IBM Research

Biometrics Short Course

Andrew Senior & Nalini Ratha

IBM T.J.Watson Research Center
Yorktown Heights, NY, USA
aws@us.ibm.com, ratha@us.ibm.com
http://www.research.ibm.com/people/a/aws

Biometrics short course overview

- Face recognition
- Speaker identification
  - Lip motion
- Signature verification
- Standardization
- Multiple biometrics
- Conclusions

Face recognition- overview

- Face recognition
  - Specific workshop this afternoon (B. Moghaddam)
  - Face & Gesture conferences (Seoul May 2004)
  - AMFG 2003 (with ICCV, Nice in October)
- Related domains
- Advantages and obstacles
- Applications
- Face detection
- Methods
  - Eigenfaces
- FERET & FRVT evaluations
- Databases

Face recognition

- Identification or verification from facial appearance
- Typically a face image from a photographic technique
  - Still or video
  - Colour or black & white
    - Non-visible wavelengths- facial thermogram
  - 3D techniques
    - Stereo
    - Structured light

Advantages

- Accepted
- Ubiquitous capture device
- Many legacy uses/databases
- Human verifiable
- Remote & unobtrusive capture- surveillance
- But
  - Sensitive to many variations
  - Easy to fake/forge
  - Not most accurate biometric

Obstacles

- Lighting
- Pose
- Age
- Facial expression
- Facial appearance
  - Glasses
  - Makeup
  - Hairstyle
  - Facial hair
- Need to constrain these factors or have an invariant algorithm

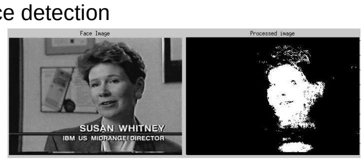
Applications

- Surveillance
- Physical/logical access
- Passive customization
- Large-scale ID- with assistance

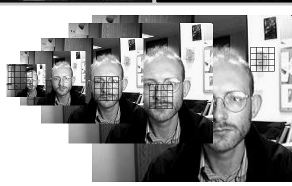
Related domains

- Face detection & tracking
  - Video indexing
  - HCI- gaze detection & focus of attention
- Facial expression recognition
- Lip motion-
  - biometric & Audio-Visual speech / lip reading
- Computer graphics – synthesis & animation
- Close links with psychology & face perception

Face detection



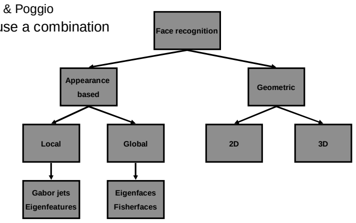
Skin-tone segmentation



Multiscale (pyramid) search with binary detector: MOG, NN, SVM, decision tree Binary classifier trained with Face samples and non-face samples Obtained by active learning, multiple classifiers through boosting

Techniques

- Many different approaches
- Broken into appearance and geometric approaches
  - see Brunelli & Poggio
- Many systems use a combination

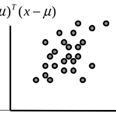


Eigenfaces- a basis for representation

- Eigenfaces are a representation technique for dimensionality reduction and generalization.
- Normalize and select a region of the face in many face images
- Treat each face sub-image as a vector (of intensities) and find moments of vectors across a database:

$$\mu = \frac{1}{N}\sum x$$

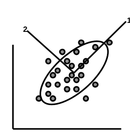
$$\sigma^2 = \frac{1}{N-1}\sum (x-\mu)^T(x-\mu)$$



Mean and Eigenfaces from eye/nose normalized face regions cropped with 41x41px square.

Eigenfaces

- Covariance matrix expresses the within-class variation of the observed faces.
- Calculate the eigenvectors or *principal components* of the matrix:
- An orthogonal coordinate system, ranking the axes in order of decreasing variance
- A basis- can uniquely decompose any vector as the weighted sum of eigenvectors
- Same as *Karhunen-Loève* transform



Eigenfaces for dimension reduction

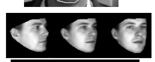
- (*Assume)* most useful information is in the first few principal components (largest variance)
- Represent faces using these components
  - Project into a subspace
- Discard lower components as "noise"
- Compact representation 100,000->~50 dim
- Generalizes well

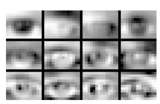
Recognition with Eigenfaces

- Compare two faces using Euclidean distance in the Eigenface subspace.
- Nearest-neighbour for closed-world identification (e.g. FERET task)
- In some cases expression or lighting may generate more variation- so first few components contain information that does not convey identity
  - In this case, these components are discarded.

Extensions to Eigenfaces

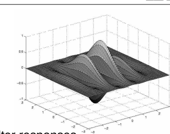
- Better ways to crop & normalize the face regions
  - E.g. "Shape-free" eigenfaces- anchor points to warp texture
  - Lighting normalization
- Combine shape and texture-
  - Active appearance models
    - Shape-free texture subspace
    - Shape subspace
    - Combine subspaces and apply PCA again
- Understand the subspaces
  - Find age/gender/expression subspaces
  - Other subspace choices e.g. ICA
  - Model separate regions of the subspace indpendently
- Apply PCA to features
  - Recognize separately





Mean eye and "eigeneyes"

Local representation- Gabor Jets



- Used in Elastic Bunch Graph methods
  - Von der Malsburg et al.
- Representation of small face regions using Gabor filter responses
  - ~4 scales x 8 orientations
- Feature points on a regular lattice, or chosen to be salient points
  - Features are "self-localizing"
- Feature points compared pairwise
  - Aggregate score gives total similarity]
- Elastic bunch graph involves displacement between features as well as feature comparisons

Handling lighting and pose

- Control the lighting or pose
- Simple normalization (e.g. mean subtraction)
- Capture lighting variability
- Enroll multiple views
- Create a 3D model
  - Use Pose correction
    - E.g. FRVT uses Blanz & Vetter's 3D morphable models as a preprocessing step for a variety of algorithms to "…substantially improve the ability to recognize non-frontal faces."
- Model the lighting- Belhumeur

Extending face recognition

- Most systems acquire faces under controlled lighting and geometry
- HID project seeks to extend that to greater distances
- How much can we improve face recognition by recognition from video vs recognition from still images
  - Choosing 'keyfaces'
  - Sequence of face images is not independent
  - Face images from a sequence may contain multiple views / lighting / expressions

Face Forgery

- Original systems were vulnerable to defeat by a photograph
- “Live” face detection
  - Detect and require movement- prompt for facial gesture
  - Test for 3-dimensionality, face-like structure
  - IR imagery to verify in unforged wavelengths, or detect a disguise
- Simultaneously check other biometrics
- Other views

FERET & FRVT

- Evaluation of face recognition performance, run by Jonathan Phillips of Army research labs/NIST
- FERET tests:
- Tests of commercial and academic face recognition systems under controlled, uniform circumstances
- FRVT 2000 & 2002
  - Temporal (> 18 months after enrolment)
  - Lighting
  - Expression
  - Resolution
  - Compression
  - Gallery up to 37,437 individuals (up to 75% correct identification)
- Test of 3D normalization system – improves recognition in 2D systems
- 50% drop in error rates on comparable tasks between 2000 & 2002

Face recognition for watchlists

- e.g. spotting terrorists at airports
- Performance not yet good enough for automated systems
  - Far too many false alarms:
  - FRVT figures (clean data):
    - 25% FRR at FAR of 1% (3 false alarms per 300-passenger plane),
- Good enough to help human operators, direct attention to interesting cases.

Face Databases

- Cheap digital imaging & growth of interest has led to an explosion of databases
  - ORL/AT&T database- often used, but limited, monochrome, high recognition rates
  - CMU PIE 68 people x 13 *Pose* x 4 *E*xpression x 43 *Illumination*: 41368 images
  - XM2VTS- video & speech from 295 people
  - Yale- lighting variation
  - FERET
  - Manchester
  - MIT
  - CMU – face detection
  - Weizman

Speaker Verification

- Identifying a person through characteristics of their speech
- Something that people do naturally all the time-
  - Even in difficult conditions of noise, channel (telephone)
  - Generally using some context

Applications

- Particularly in telephony
  - Verifying telephone transactions
  - Home banking, mail order, travel agent
- Speaker labelling
  - Surveillance e.g. Switchboard / EARS task
  - Video indexing
  - Teleconferencing/ meeting mining
- Forensic
  - Is this Osama Bin Laden?
- Computer log-on, physical access

Advantages

- Lowest cost biometric device
- Device is ubiquitous- any telephone, most computers
- Contact-free
- Well accepted
- Incorporates a revocable knowledge component
- Acquisition can be passive, even covert

Disadvantages

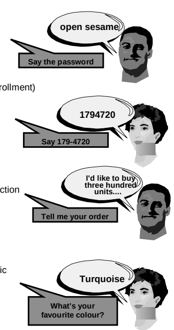
- Easy to record and replay speech
- Subject to noise & channel

Obstacles

- Subject to background noise, crosstalk
- Subject to channel variations:
  - Microphone
  - Telephone bandwidth
  - Echo
  - Cellular codec
- Individual changes
  - Aging
  - Health
  - Mood
- Defeating replay attacks
  - Recording & replaying a fixed text password is very easy

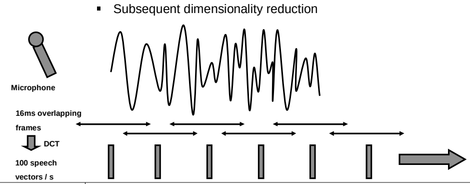
Domains

- Fixed text
  - User says name, a recorded password etc
  - Simple matching techniques can be used
  - Simply attacked with replay attacks
  - Password text is verified and can be revoked (requires reenrollment)
- Prompted text
  - System generates a piece of text that user reads
  - More complex modelling
  - Replay attacks much harder
- Text independent
  - "Passive" approach- verified during some other voice transaction
  - Language independent
  - Use arbitrarily long speech segment to increase confidence
- Conversational biometrics
  - Text independent speech with random questions
  - Simultaneously testing for knowledge as well as the biometric



Features

- Signal sampled and digitized (PCM)
- Most people use standard features as used for speech recognition that encode frequency distribution of sound energy in short sections (frames) of speech
  - Linear Prediction Coefficients Cepstra
  - MFCC (Mel frequency cepstral coefficients)
  - Time variation with delta and delta-delta
  - Subsequent dimensionality reduction

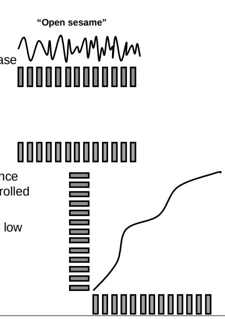


Methods

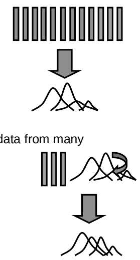
- Train a (probabilistic) model for the features seen a training
  - Train one model per speaker
- Find the likelihood of test data given each model
  - Accept claimed model if score greater than threshold (verification)
  - Choose maximum likelihood model (closed world identification)
  - Accept max if over threshold (open world identification)
- Model might be a single Gaussian, a VQ (Vector Quantization) codebook, a HMM sub-phone speech recognition system

Fixed text by Dynamic Time Warping

- **Train**
  - Record speaker uttering passphrase
  - Store sequence of features
- **Test**
  - Find features of test utterance
  - Find minimum accumulated distance alignment of test features with enrolled features (Viterbi search)
  - Accept if accumulated distance is low



Text independent GMM (IBM 2002 NIST Cellular evaluation)



- **Train Universal Background Model**
  - Train Gaussian Mixture model by clustering on training data from many speakers
- **Create speaker models by adapting UBM**
  - MAP adaptation on one speaker's data
- **Test**
  - Find likelihood ratio of test data for speaker model vs UBM
  - Normalize scores by using scores of other models tested under same conditions

Cohorts / Universal background model

- Cohorts/UBM provide a way of normalizing signal to eliminate the effects of person-independent variation
- Compare one speaker with a group of known similar speakers, equally affected by conditions.

NIST Evaluations

- Recent evaluations on large populations by NIST
- 2000:
  - 110-130s enrollment
  - One speaker detection (15-45s)
  - Two speaker detection (1min conversation)
  - Target speaker tracking (give time info, 10min, 10 speaker)
  - Speaker segmentation (unknown speakers)
  - Spanish
- 2002
  - Two of four tasks on cellular data

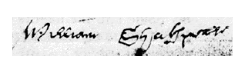
NIST Evaluations

- Weighted cost function:
  - $C_{total} = \sum C_{Error|Cond} P_{Error|Cond} P_{Cond}$
- e.g. (One speaker detection task)
  - $C_{Det} = C_{Miss} \times P_{Miss|Target} \times P_{Target} + C_{FalseAlarm} \times P_{FalseAlarm|NonTarget} \times (1-P_{Target})$
  - With $C_{Miss} = 10$, $C_{FalseAlarm} = 1$, $P_{Target} = 0.01$

Lip motion verification

- Identify a person by characteristic lip motions
- Commercial product by BioID-
  - Combined with speakerID and face recognition
  - Natural, practical combination
  - Makes forgery very difficult
- Database XM2VTS for testing
- Close ties with audio-visual speech recognition, new work on detecting synchronization of lips and audio [Nock, Darrell,…]

Signature Verification



- Verify a person by their signature
  - Equivalent to fixed text speaker ID
- Writer identification
  - Corresponds to text-independent speaker ID
  - Little researched

Signature acquisition

- Offline
  - Scan document
- On-line
  - Capture with an XY (+`pen-up') tablet
    - –Computer graphics tablet
    - –POS terminal (Sears etc.)
    - –Handheld device (PDA, FedEx etc.)
  - Acquire XY + pen angle, pen pressure with specialist tablet
    - –Much harder to forge
  - Use an instrumented pen without a tablet
  - Track pen tip through vision [Munich & Perona98,99]

Applications

- Most commonly captured biometric
  - Cheques
  - Credit card authorizations
  - Shipping company delivery record
  - Driving license, passport etc.
- Often not verified unless a transaction is queried

Verification

- Tablet samples X,Y,p,θ,ψ uniformly in time (e.g. 100Hz)
  - Resample uniformly in space or at salient (e.g. turning) points
  - Transform features- smoothing, dx/dt, curvature.
  - Model user's vectors over time as for speech
- Estimate global characteristics
  - Number of turning points, moments, aspect ratio, etc.

Signature forgery

- Signature could be considered to be a “knowledge”-based authentication method
  - A forger can learn and reproduce a signature
  - Could be revoked and changed
- Only domain with significant emphasis on forgery
  - Zero-effort
  - Home-improved
  - “over the shoulder”
  - Professional

Current & Future Advances in Biometrics

- New biometrics being devised
- Better sensor technologies
  - costs dramatically reduced (quantity helps)
- More research: e.g. DARPA HID program
- More robust acquisition & better handling of poor data
  - Distortion handling
- Integration with other security & authentication methods; coherent design
  - No longer just retrofitted
  - Better frameworks - BioAPI, UVM, CDSA, PAM
  - Cancelable biometrics
- More use studies & practical experience
- Combinations of biometrics

Standardization and consolidation

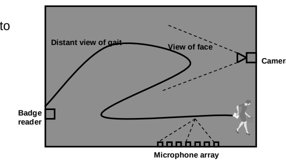
- A number of standards efforts in biometrics technology united to form BioAPI
- Uniform standards
  - Common biometric file formats
  - Interoperability, Performance and Assurance Working Group
- Related standards in computer security, PKI, smartcards.
- BioAPI (on NCITS fast track to become an ANSI standard)
- Intel CDSA/HRS working with BioAPI
- BAPI (forerunner of BioAPI) chosen by Microsoft
- ANSI X9F4 (financial services)
  - X9.84 Biometric Information Management and Security
- SVAPI
- AAMVA DL/ID 2000 minutia exchange standard
- NIST Common Biometric Exchange File Format

Multiple biometrics

- System takes evidence from multiple sensors
  - e.g. BioID: Face, Voice, Lip Motion
  - Very hard to forge all modes simultaneously
- Different modes for:
  - different situations
    - face & voice for ATM & home banking
  - different people
    - People with poor/no fingerprints, no eyes
    - People the system decides are “goats” for one biometric
  - Exploit *continuity of identity* with tracking to integrate ID information over time

Continuity of identity

- Track a person over time
- Acquire multiple biometrics during that period
- *Continuity of identity* allows us to integrate all the identity information together
  - Multiple views of a face
  - Different biometrics
  - Other authentication methods
    - Badge
    - Password
- Can be used for automatic enrollment



Conclusions

- Biometrics is an essential mode of authentication
- A lively area of pattern recognition research
  - No “completely solved” biometric
- Many commercial applications
- Deployment in real world happening slowly
- No single “Best” biometric
- Accuracy is important – under fair conditions
  - Measuring accuracy is a science in itself
  - Accuracy only measures “accidental” errors not forgery/attacks
- Many issues besides accuracy are involved in deploying a successful system

- Slides will be available on http://www.research.ibm.com/people/a/aws
- See our book
  - “Guide to Biometrics” (forthcoming from Springer)