

Feature-level transforms for privacy enhancement and template protection



Nalini K. Ratha, Ph. D.*
Exploratory Computer Vision Group
IBM T. J. Watson Research Center
Hawthorne, NY 10532
ratha@us.ibm.com

***joint work with members for the computer vision group**

Outline

- Motivation
- Cancelable biometrics
- Registration-based method
- Registration-free method
 - Meta feature based
 - Patch based
- Conclusions

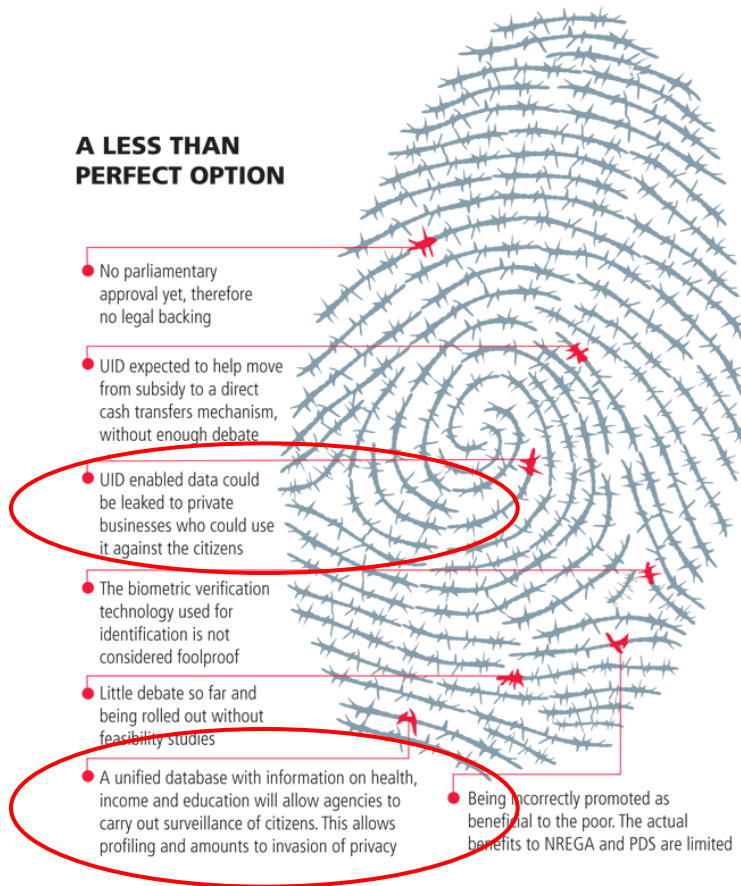
Privacy Issues



- Biometrics is **secure** but not **secret**
- It is the **strongest** form of personally identifying information
- You are giving away a part of yourself.
- **Cross matching** can be used to track individuals without consent (Allows institutions to share data. Allows attack at the weakest point)
- If a biometric is lost or stolen, it is compromised **forever**
- It cannot be **revoked or replaced** (while you can change your phone number or credit card number)

Is there a backwards compatible solution for privacy preserving authentication?

Public concerns with biometrics



- Master Card: The UID Faces Opposition (<http://business.in.com/article/resolution/master-card-the-uid-faces-opposition/21272/1>)
- Setting aside policy questions, the key concerns expressed in the article
 - Template data leakage/compromise
 - Cross matching

From Master Card: The UID Faces Opposition

(<http://business.in.com/article/resolution/master-card-the-uid-faces-opposition/21272/1>)

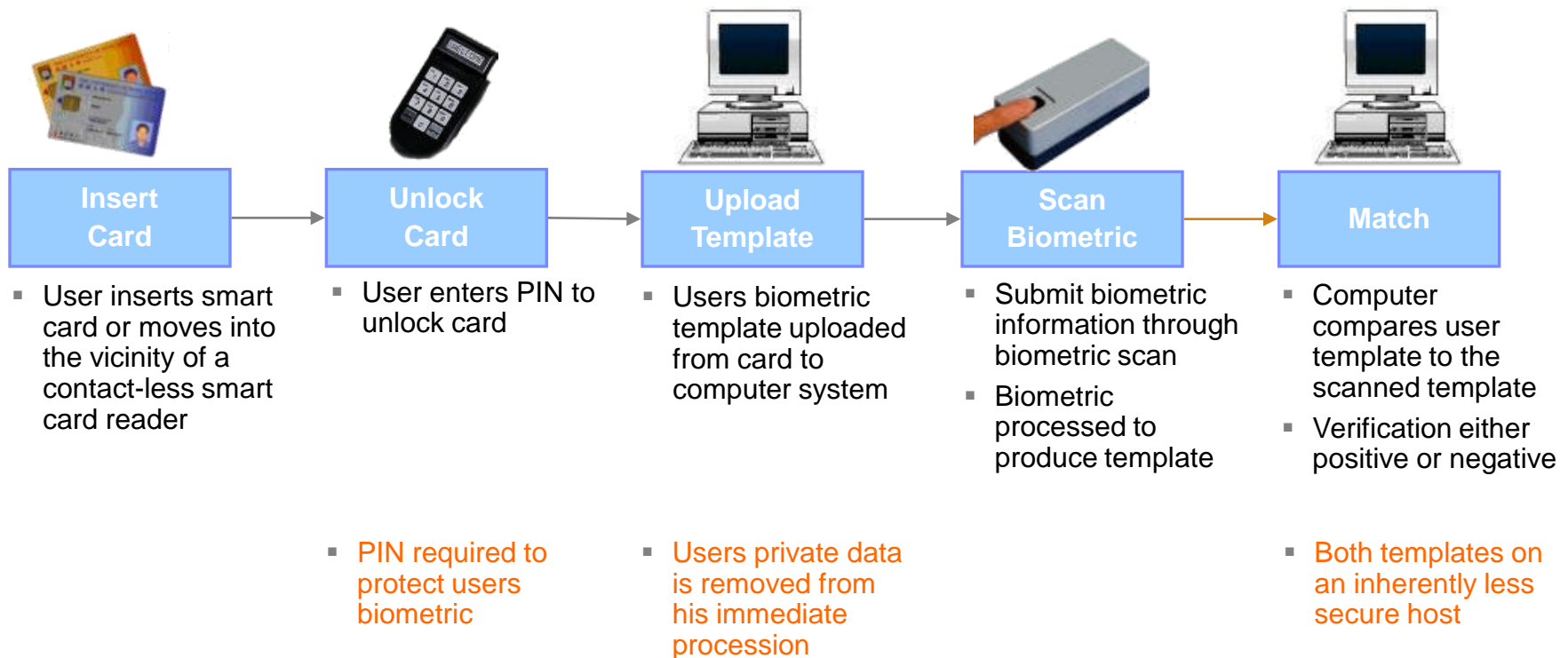
IBM 4758/4764 PCI Cryptographic Coprocessor



- Performs high-speed cryptography
- Provides secure key storage
- ***Tamper-resistant, sensing and responding:***
 - detecting physical attacks (probe, voltage, temperature, radiation)
- ***Programmable***
- Secure configuration and field updates
- US and Canadian government certified:
FIPS 140-1 overall level 4
- Support for Windows, AIX, OS/390, OS/400 (Solaris, Linux)

Smart cards and biometrics

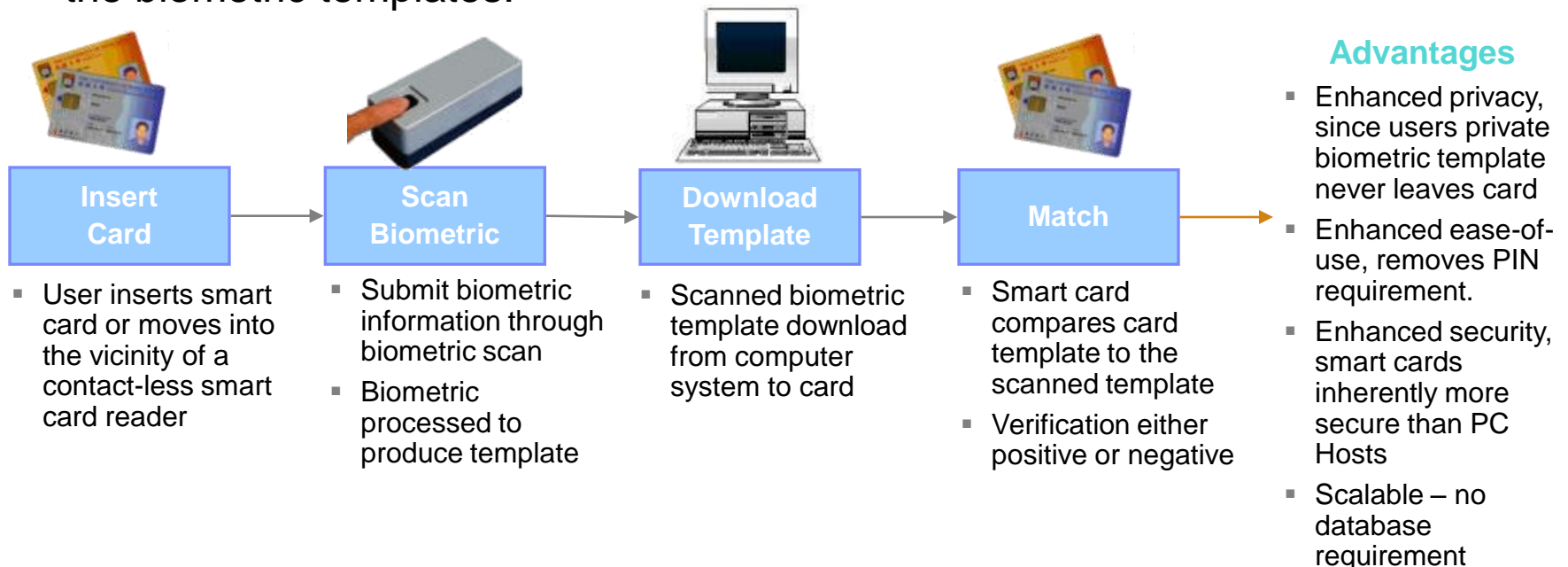
Off-card biometric match process



Smart cards and biometrics

On-card biometric match process

- A secure applet on the java card uses the Java BioAPI to compare and verify the biometric templates.



Cancelable Biometrics

- Intentional **repeatable** distortion
 - Generates a similar signal each time for the same user
- Compromised scenario:
 - a new **distortion** creates a new biometrics
- Comparison scenario:
 - **different** distortions for different accounts
- **Backwards compatibility**
 - Representation is not changed.



© New Yorker Magazine (Charles Addams)

Cancelability requirements of the transform

1. The intrinsic strength (individuality) of the biometric should not be reduced after transformation. (Constraint on FAR)

$$D(x_1, x_2) \geq t \Rightarrow D(T(x_1), T(x_2)) \geq t$$

2. The transformation should be tolerant to intra-user variation (Constraint on FRR)

$$D(x_1, x_2) < t \Rightarrow D(T(x_1), T(x_2)) < t$$

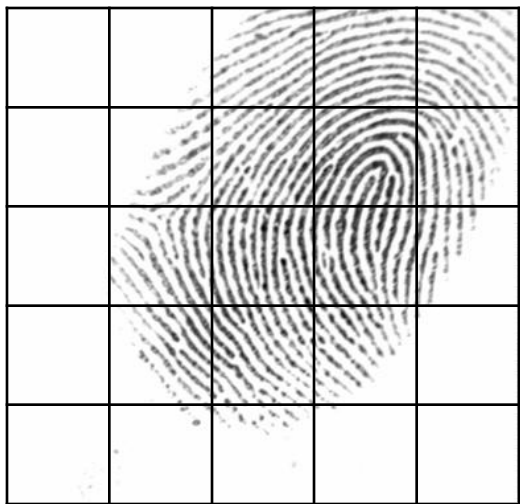
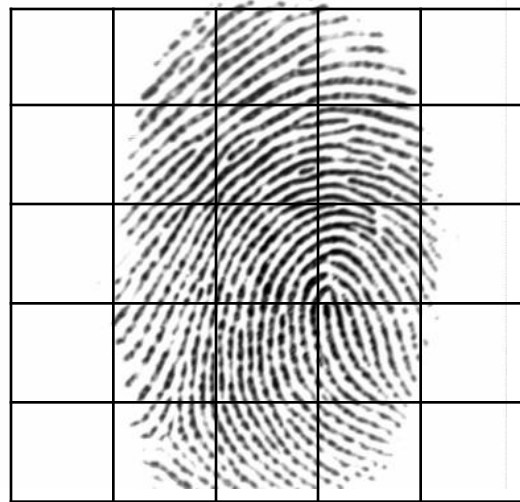
3. The original should not match with the transform,

$$D(x, T(x)) \geq t$$

4. Different transforms of the same user should not match with each other

$$D(T_1(x), T_2(x)) \geq t$$

Challenges



Registration

Transformation

T



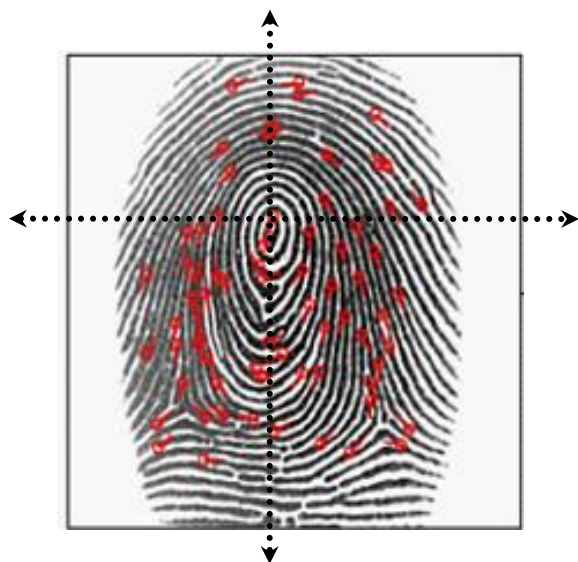
T



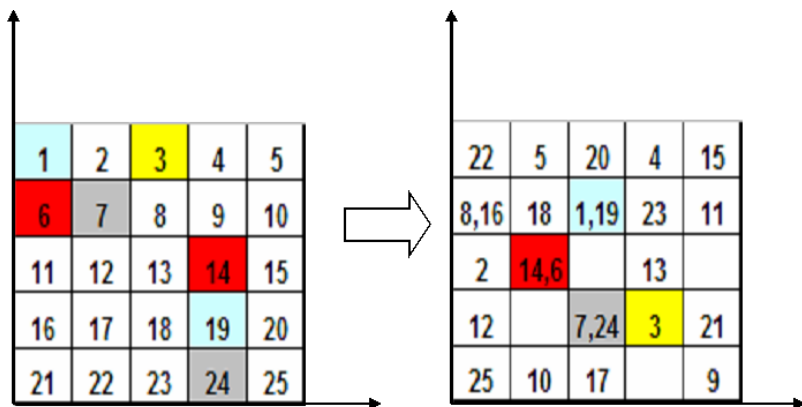
Same?

Intra-user variation

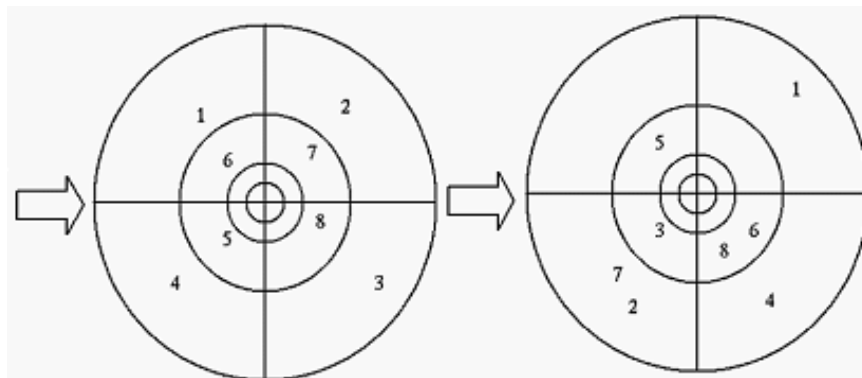
Feature Domain Transformation



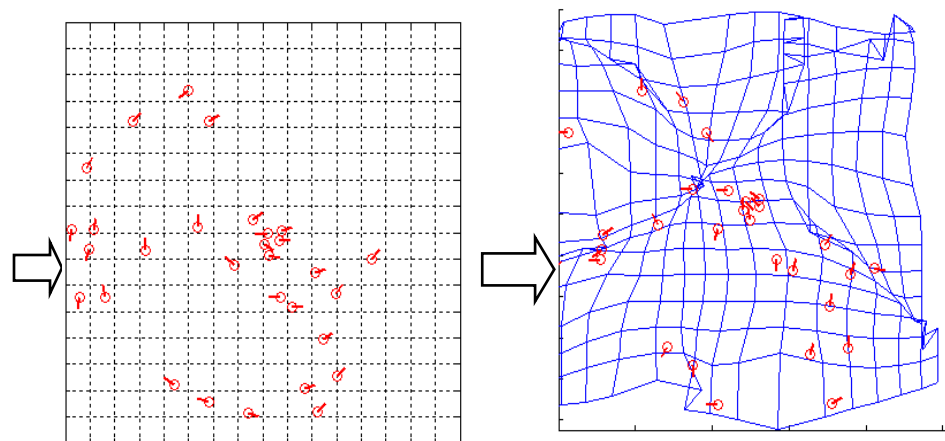
Feature Extraction



Cartesian Transformation



Polar Transformation



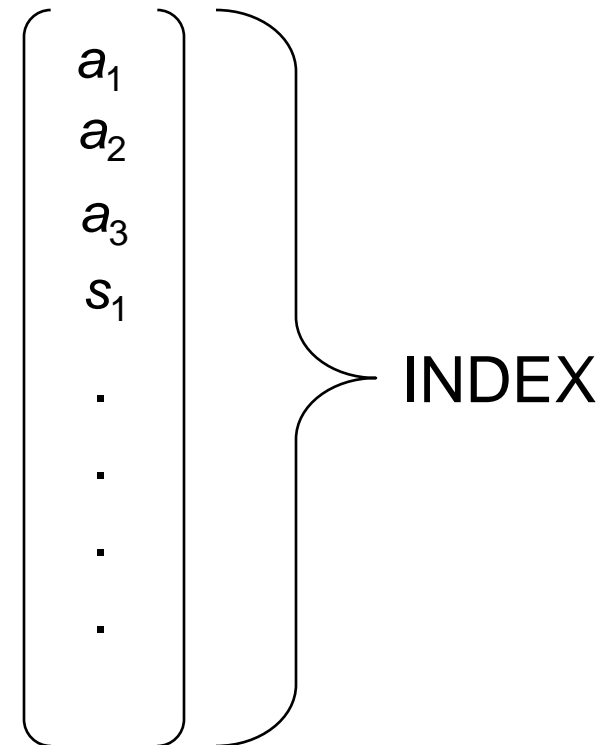
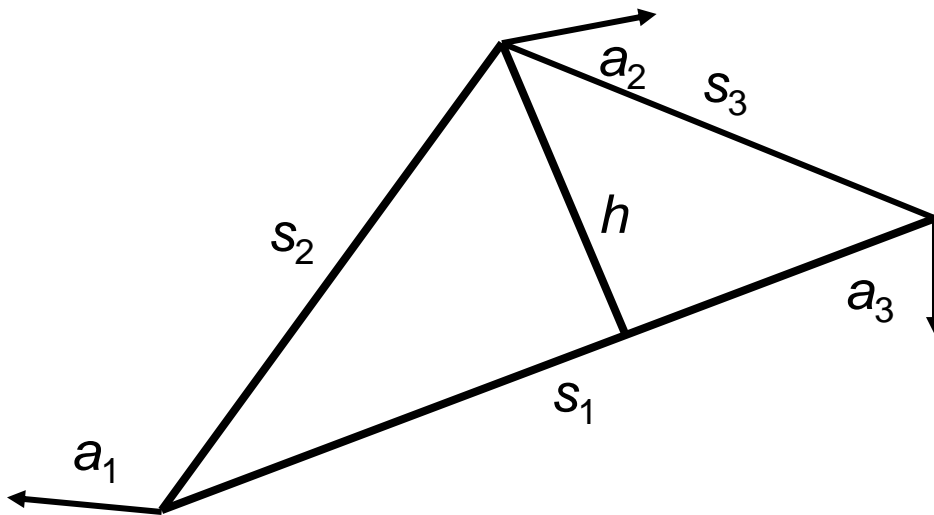
Surface Folding Transformation

Motivation for new constructions

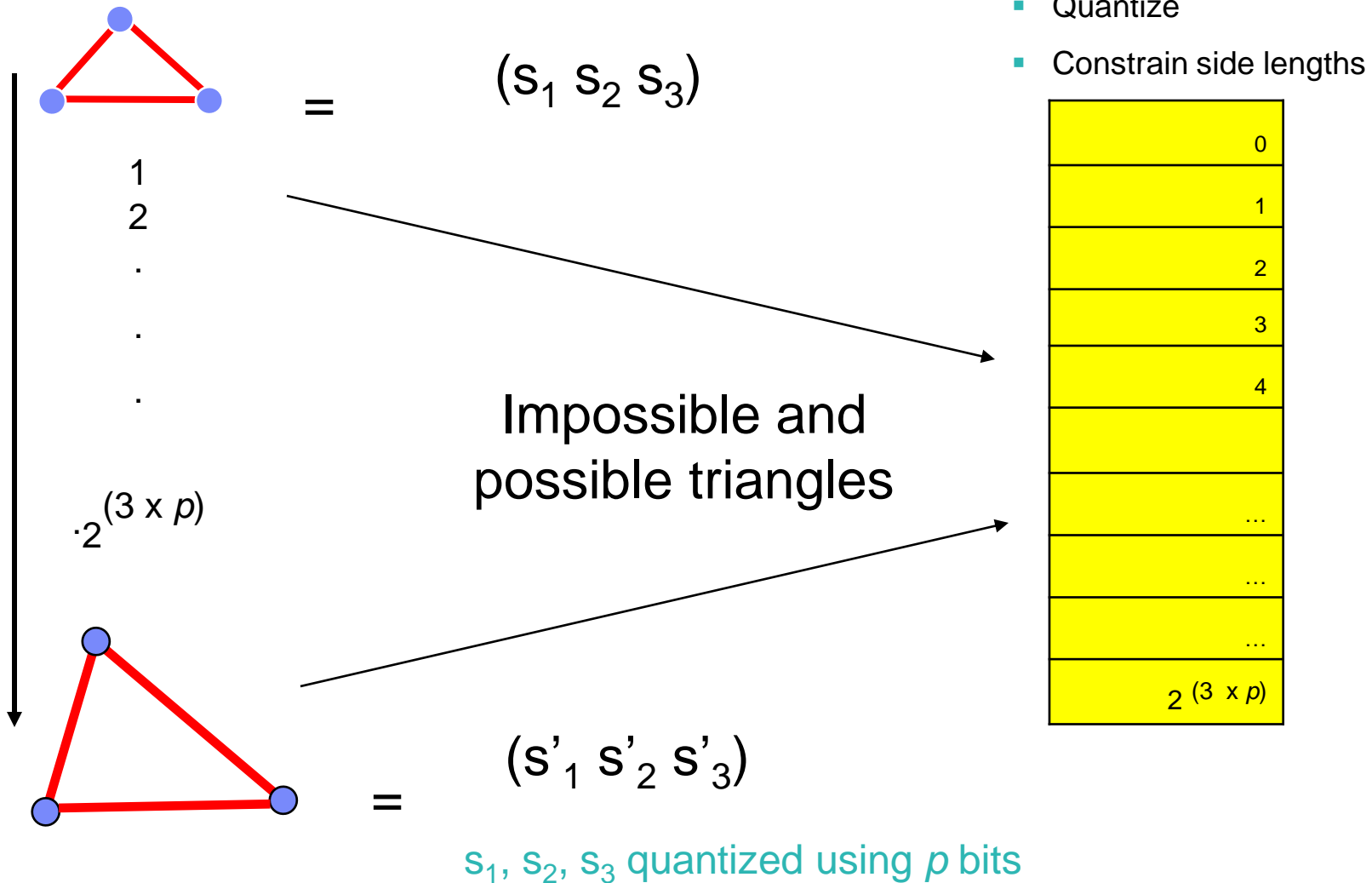
- Existing construction
 - Current construction requires **accurate registration**
 - **Proofs are empirical**-as hard as the fingerprint individuality problem
- Improved construction
 - Can we **avoid registration**?
 - Can we avoid **reduction in accuracy**?
 - Can we have a **provably secure** construction?

Invariant features

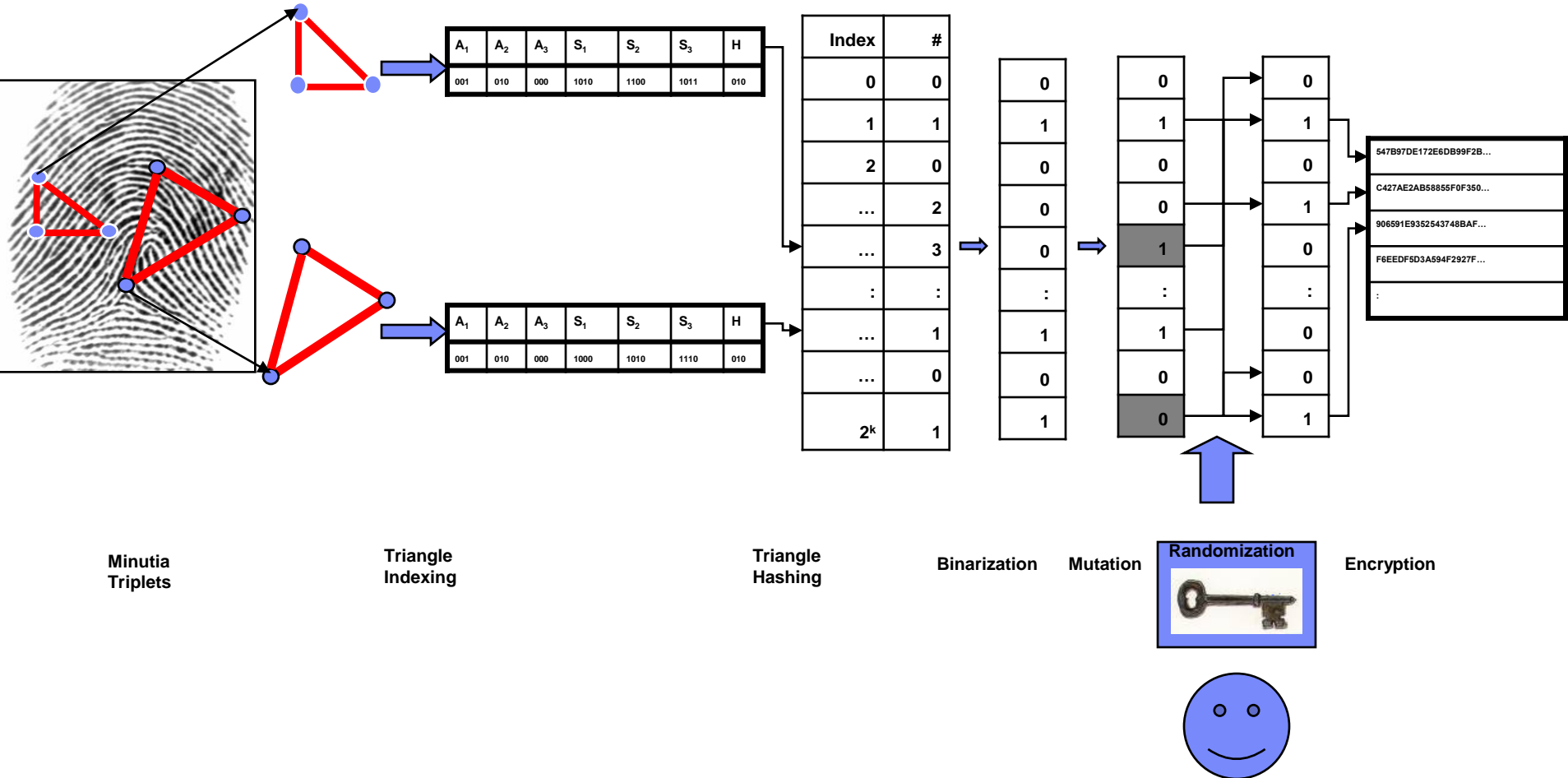
- Independent triangle features
 - The sides
- Dependent triangle feature
 - Height at largest side
- Fingerprint features
 - Minutiae angles with respect to triangle



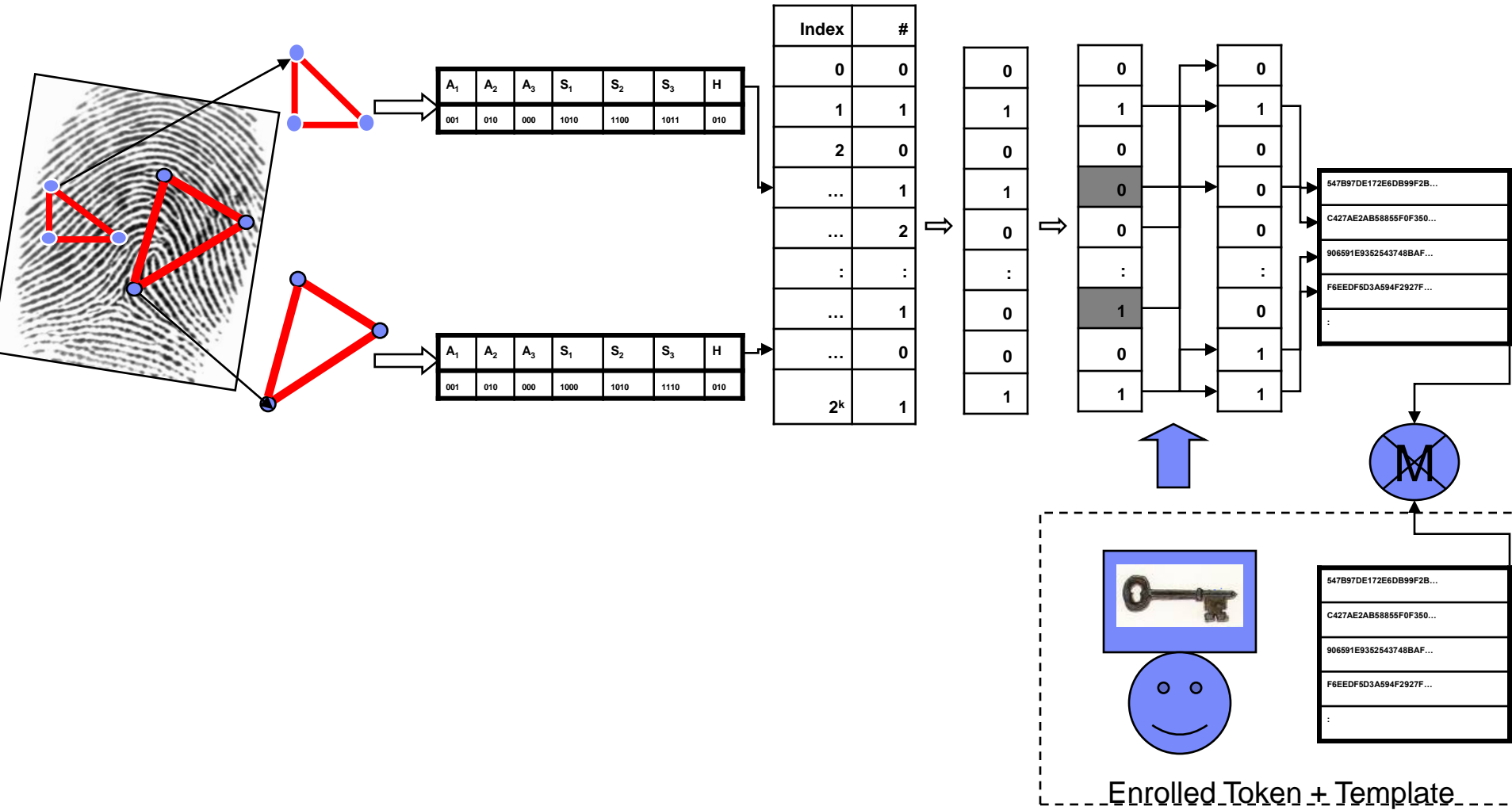
Triangles can be enumerated



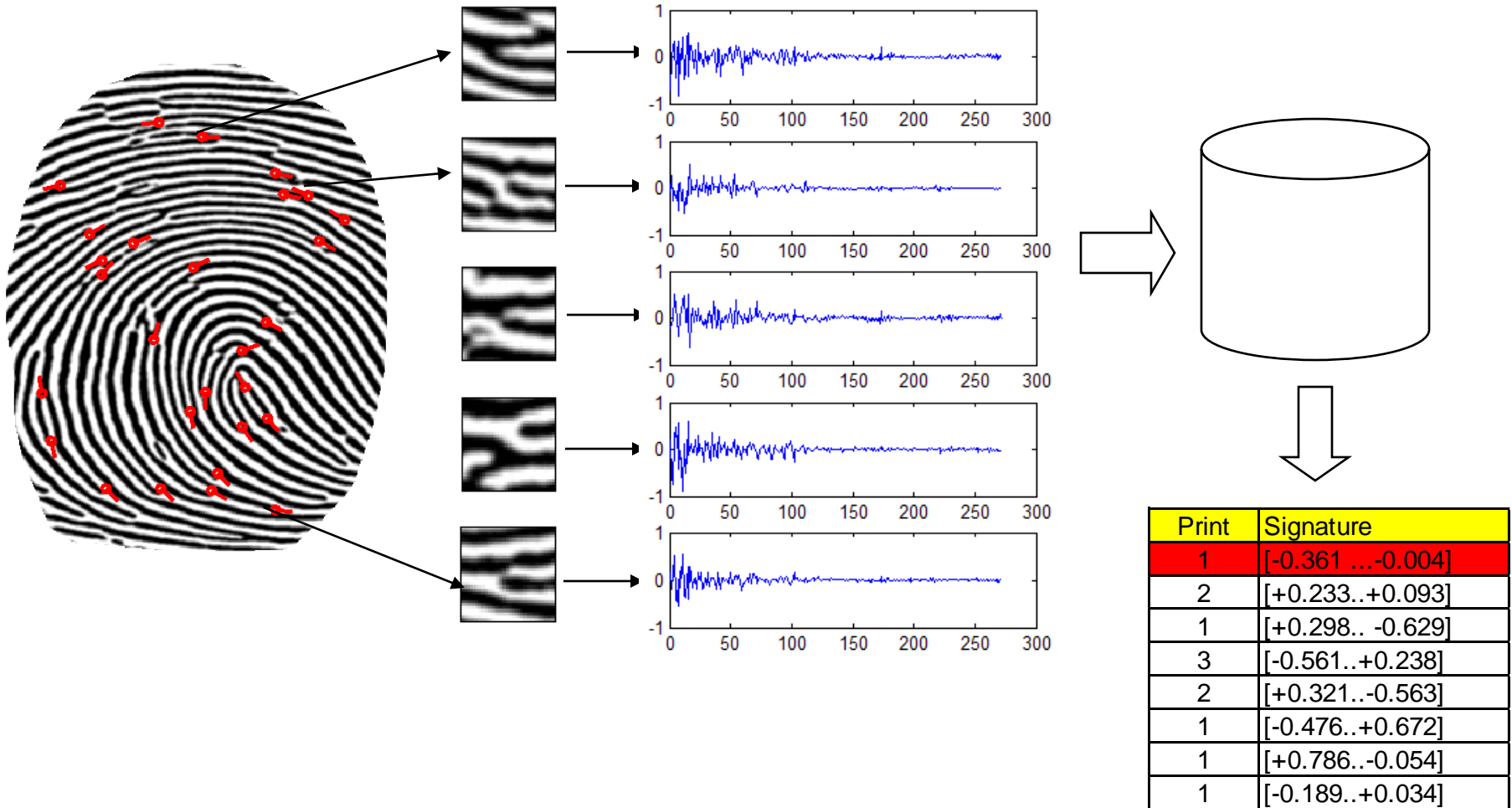
Enrolment



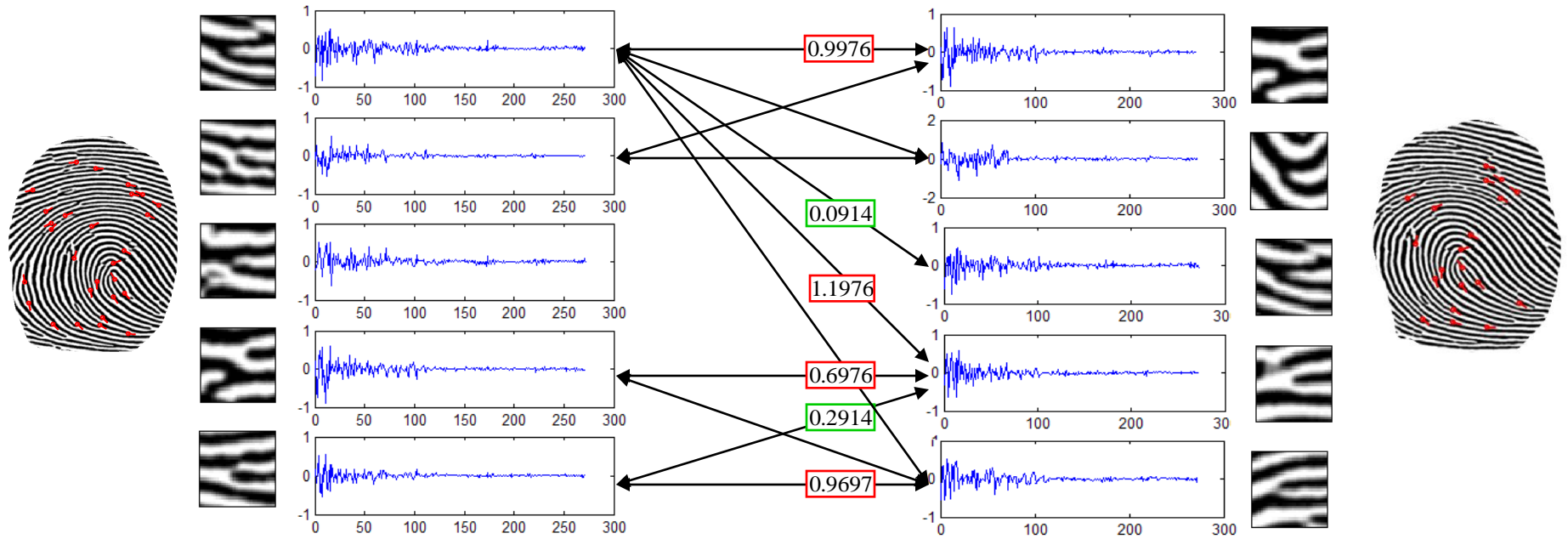
Verification



Patch based features



Verification



Patch signatures: Gabor expansion

$$I(x, y) = \sum_n a_n G_n(x, y)$$

$$f(a_1, a_2, \dots, a_n) = \sum_{(x,y)} \left[I(x, y) - \sum_n a_n G_n(x, y) \right]^2$$

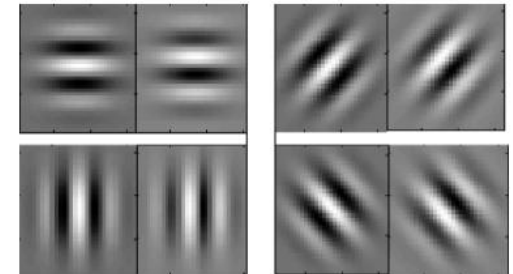
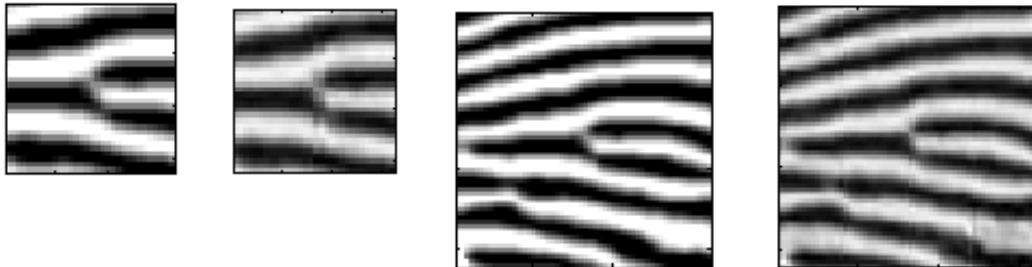
$$\Delta a_n = \sum_{(x,y)} \left(G_n(x, y) \cdot I(x, y) - G_n(x, y) \cdot \sum_k a_k G_k(x, y) \right)$$



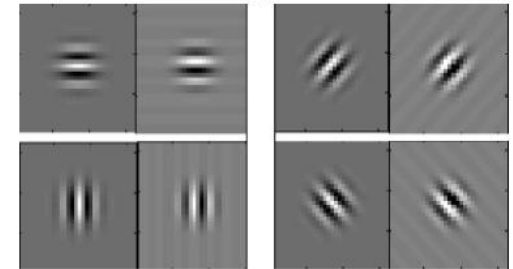
(a)



(b)



(a)



(b)

$$\psi_{mpq\theta}(x, y) = 2^{-m} G(x', y')$$

$$x' = 2^{-m} \lfloor \cos\theta + y \sin\theta \rfloor p$$

$$y' = 2^{-m} \lfloor x \sin\theta + y \cos\theta \rfloor q$$

$$(p, q) \in [0, 2^m, 2^{2m} \dots N]$$

Verification

- How do we get the optimal pairs?
 - The patches are paired by solving a 'bipartite-graph matching' (Hungarian assignment) problem.
- How do we measure the degree of correspondence?

$$\text{Simple count} \quad : \text{score}(x, y) = \begin{cases} 0 & \text{if } d(x, y) > t (= 0.35) \\ 1 & \text{otherwise} \end{cases}$$

$$\text{Log weighting} \quad : \text{score}(x, y) = \begin{cases} -\log(d(x, y)) & \text{if } d(x, y) < t (= 0.5) \\ 0 & \text{otherwise} \end{cases}$$

$$\text{Inverse weighting} \quad : \text{score}(x, y) = \begin{cases} \frac{1}{d(x, y)} & \text{if } d(x, y) < t (= 0.5) \\ 0 & \text{otherwise} \end{cases}$$

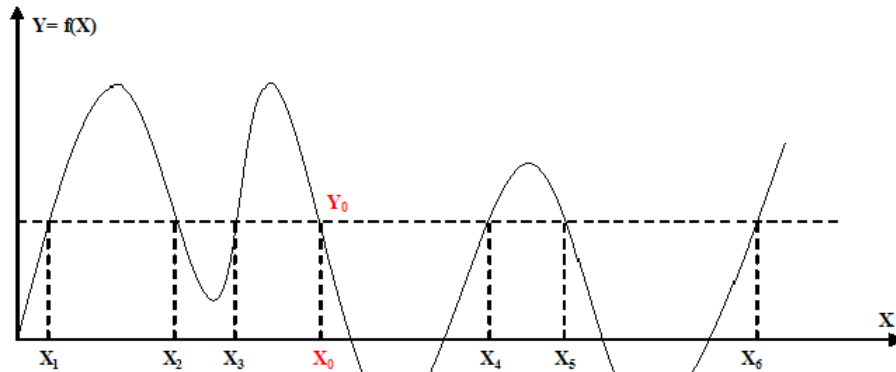
- Combining scores

$$\text{match score}(M, N) = \frac{\left(\sum_{x \in M} s(x, y) \right)^2}{|M| \times |N|}$$

$$M - \text{probe}, N - \text{gallery}, |M| < |N|$$

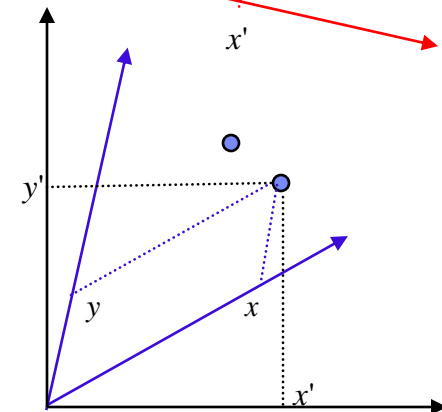
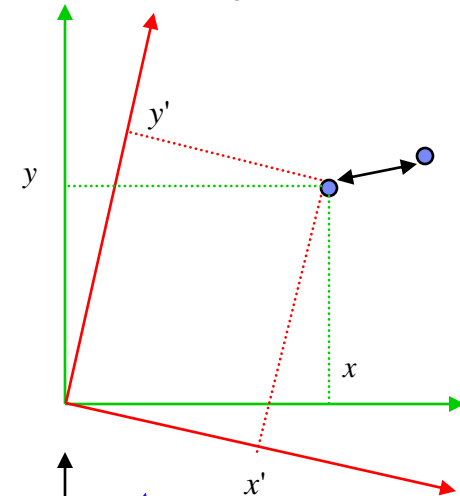
Cancelable methods

- Can we **avoid storing the original patch signatures**?
- Ways to transform/hide the feature vector
 - Encryption - representation too unstable for encryption
 - Polynomial transformation
 - Random projection- **fits well with NDP distance**



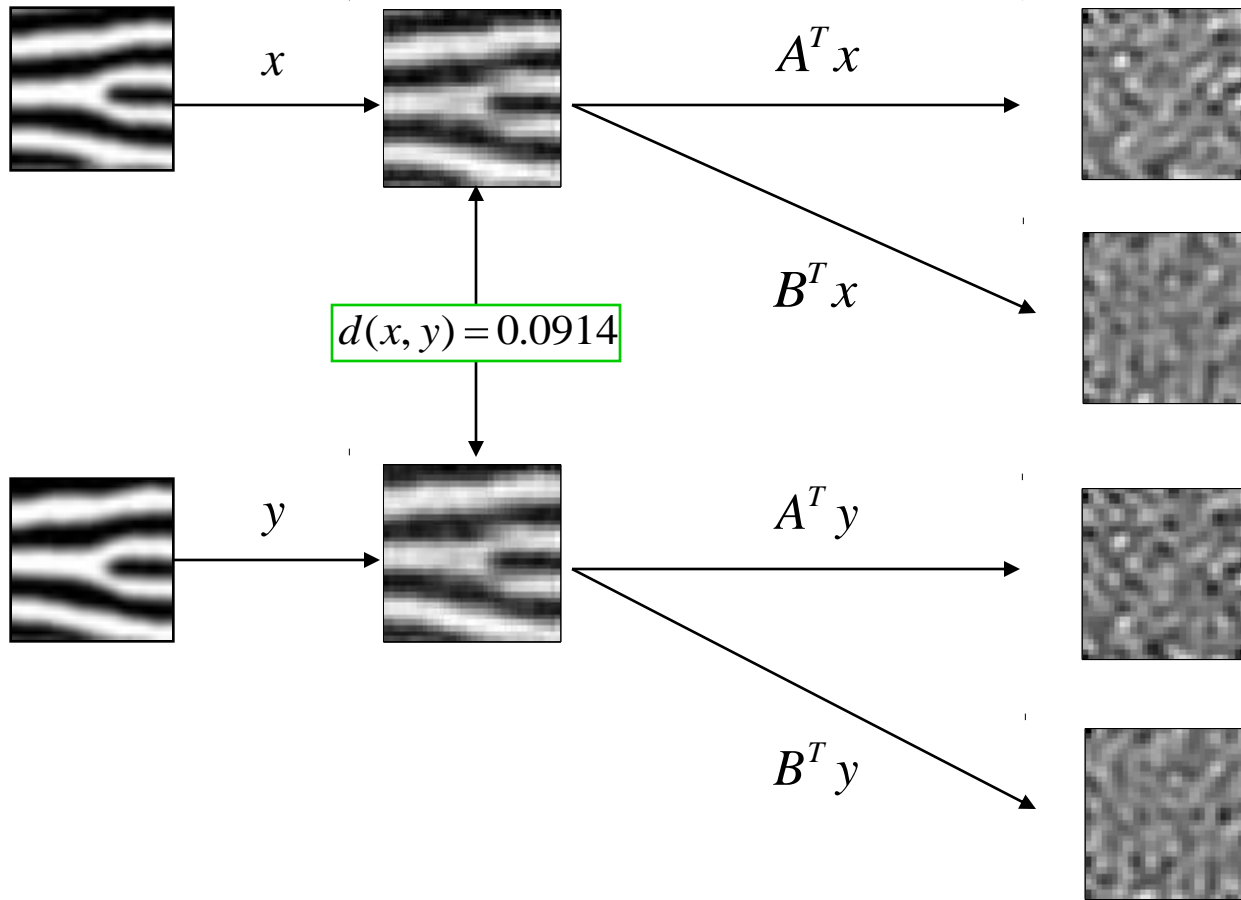
Polynomial transformation

Preferred: Ortho normal projections



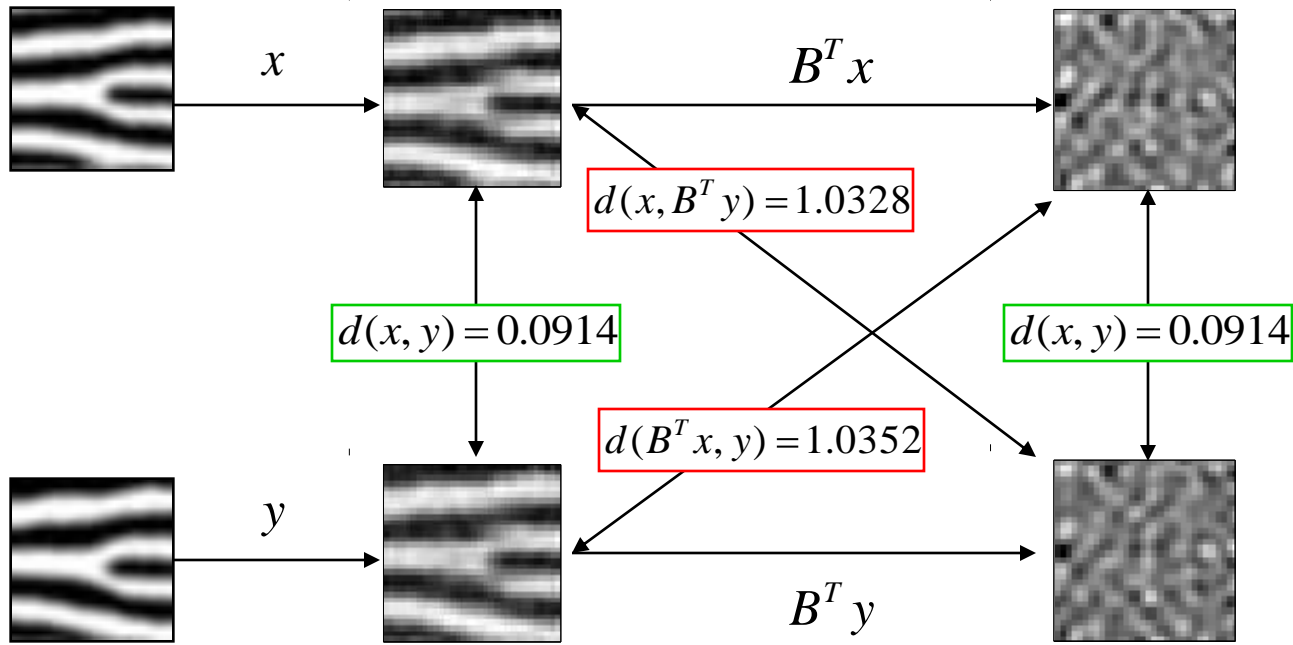
Random Projections

Patch transformation



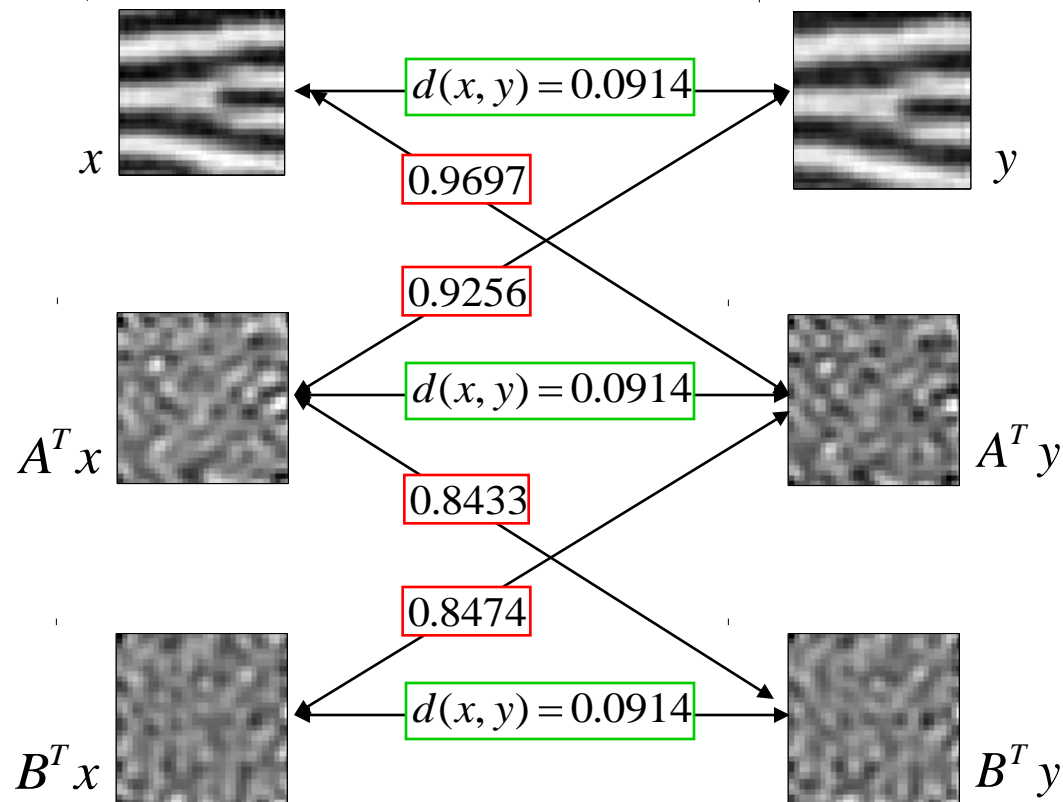
- Each patch can be used to produce **multiple transforms**

Transformed patch matching



- Original **match** among themselves
- Transforms **match** among themselves
- Transform **does not match** with original

Patch transform cancelability



- Score more than 0.5 is a mismatch
- Different Transforms **don't match** with each other

Increasing security: Two factor transformation

- The current construction **is invertible**

If we have the projecting matrix B , and the transform $T(x) = B^T x$

$x = BT(x) = BB^T x$, can be recovered

- **Can we increase security?**

- Two factor transformation

- The projection matrix B is constructed using **two** orthonormal matrices U, V

$$B = UV^T$$

$$UU^T = U^T U = VV^T = V^T V = I$$

$$BB^T = (V^T \overleftarrow{V}) U^T = U (U^T V \overrightarrow{U})^T = I$$

U, V are obtained by performing SVD on a random matrix $R = USV^T$

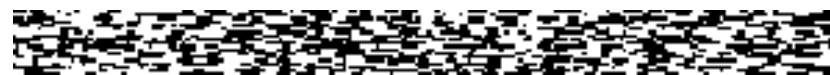
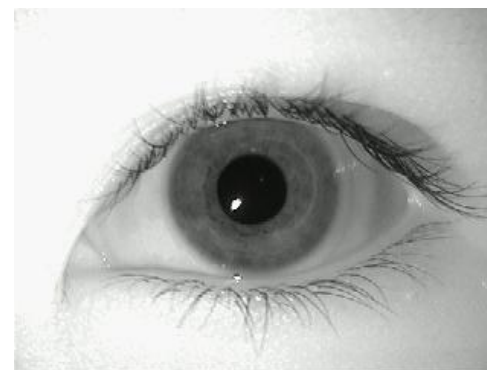
S is not recorded anywhere in the system.

U, V do not leak information about each other

- U and V can be separately stored **separately** (e.g. split between user and application?)
- **Symmetric key, public key comparison**

Steps in building a cancelable iris system

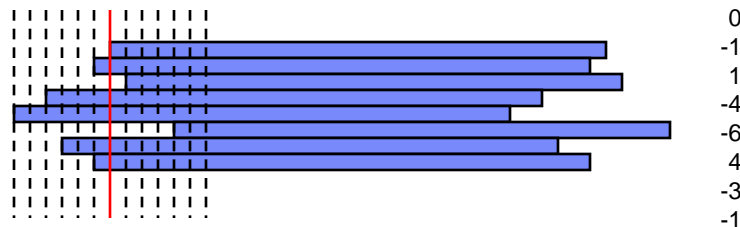
- **Segmentation**
- **Feature extraction**
- **Cancelable techniques**



Method 1: GRAY COMBO

- **template based row shift and combination**

- Step 1: for each row shift circularly:



- Step 2: combine two rows together to get a new one:

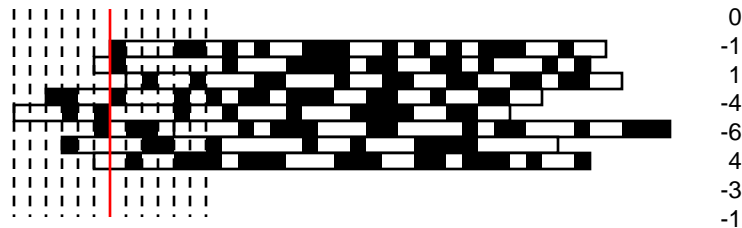
- Intensity +, -
- One row can be used more than once
- Easy methods: odd+even, fold like a mirror

Combine rows 1, 3 to the new 1st row
 Combine rows 2, 8 to the new 2nd row
 Combine rows 4, 6 to the new 3rd row
 Combine rows 5, 7 to the new 4th row

Method 2: BIN COMBO

- code based row shift and combination

- Step 1: for each row shift circularly:



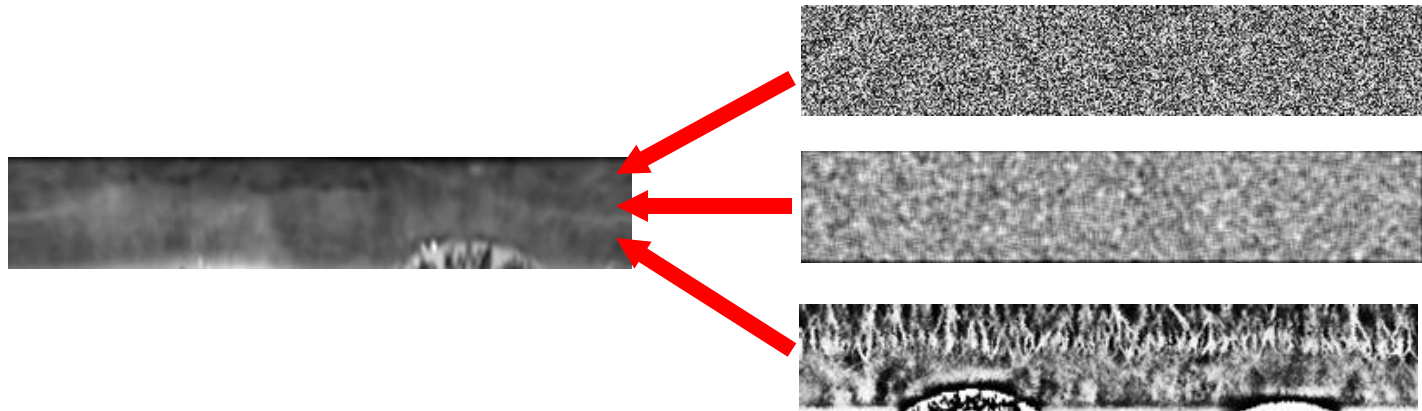
- Step 2: combine two rows together to get a new one:

- Binary XOR, or NXOR
- One row can be used more than once
- Easy methods: odd+even, fold like a mirror

Combine rows 1, 3 to the new 1st row
 Combine rows 2, 8 to the new 2nd row
 Combine rows 4, 6 to the new 3rd row
 Combine rows 5, 7 to the new 4th row

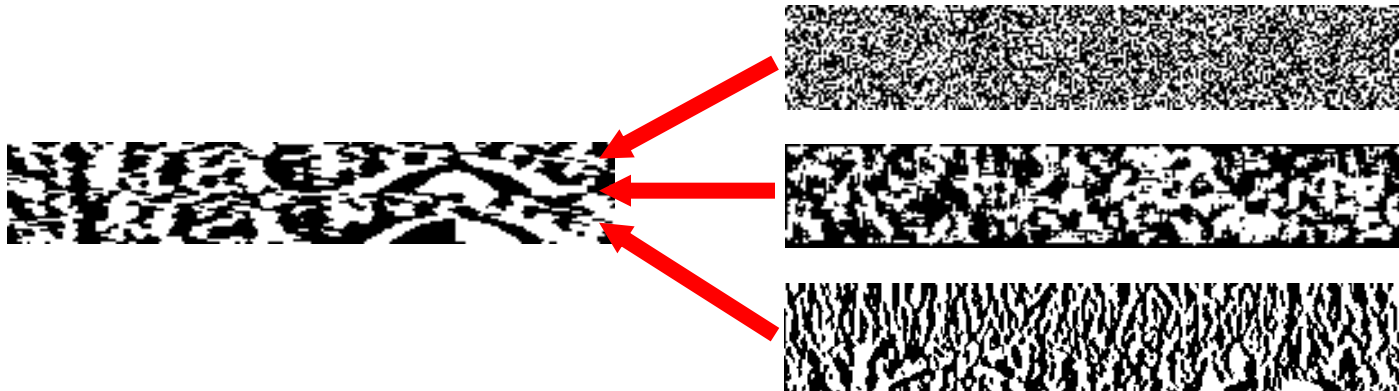
Method 3: GRAY SALT

- **template based salty noise**
 - Just plus a unique pattern --- random noise, random pattern or random synthetic iris texture
 - Generate new code according to the new texture



Method 4: BIN SALT

- **code based salty noise**
 - Just plus a unique binary pattern --- random noise , random pattern or random synthetic iris code

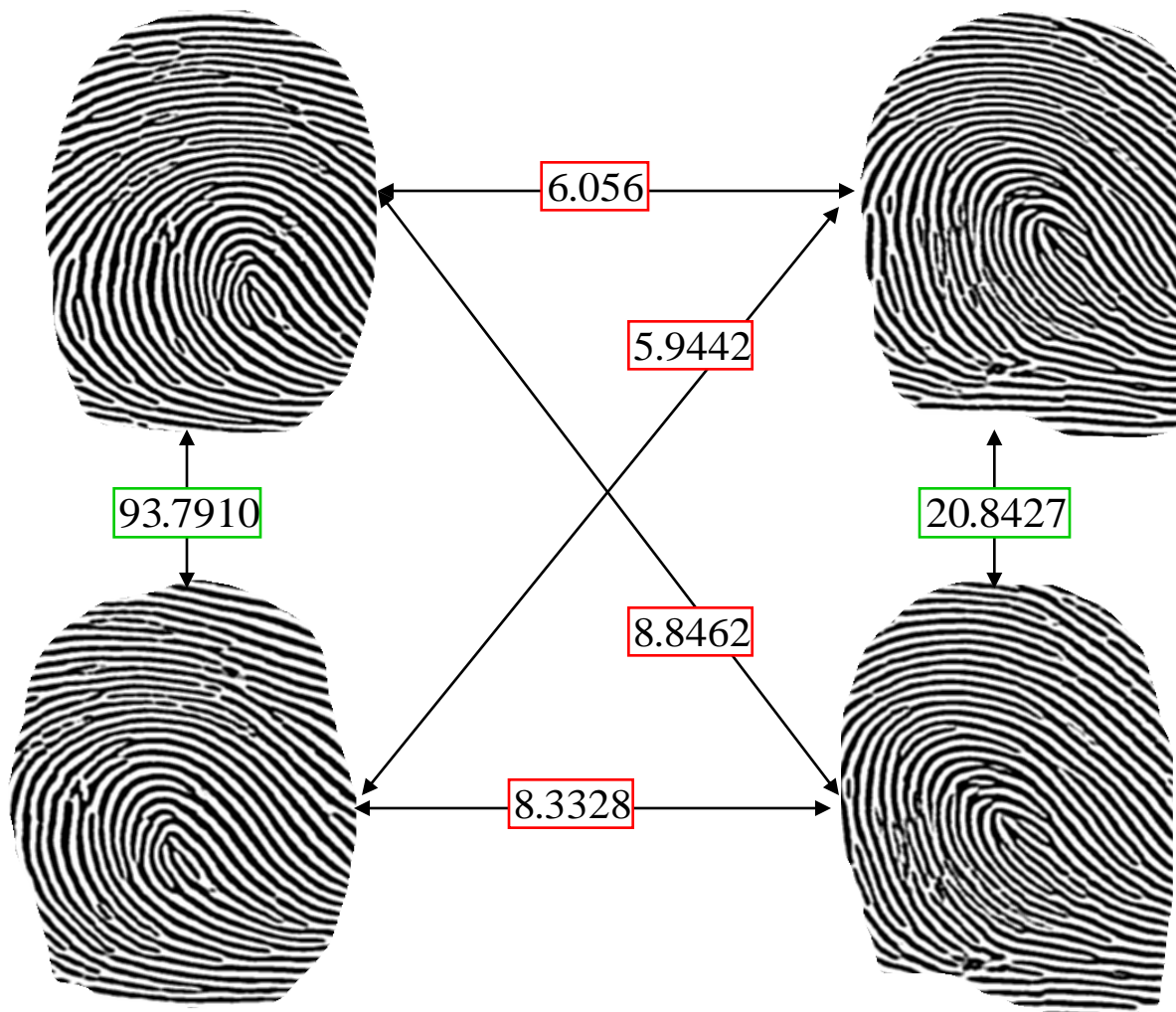


Conclusions

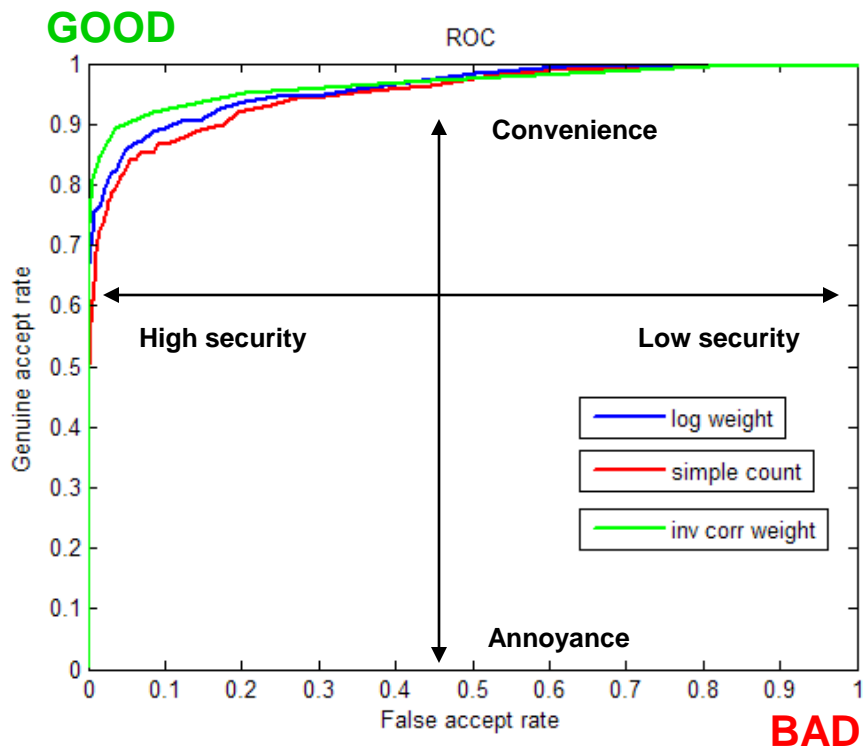
- **There is no substitute to biometrics for positive identification in integrated security applications**
- **Biometrics offer a convenient and efficient method to solve many of the toughest real-life person identification problems**
- **When properly used, biometrics enhance and protect individual privacy**
- **Design of each stage of biometrics based identification system is challenging and entails novel engineering approaches to be more secure**
- **There are a number of technical, political and societal barriers to overcome before a widespread acceptance of biometrics in our society**

Backup

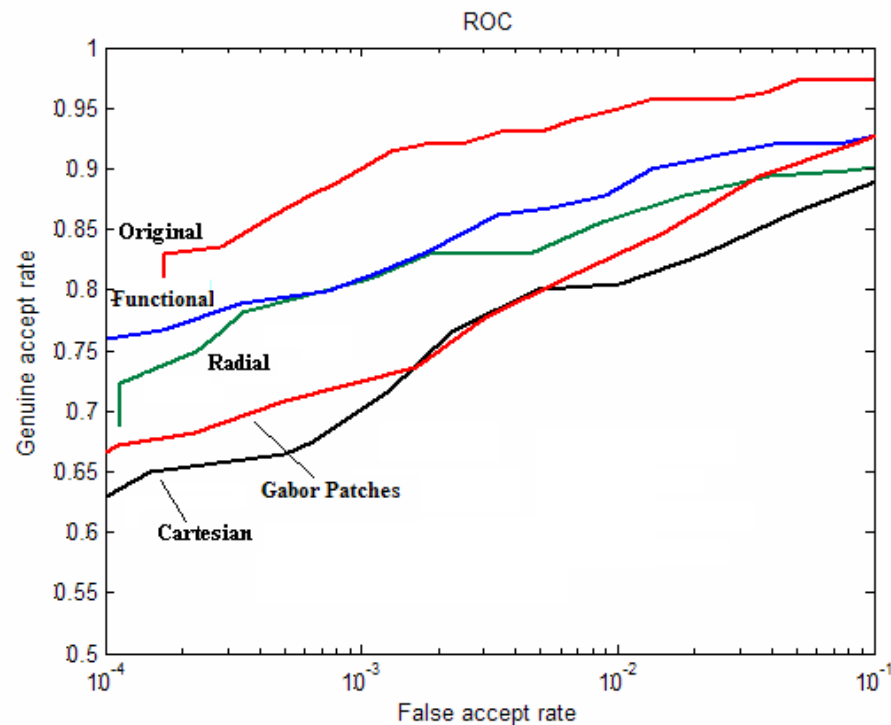
Example



Empirical Evaluation



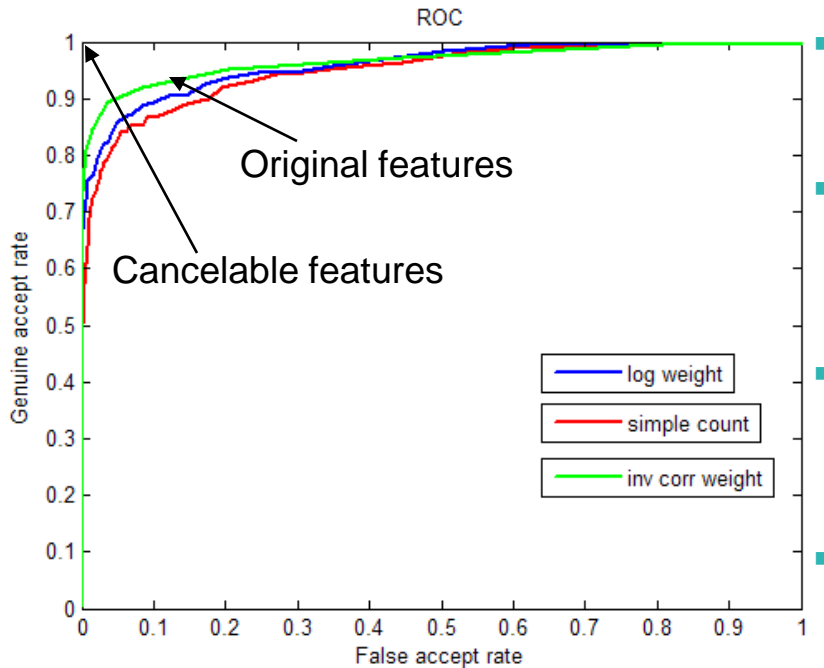
Patch-based verification



Geometry-based verification

- The performance was evaluated over **188 pairs** of fingerprints from the IBM optical database
- Results are over 188 genuine and 17578 impostor comparisons

Empirical Results (1)



■ Patch based verification

- Performance is less than geometry based matchers (62% GAR at 0.01% FAR)

■ Cancelability

- **Complete separation** (100% GAR, 0% FAR) achieved by having **separate** transforms for **separate** individuals

■ Diversity of key space

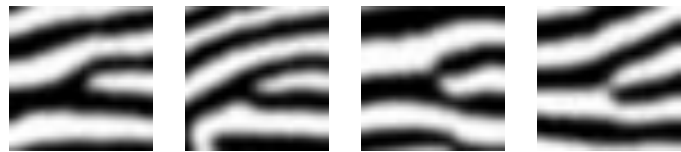
- **Complete separation** (100% GAR, 0% FAR) achieved for **separate** (188) transforms of the **same** individual.

■ Non invertibility

- **Complete separation** (100% GAR, 0% FAR) achieved for **non-invertible** construction as well

- **Perfect performance because uses entropy from key also**
- **If everyone uses the same key performance will not change because distances are preserved**

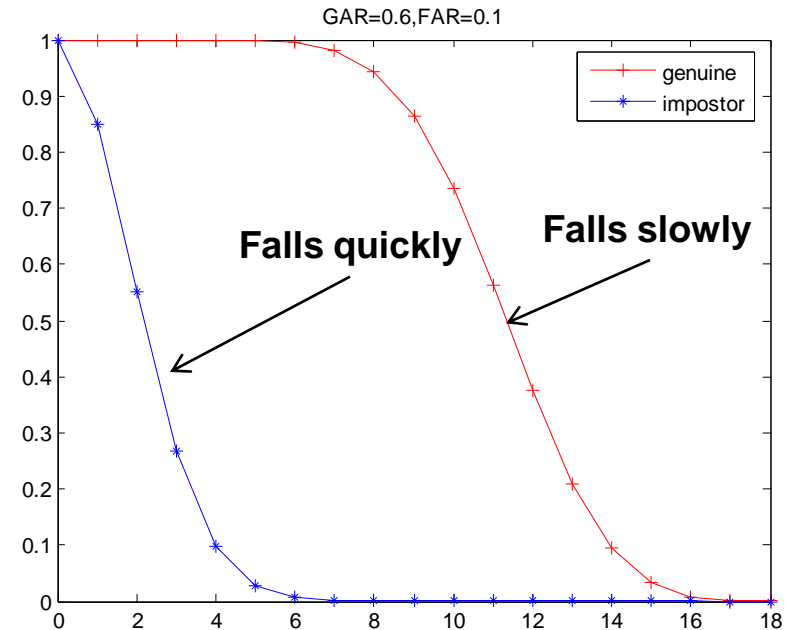
Why does this work?



Low Entropy



High Entropy



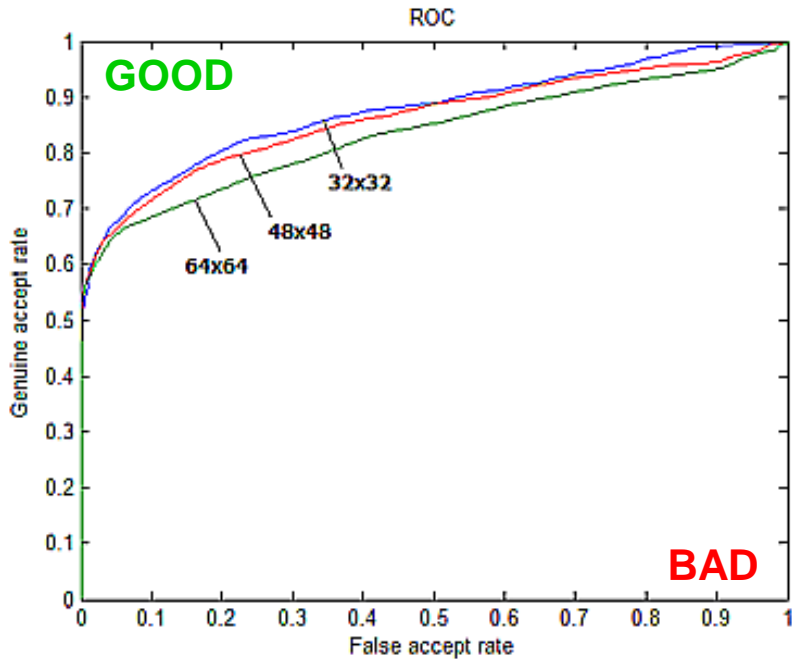
- High entropy patches have high association with the individual
- Given GAR and FAR for **individual matches (using ground truth)**, what is the probability that more than k minutiae will match? (N=50)

$$P_K(k) = 1 - \sum_{n=0}^{k-1} \binom{N}{n} p^n (1-p)^{N-n}$$

N – total number of minutiae

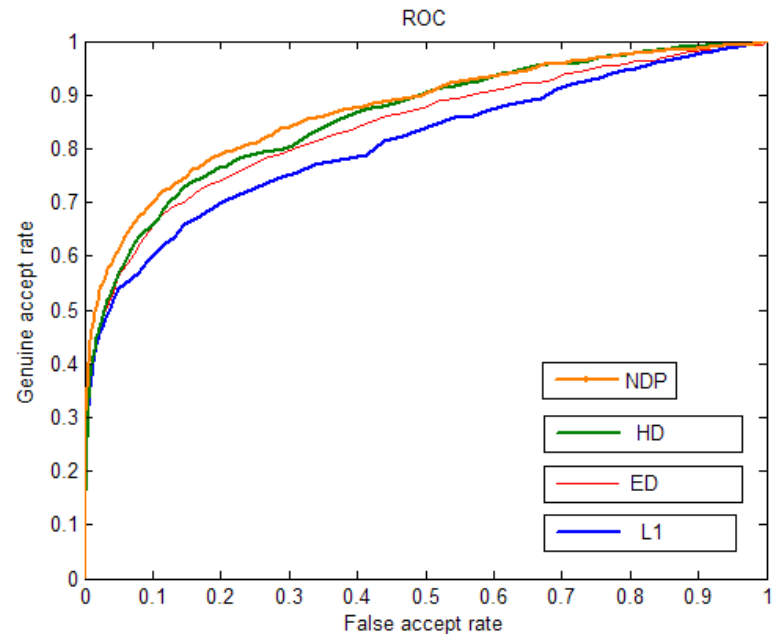
p – probability that a single patch will match

Selecting optimal parameters



Selecting **patch size**

Using ED as the distance measure



Selecting patch **similarity measure**

$$ED: d(\vec{x}, \vec{y}) = \|\vec{x} - \vec{y}\|^2 \quad \vec{x} = (x_1, x_2 \dots x_{272})$$

$$L1: d(\vec{x}, \vec{y}) = |\vec{x} - \vec{y}| \quad \vec{y} = (y_1, y_2 \dots y_{272})$$

$$HD: d(\vec{x}, \vec{y}) = |\text{sgn}(\vec{x}) - \text{sgn}(\vec{y})| \quad \text{Patch size} = 32 \times 32$$

$$NDP: d(\vec{x}, \vec{y}) = 1 - \frac{\langle \vec{x}, \vec{y} \rangle}{\sqrt{\langle \vec{x}, \vec{x} \rangle \langle \vec{y}, \vec{y} \rangle}}$$