Elementi Caratteristici di un Sistema Biometrico

Parte II

Indice

Capitolo 1

Struttura di un sistema biometrico

1.1 Struttura di un sistema biometrico (generale)

1.1.1 Fase di enrollment

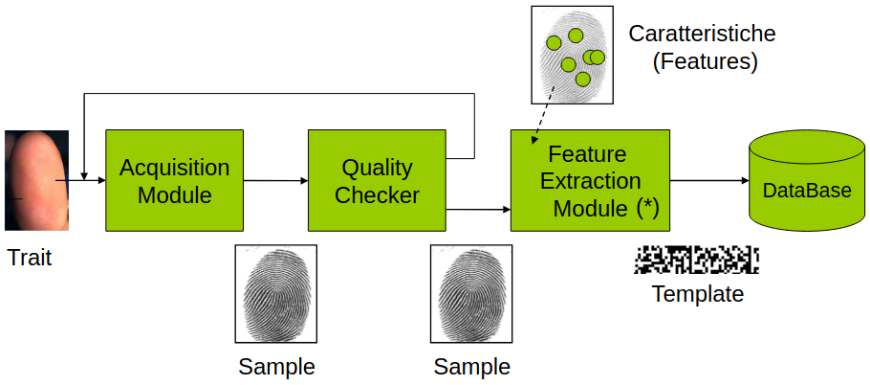


Figura 1.1: Enrollment: (template) –> DB

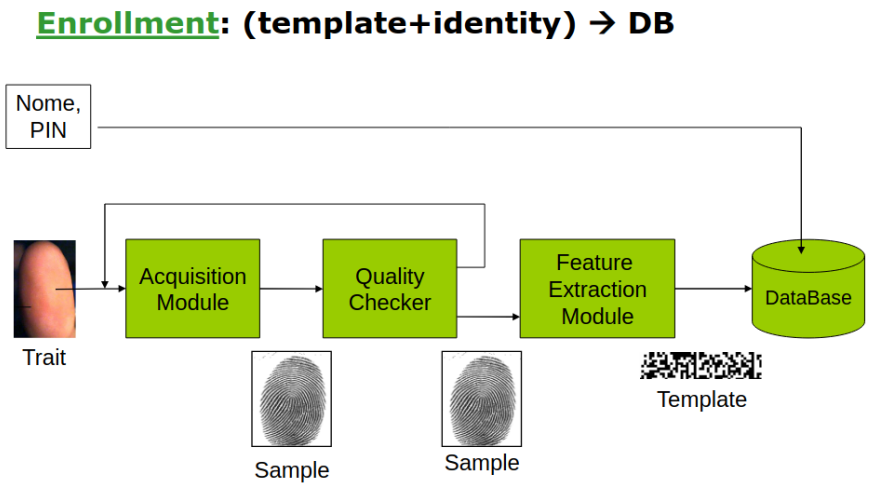


Figura 1.2: (template + identity) –> DB

1.1.2 Verification usando un DB

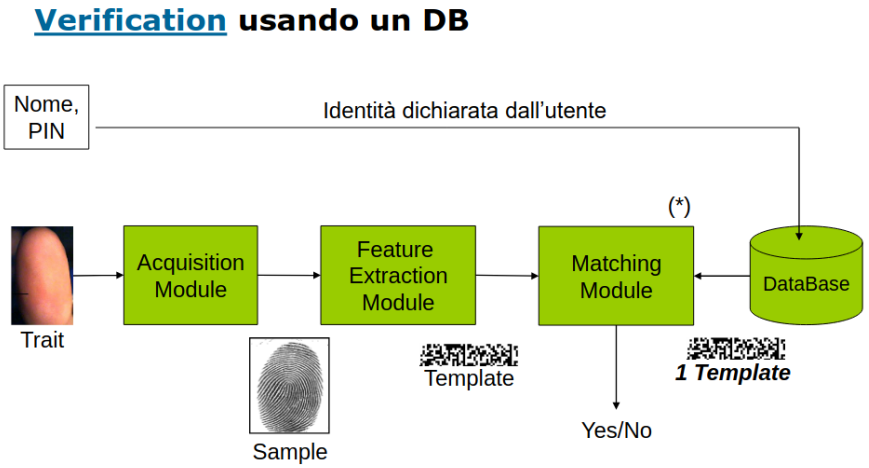


Figura 1.3: Verification usando un DB

1.1.3 Identification

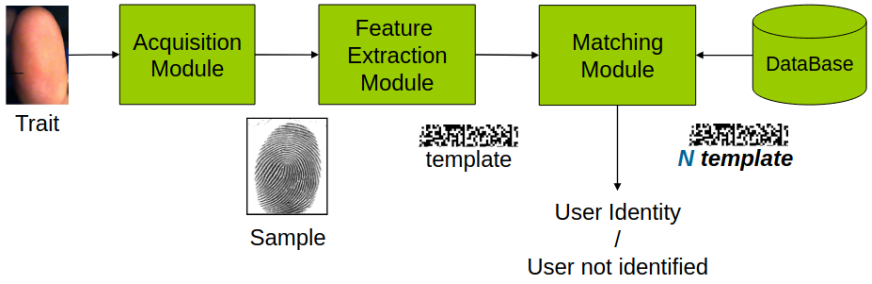


Figura 1.4: Identification

1.2 Struttura per documenti biometrici

1.2.1 Fase di enrollment

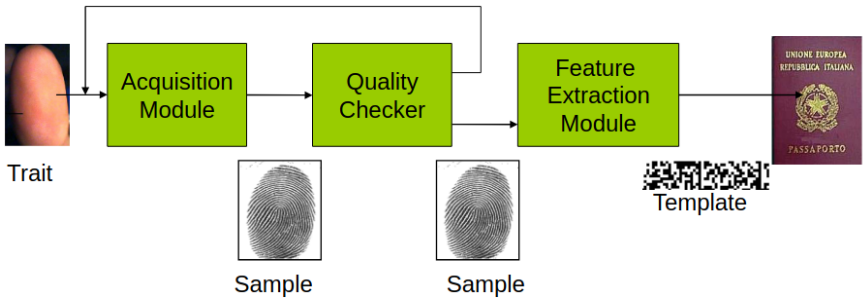


Figura 1.5: (template) –> Documento

1.2.2 Verification (con documento biometrico)

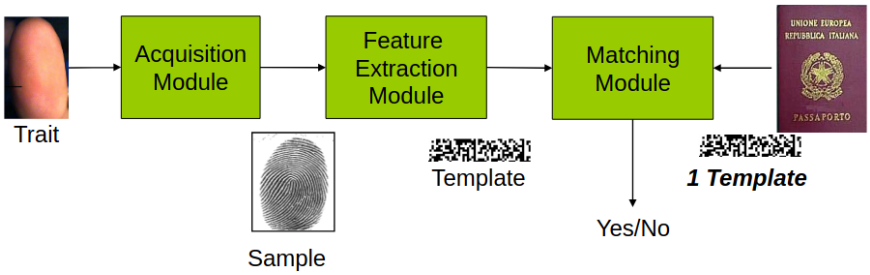


Figura 1.6: Verification con documento biometrico

1.3 Struttura dei sistemi multimodali

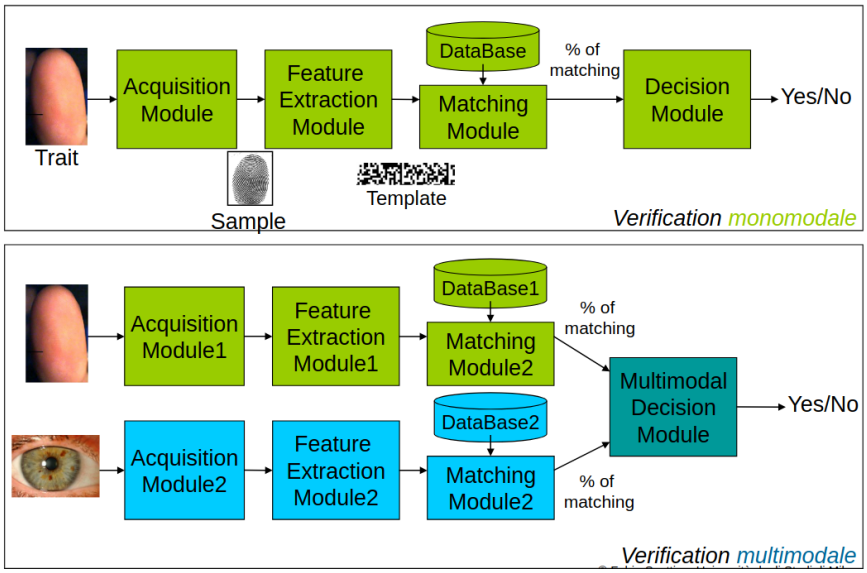


Figura 1.7: Confronto tra la struttura di un sistema monomodale e multimodale

1.4 Struttura dei sistemi biometrici distribuiti

Il termine "distribuito" si riferisce ad un sistema biometrico quando i moduli componenti sono separati e collegati in rete. Piuttosto raro quando si parla di sistemi di autenticazione; è invece comune quando si parla di sistemi di identificazione di grosse dimensioni.
Solitamente è il modulo dei database ad essere separato dai terminali.

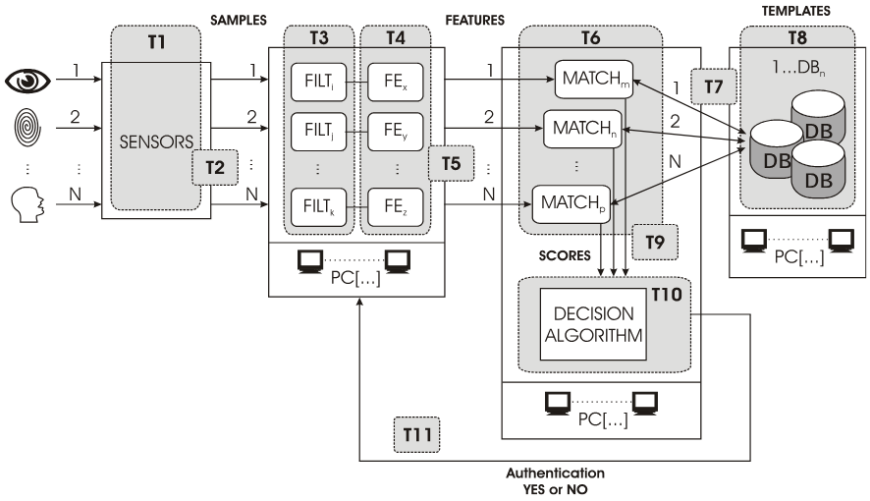


Figura 1.8: Struttura di un sistema biometrico distribuito

1.5 Sistemi biometrici on card

Il termine on card si riferisce al fatto che il template biometrico risiede su una smart card.

Capitolo 2

Aspetti analitici di un sistema biometrico

2.1 Variabilità intraclasse

Si intende la variazione del **sample** o delle **feature** dello stesso individuo tra acquisizioni effettuate in istanti di tempo diversi. Può essere dovuto a:

- effetti casuali (rumore del dispositivo)
- varizioni dello sfondo
- variazioni del tratto (invecchiamento, posizione, espressioni, ecc.)

2.2 Similitudine interclasse

Particolare vicinanza dei **sample** o delle **feature** acquisiti da individui diversi.

Capitolo 3

Regole Generali di Progettazione

3.1 Acquisizione del tratto

Analizziamo i metodi per **progettare** ed eventualmente **migliorare** il modulo di acquisizione.
L'acquisizione di una informazione rilevante **è un processo critico** e non sempre adeguatamente studiato.
La cura nel processo di acquisizione **influenza pesantemente l'accuratezza** finale del sistema.

Il processo di acquisizione si divide in **due fasi:**

- **Valutazione della qualità:** controllo automatico sulla correttezza dei dati in ingresso coerentemente alle successive elaborazioni
- **Segmentazione:** separazione dei dati in ingresso nell'oggetto di interesse (foreground) e nello sfondo/informazione non rilevante (background)

Estrarre molte informazioni

È buona prassi cercare di estrarre maggiori informazioni possibili per migliorare le performance del sistema biometrico.

- **Acquisire anche il contesto** attorno al sample; permette di trovare meglio il vero volto per sottrazione di frame senza elaborazioni troppo complesse
- **Evitare sul nascere di fare cattive acquisizioni** per non dover richiedere il sample (ad esempio, controllare se il soggetto è in movimento o alla distanza corretta prima di elaborare il frame)

3.1.1 Controllo della qualità

Dopo l'acquisizione molti sistemi attuano un controllo automatico della qualità del tratto rilevato per evitare problemi di funzionamento.
I sistemi di controllo della qualità producono un **indice di qualità** del sample acquisito:

- se l'indice di qualità è sufficientemente alto si prosegue
- altrimenti si torna ad acquisire un altro sample

Il concetto di base è semplice, ma la progettazione e realizzazione dell'indice di qualità non lo è; alcuni punti sono che:

- non sempre esiste un **modello rigoroso e realistico** della misura in ingresso da usare per calcolare l'indice; ad esempio, se potessimo definire come dovrebbe essere un'immagine ottimale di un'impronta, potremmo esprime l'indice di qualità come la "distanza" dell'immagine in ingresso da quella ottima
- non sempre esistono **metriche rigorose e robuste** per misurare la distanza del sample in ingresso con il riferimento ottimale

Signal/Image enhancement

In alcuni casi non è possibile rifiutare un sample perché il suo indice di qualità è basso (ad esempio database giudiziari); in questo caso, il sistema cerca di estrarre le informazioni (foreground) dal rumore (background) in modo tale da far funzionare il resto della catena di moduli del sistema.

Solitamente questa fase è ad alta complessità computazionale. Può generare i cosidetti **artefatti**; ad esempio, data un'impronta rumorosa genere delle minuzie che non erano presenti nell'immagine originale.

3.2 Rappresentazione

Un'acquisizione di un sistema non elaborata, chiamata sample, è:

- **non invariante** rispetto al momento dell'acquisizione
- **non discriminatoria** (sono tutte diverse)

In un sistema biometrico occorre **studiare come rappresentare al meglio l'informazione** per rispondere alla domanda: *"Quale rappresentazione machine-readable cattura* ***completamente*** *l'informazione* ***invariante*** *e* ***discriminatoria*** *della misura in ingresso?*
Il problema della rappresentazione consiste nel determinare uno spazio di misura che sia:

- **invariante** (meno variante) rispetto ad acquisizioni dello stesso individuo

- che si **differinzi massivamente** dalla acquisizioni di individui diversi

In altre parole, si può dire che la rappresentazione deve fornire:

- **alta variabilità interclasse** (io diverso da tutti gli altri)
- **bassa variabilità intraclasse** (io simile a me stesso nei miei sample)

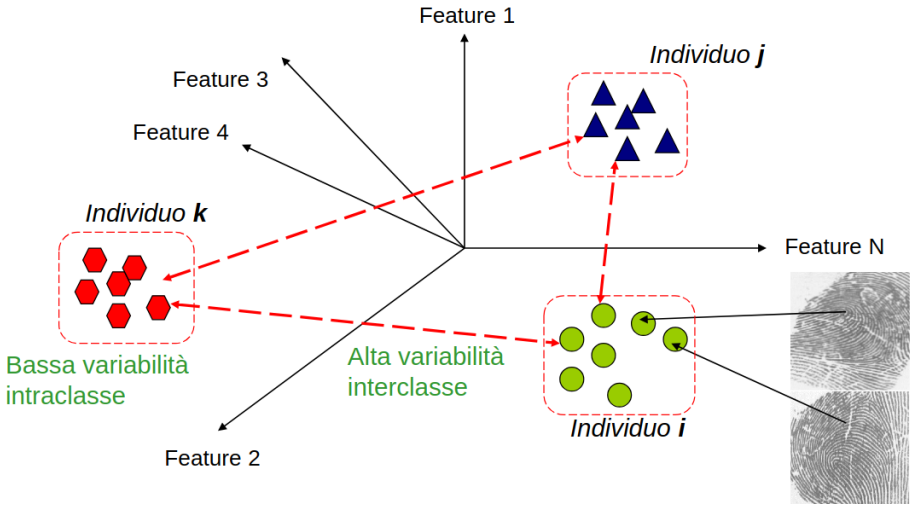


Figura 3.1: Visualizzazione del problema della rappresentazione in un spazio delle features N-dimensionale

Il problema della rappresentazione si suddivide in:

- rappresentazione del sample
- estrazione delle caratteristiche
- rappresentazione del template

3.2.1 Rappresentazione del sample

Si riferisce alle caratteristiche tecniche del processo di acquisizione e memorizzazione del sample. Varia a seconda del tratto biometrico, ci si riferisce a questi dati come ***raw data***.

3.2.2 Estrazione di caratteristiche

L'estrazione delle caratteristiche impatta sul modulo evidenziato presente sia in fase di enrollment che di verification/identification.

Avendo i dati raw provenienti dalle misurazioni occorre ora **estrarne la rappresentazione nello spazio delle caratteristiche.** Questo non è mai un problema semplice, specialmente con dati rumorosi.

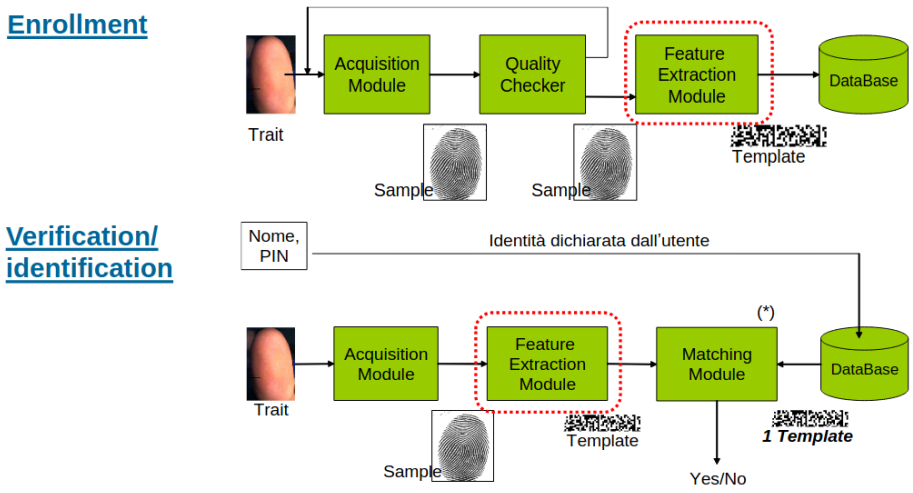


Figura 3.2: Impatto dell'estrazione delle caratterisiche

Può essere fatta in maniera manuale da un operatore oppure utilizzando un sistema automatico; **lo spazio delle caratteristiche di un sistema automatico tende ad essere diverso** da quello di un sistema con estrazione manuale.

3.3 Matching

Il matching impatta sul modulo evidenziato solo nella fase di verification /identification.

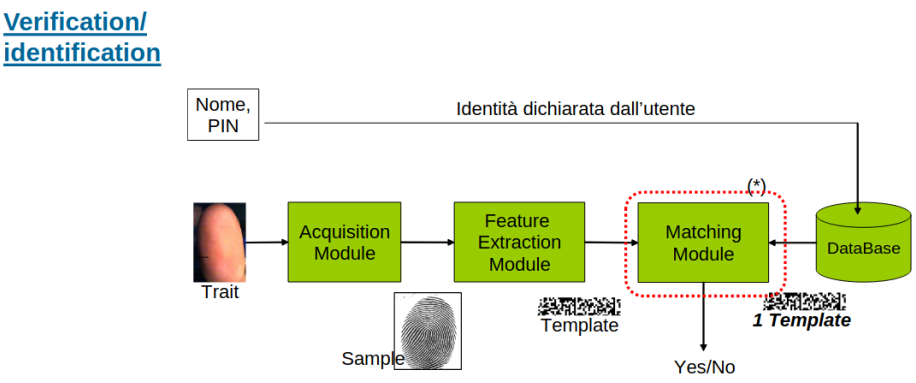


Figura 3.3: Impatto del matching

3.4 Ricerca ed organizzazione dei DB biometrici

Scalabilità

I sistemi che devono gestire una grande quantità di identità dovrebbero essere in grado di operare efficacemente quando il numero di utenti registrati nel DB aumenta.
L'introduzione della biometria in un sistema di identificazione di grandi dimensioni produce dei vantaggi:

- non soffrono del problema della produzione e del rinnovo dei documenti di identità
- sono competitivi in termini di costo e mantenimento

3.4.1 Organizzazione del DB (indexing) e tasso di prenetazione nel DB

L'obiettivo di gestire efficacemente la complessità delle ricerche rispetto all'incremento del numero di template nel DB del sistema può essere raggiunto solo con una attenta organizzazione dei DB (indexing); un DB organizzato permette di non confrontare un template in ingresso con tutti i template nel DB ma solo con quelli contenuti in una partizione.

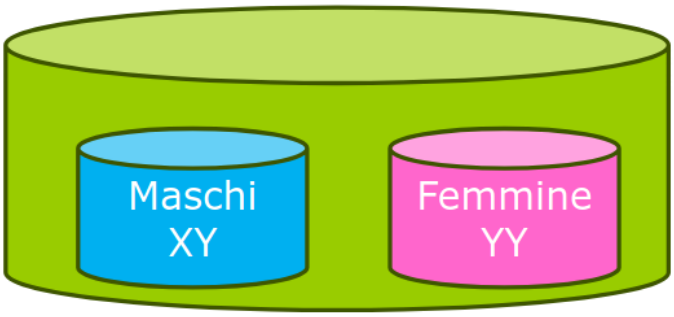


Figura 3.4: Esempio di indexing

In generale si può definire il tasso di penetrazione come la percentuale del database totale da esaminare in media per ogni ricerca. Più basso è il tasso di penetrazione, più efficiente è il sistema.
Tuttavia, per i sistemi biometrici serve fare un distinguo:
la proporzione attesa dei template da cercare su tutti i campioni di input secondo la regola che la ricerca prosegue attraverso l'intera partizione, indipendentemente dal fatto che venga trovata una corrispondenza o meno.

3.4.2 Binning

Per giovare delle partizioni del DB occorre disporre di un **algoritmo automatico molto robusto** per la classificazione dei template; quando il DB viene creato, i template vengono disposti nelle partizioni (***bins***).

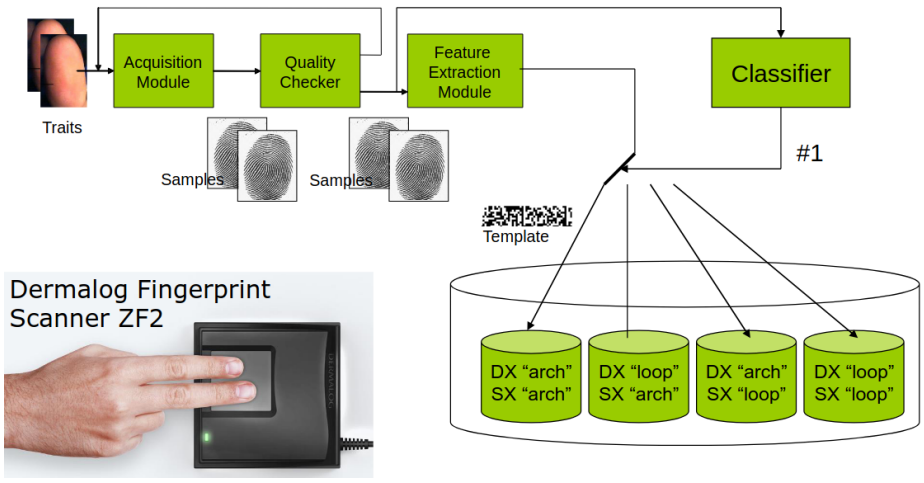


Figura 3.5: Esempio di suddivisioni in bins di un DB

Calcolo del numero di bin ottimale

Se ho **N** impronte disitnguibili (indice, medio...) tutte divisibili **M** tipi (arch, whorl, ...), allora numero migliore di bin è = M^N.

Binning error

I problemi nascono quando un individuo presenta i propri tratti biometrici al sistema e l'algoritmo di classificazione del tratto sbaglia il bin (binning error); se l'indivduo era registrato nel DB, probabilmente non si avrà un match in quanto i template che matchano sono in un bin differente.

Calcolo del binning error

Definiamo:
Binning error = **Prob(**fare almeno 1 errore**)**
p = **Prob(**sbagliare una classificazione**)**

Allora:
Binning error = **Prob(**fare almeno 1 errore**)**
= 1 - **Prob(**0 errori**)**
= 1 - $(1-p) * (1-p)$
= 1 - $(p^2 - 2p + 1)$
= $2p - p^2$ = 2p - *molto poco (se p è piccola)*

Ne consegue che:
Binning error $\rightarrow$ **circa** $N * p$ **se si usano N impronte** se:
- p è piccola (vero)
- classi distribuite (simile numero di impronte nei bin, vero)

Se si **abbassano il numero delle classi:**

- **Si alza il P.R.** $\rightarrow$ **si abbassa la qualità del P.R. ottenibile**
- **Si abbassa l'errore di classificazione**

Capitolo 4

Introduzione alla misura dei parametri

Nel caso dei sistemi biometrici non è banale rispondere alla domanda: ***"qual è il tasso di errore di verifica/identificazione?"***
Per descrivere le performance del sistema è necessario disporre di **un insieme di dati** e **curve di funzionamento**; per questo motivo diventa difficile comparare due sistemi in quanto non basta confrontare 2 numeri.

Genuini ed impostori

Si impiegano i termini:

- **genuino** per indicare un individuo che accede al sistema e ha titolo per farlo
- **impostore** per chi prova ad accedere senza averne titolo

La formulazione del problema risulterà diversa a seconda che il sistema funzioni in **verification** o **identification**.

4.1 Verifica e Identificazione

4.1.1 Verifica

"tu sei chi dici di essere?"
Il problema in questo caso si riconduce ad un caso di classificazione binaria.

Problema della verifica

Dato in ingresso un insieme di caratteristiche X_Q e la dichiarata identità I, occorre determinare se (I, X_Q) appartiene a w_1 o w_2, dove:

- w_1 indica che la richiesta è vera (utente genuino)
- w_2 indica che la richiesta è falsa (un impostore)

Tipicamente, le caratteristiche X_Q vengono controllate con le caratteristiche X_I (il template associato alla identità I).

Regola di decisione per la verifica

Si tratta di una comparazione con soglia:

$$(I, X_Q) \in \begin{cases} \omega_1 & se\ S(X_Q, X_I) \geq T \\ \omega_2 & altrimenti \end{cases}$$

Dove:

- S è la funzione che misura la similitudine tra X_Q e X_I
- T è la soglia prefissata
- $S(X_Q, X_I)$ prende il nome di **match score**

4.1.2 Identificazione

"il sistema controlla se i tuoi dati biometrici corrispondono ad un insieme di identità registrate"

Problema di identificazione

Dato in ingresso un insieme di caratteristiche X_Q, determinare l'identità I_k, con $k \in \{1, 2, 3, ..., M, M+1\}$ dove $I_1, I_2, ..., I_M$ sono le M identità memorizzate nel sistema e $I_M + 1$ rappresenta il **caso di reiezione**.
Nel caso di reiezione nessuna delle M identità registrate è sufficientemente simile al dato in ingresso.

Regola di decisione per l'identificazione

Si tratta di M comparazioni con soglia con la seguente regola di decisione:

$$X_Q \in \begin{cases} I_K & se\ K = \arg\max_k \left\{S(X_Q, X_{Ik})\right\} and\ S(X_Q, X_{Ik}) \geq T \\ I_{M+1} & altrimenti \end{cases}$$

Dove:

- $X_I k$ è il template corrispondente alla identità I_k
- T è la soglia prefissata
- $S(X_Q, X_I)$ prende il nome di **match score**

In alcuni casi ci si riferisce ad una **misura della distanza** fra X_Q e X_I; una **grande distanza** fra i vettori di features porta ad un **basso match score**.

4.2 Distanza tra template

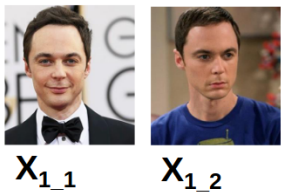
N template, provenienti dalla stessa persona ma acquisiti in tempi diversi, NON sono **mai uguali**.
Esiste sempre una distanza nello spazio delle features che separa i template anche della stessa persona (rumore, posa del soggetto, illuminazione, condizione ambientiali, ...)
Questa comporta che **la soglia T non può esser arbitrariamente abbassata**, altrimenti nessuno sarebbe identificato.
Se si riscontrasse una **distanza nulla** fra X_Q e X_I (quindi $S(X_Q, X_I) = max$), probabilmente saremmo di fronte ad un **replay attack**: una copia illecita di un template memorizzato che viene riproposto per frodare un sistema.

Genuini ed impostori

Si parla di **genuine score** quando si confrontano le distanze tra i template dello stesso individuo.
Si parla di **impostor score** quando si confrontano le distanze tra i template di indivdui diversi.

4.2.1 Distribuzioni dei match score

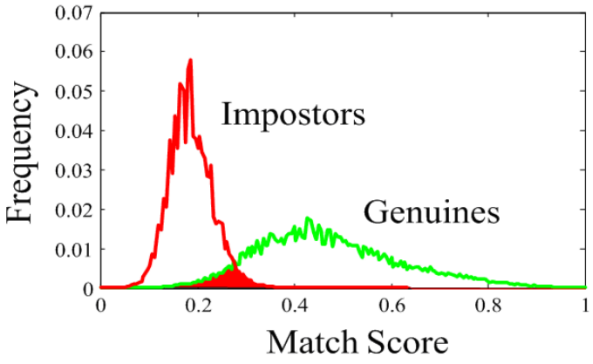
- Sia X_{i_j} il j-esimo template dell'individuo i-esimo



X_{1_1} X_{1_2}

Match Genuini
$S(X_{1\_1}, X_{1\_2}) = 0.7$
$S(X_{1\_1}, X_{1\_3}) = 0.8$
$S(X_{2\_1}, X_{2\_2}) = 0.4$
$S(X_{2\_1}, X_{2\_3}) = 0.5$

Match impostori
$S(X_{1\_1}, X_{3\_2}) = 0.11$
$S(X_{4\_1}, X_{3\_1}) = 0.21$
$S(X_{5\_2}, X_{1\_2}) = 0.001$
$S(X_{2\_2}, X_{1\_2}) = 0.19$



4.3 False Match e False Non-Match

- **False Match:** *il ladro entra in casa perché il sistema lo ha scambiato per noi; impostorScore* $> T$
- **False Non-Match:** *non entrate in casa perché il sistema ritiene che il template non assomigli abbastanza a quello/i registrati; genuineScore* $< T$

4.3.1 FM Rate, FNM Rate

Supponiamo di poter variare la soglia e di fissarla a un valore T in mezzo fra il picco degli impostori e quello dei genuini. Notiamo che:

- un certo numero di persone appartenenti al gruppo dei **genuini** sono **sotto la soglia T**; non saranno autorizzati e daranno errore di False Non-Match (FNM).

  $FNMR(T) = FNM(T) \ / \ numGenuini$

- una parte degli **impostori** hanno valori di match **sopra la soglia T**; saranno autorizzati e daranno errori di False Match (FM).

  $FMR(T) = FM(T) \ / \ numImpostori$

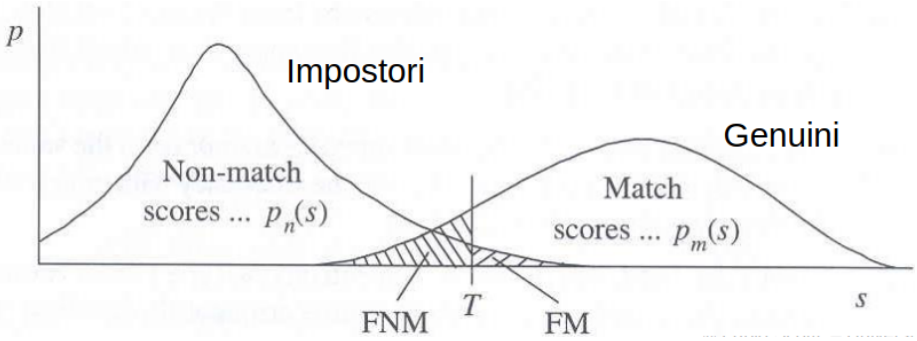


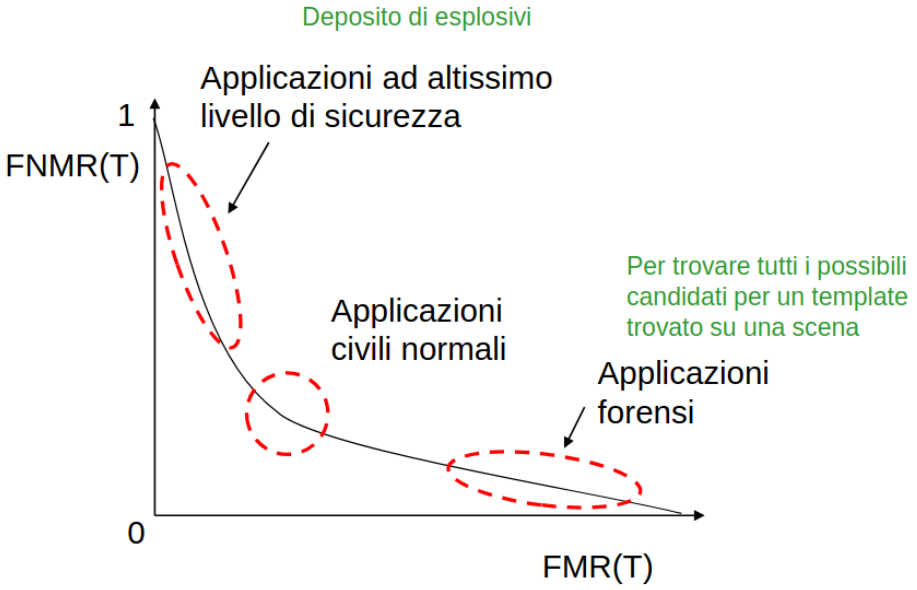
Figura 4.1: FNM e FM

Il funzionamento di un sistema biometrico dal punto di vista degli errori commessi è descritto dai tassi *FMR(T)* e *FNMR(T)* per tutti i valori della soglia.

4.4 Decision Error Tradeoff (DET) e Receiver Operating Characteristic (ROC)

DET: Regioni di funzionamento

Regolando la soglia T possiamo regolare il livello di sicurezza.



DET: Equal Error Rate

L'EER è il tasso di errore corrispondente all'unico punto nel quale si ha FNMR = FMR.
Si tratta dell'unico numero singolo che può riassumere il funzionamento del sistema.

ROC = 1 - DET

La curva DET e ROC mostrano le stesse informazioni. La DET mette l'attenzione sul FNM (genuini che non entrano), mentre la ROC mette l'evidenza su 1 - FNM (quanti genuini riescono ad entrare).

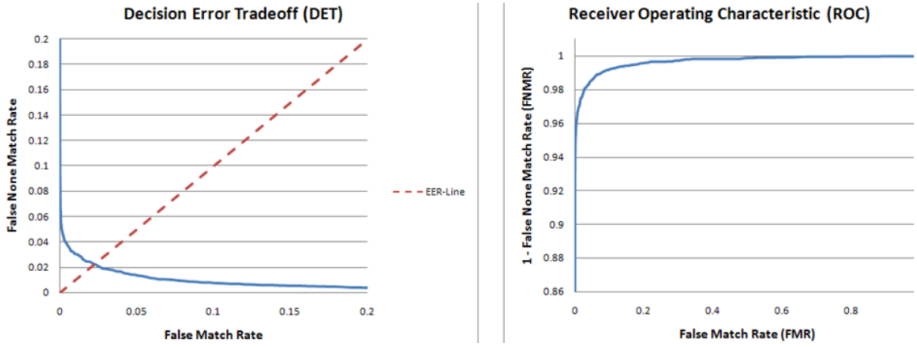


Figura 4.2: Curve DET e ROC

4.5 Metodi statistici per la stima dei parametri in un sistema biometrico

4.5.1 Che modello usiamo?

Chiariamo il concetto di **Prova/Esperimento di Bernoulli:**

1. ad ogni singola prova si hanno solo due esiti possibili
   - successo (1)
   - insuccesso (0)
2. La probabilità $\boldsymbol{p}$ dell'evento 'successo' è costante
3. i risultati delle prove sono **indipendenti**

Date due impronte il SB in autenticazione mostra un tasso di errori stimabile e fissato $\boldsymbol{p}$.
L'evento che il SB sbagli una autenticazione è un **esperimento/prova di Bernoulli**.

Un SB usato in identificazione (1:N) si può modellizzare come un N prove di Bernoulli, ovvero un **processo di Bernoulli**.

4.5.2 Regola dei 3

"qual è il tasso di errore più basso p che può essere stimato con un esperimento di comparazione di N campioni indipendenti?"
Se abbiamo un sistema che commette 0 errori su N prove non dobbiamo pensare di avere un sistema con $p = 0$, ma con il 95% di confidenza abbiamo un sistema che ha $p \approx 3/N$.

Esempio

Se faccio 300 prove e ho 0 errori, allora posso dire con confidenza del 95% che il sistema ha un tasso di errore stimato del $p \approx 3/N = 3/100 = 1\%$

4.5.3 Regola dei 30

La regola dei 30 è utilizzata per determinare la larghezza del campione biometrico in questo modo:

- per essere sicuro con intervallo di confidenza del 90% che il tasso di errore **vero** sia tra il $\pm 30\%$ del tasso di errore **osservato**, ci devono essere almeno 30 errori.

Esempio

Se abbiamo 30 FNM in 3000 comparazioni, possiamo dire (con intervallo di confidenza del 90%) che l'errore vero sia tra $0,7\%$ e $1,3\%$.