vulnerability
  - what is the public perception of spoofing vulnerability
- ways to mitigate risk
  - multi-factor authentication - password, smart card
  - multi-biometrics - require multiple biometrics
  - liveness detection or anti-spoofing

9.3.5 Liveness Detection

- also termed vitality detection or anti-spoofing
- definition: to determine if the biometric being captured is an actual maesurement from the authorized, live person who is present at the time of capture
- "It is liveness, not secrecy, that counts" - Dorothy Dening
  - fingerprint is not secret
  - cannot reasonably expect it to be absolutely secret
  - therefore, must ensure measurement is of the real biometric and not a replica
  - true for most other biometrics, with some exceptions to be discussed
- typically treated as a two class problem - live or spoof
- rarely do biometric sensors measure liveness, that is, liveness is nt necessary to measure the biometric
- hardware-based requires specialized hardware design, it is integrated with biometric sensor (examples: temperature, pulse, blood pressure, odor, electrocardiogram, multi-spectral imaging, spectroscopy)
- software-based uses information already mesured from biometric sensor, additional processing is needed to make a decision (examples: skin deformation, elasticity, pores, perspiration pattern, power spectrum, noise residues in valleys, combining multiple features)
- liveness inherent to biometric - must be live to measure it, e.g. electrocardiogram