Public Domain Biometric Similarity Scores

And five reasons to get them

Patrick Grother
NIST
22 September 2004

Similarity Scores

	User sample	Impostor sample
Enrolled sample		
	0.84	0.22

Background

- NIST has c. 15M privacy-protected images it cannot release.
- NIST has conducted very large public (and private) tests.
- Developers, integrators and academics have limited imagery for research and development.
- NIST has 10^{12} sample comparison similarity scores
- Scores are useful beyond just performance measurement!
- Score-level fusion of multiple images and/or technologies is a potent means of improving recognition performance.
- System integrators most readily implement fusion at the score level.
- Score level fusion is subordinate to feature and image level fusion.

www.nist.gov/biometricscores

- What is it?
  - Downloadable face and fingerprint similarity scores generated during offline recognition tests
- Why? To advance techniques for:
  - Multimodal, Multi-sample, Multi-algorithm Fusion
  - Prediction of how systems scale to very large populations
  - Considering aspects of whether systems are vulnerable (zoo)
  - Confidence interval estimation
  - Extreme value statistics
  - Anything else that persons find worthwhile

Application I : Fusion

Seek more performance by aggregating information:

- Multi-modal:
  - Face + 4 “slap” Fingers
  - Face (vis.) + Face (IR)
- Multi-algorithmic:
  - Finger (by Co. X) + Finger (by Co. Y)
- Multi-sample:
  - Five faces (up, down, left, right, frontal)
  - Two irides (left + right)
- Repeated sample:
  - Face (now) + Face (then) + Face (way back then)

Application I : Fusion

Level	Image	Feature / Template	Score	Decision
Exam-ple	Face, IR+Visible But how to fuse face + finger?	Concatenation of KL coefficients in PCA face	F(finger) + G(face)	Voting: Face + Finger say yes, iris no
OK	Theoretically better at earlier (higher dimensional) stages			
Avail-ability	Populations Public: 10^3 Gov: 10^8 Industry: $10^?$	None: proprietary or secret templates, features are not retained,	Now!	Not natively, but by derivation from scores
Privacy	Major (obvious) problems	Possible problems. 1. PCA recon. from KL coefficients 2. Minutia templates	None	None
Why / Why not	Algorithmically immature(?)	Potent, easier within one mode, or in vector spaces. But more difficult integration	Relatively easy integration (given the statistics) of COTS tech.	Very easy integration. Suitable when sample from one mode is missing.

Application I : Fusion (score level)

Normalization: Put scores on same (or at least fuse-able) scales:

- $s' = (s - s_{min}) / (s_{max} - s_{min})$
- $s' = (s - s_{mean}) / s_{stdv}$
- $s' = M(s)$
- $s' = N(s)$
- $s' = \log m(s) / n(s)$

Where statistical info. is known apriori or empirically

- M = match score CDF = FRR
- N = non-match score CDF = 1-FAR
- m = match score PDF
- n = non-match score PDF

Fuse: Combine (or classify) two or more scores

- Sum
- Product
- Sum (with weights)
- Product (with powers)
- Min
- Max

Widely researched:

- Kittler, PAMI 1998
- Jain et al. 2001, 2002, 2004
- FRVT Supplemental 2004

Application II : Large Populations

- Seek two quantities
  - FNMR (P, k, t) = fraction of genuine samples whose enrolled match is either below threshold, t, or below rank k when the enrolled population is of size P.
  - FMR(P, t) = fraction of impostor samples matching any of P enrollees at or above threshold, t.

- Measure empirically, or model them:
  - $FMR(P, t) = 1 - [1 - FMR(1,t)]^P = P\ FMR(1, t)$ i.e. linear
  - $FNMR(P, P, t) = FNMR(1, P, t)$ i.e. match score does not depend on other enrollees

- But:
  - Rank causes model to break (somewhat)
  - FNMR(P, 1, t) = declining function of P
  - Models are sensitive to the right tail of the non-match distribution

Application II : Large Populations

- Seek closed-form formulae for identification performance in populations of size P, at threshold t, and rank k.
  - $FNMR(P, k, t) = 1 - \int_{t}^{\infty} I_{N(s)}(P-k, k)\ m(s)\ ds$
  - $FMR(P, t) = 1 - N(t)^{P}$

I = incomplete β func

- Special Case is Cumulative Match Characteristic for closed-set
  - $CMC(P, 1) = \int_{\infty}^{\infty} N(s)^{P-1}\ m(s)\ ds$
- Driven by (empirically derived) statistics:
  - N(s) = non-match CDF
  - m(s) = match PDF
- See:
  - Grother and Phillips “Models of Large Population Recognition Performance” in IEEE’s Proceedings of CVPR 2004
  - Madigan et al. “Extreme value theory applied to document retrieval”, Annals of Stats, 2004.

Application III : Vulnerabilities

- Consider persons with “in-the-tails” samples: twins, lookalikes, and impostors. This goes to vulnerability. Extreme value statistics.

- The Biometric Zoo: are certain people easy to pose as (lambs)? Or apt to gain access (wolves)?
  - Existence of is not proven on large datasets
  - Is this of major importance for security?
  - Can anything about be non-zero effort impostors be learned?
  - Do low quality samples give increased FAR?

Application IV : Confidence Intervals

- How sure are we that the measured performance will hold up. What's the sample variance?
- How sure are we that the measured performance beats a performance requirement?
- How many samples need to be used for this?
- Do transactions produce independent scores?

Application V : etc etc

- Are face and finger truly de-correlated?
  - Can we just multiply false match rates?
  - Will these data sets answer that?
- Secondary uses:
  - clustering of humans on demographics (age, sex etc.) if we're able to provide such metadata
  - Failure mode analyses, do systems fail on certain persons?
- Unintended uses:
  - Hopefully (inevitably?) useful in other ways

Next Steps …

- Download: www.nist.gov/biometricscores
- Advertise:
  - Please socialize this: tell the engineers and the statisticians.
  - Cite us if you publish.
- Give the sponsors feedback on:
  - What you did with the data
  - what (else) you want – this project is direct response to requests from industry and academia.
  - what should be tested – fusion schemes?
  - what should be provided – more scores? metadata?
  - what should be standardized – interchange markups?

Summary

- Face and fingerprint scores from c. 10^4 persons
- www.nist.gov/biometricscores
  - Register
  - Download
  - Cite
  - Feedback
- Check website for more later (possibly)

?

patrick.grother@nist.gov