

# Introducción a la Biometría

Information and Communications Technologies for Verticals

Grupo de Lectura: Reconocimiento de Patrones

Javier Preciozzi - Dirección Nacional de Identificación Civil

2 de Setiembre de 2015

## Agenda

- Quienes somos
- Marco teórico: Basado en el libro "Introduction to Biometrics" - A. Jain, A. Ross y K. Nandakumar"
- Caso de estudio: Dirección Nacional de Identificación Civil

## Quienes somos

- Grupo de Tratamiento de Imágenes - Instituto de Ingeniería Eléctrica (Facultad de Ingeniería)
  - Alicia Fernández
  - Federico Lecumberry
  - Luis Di Martino (maestría)
  - Gabriel Lema (maestría)
- Dirección Nacional de Identificación Civil - Ministerio del Interior
  - Javier Preciozzi
  - Luis Di Martino

## Antecedentes

- 2001 - Convenio DNIC-FING
- 2005 - CSIC: Fusión de información biométrica
- 2102 - Proyecto de grado: Faceval - Un sistema de reconocimiento facial
  - Primer premio en Ingeniería DeMuestra 2012
  - Primer puesto en concurso de proyectos de grado 2012-2013 organizado por la Academia Nacional de Ingeniería
- 2014 - Maestría de Luis Di Martino: Estudio, implementación y validación de algoritmos de fusión biométrica
- 2014 - Maestría de Gabriel Lema: Reconocimiento facial en condiciones no controladas (Faces in the Wild)
- 2015 - Fondo María Viñas (ANII): Fusión biométrica: Aplicación a una base de identificación civil

## Publicaciones

- 2001: Performance evaluation of an automatic fingerprint classification algorithm adapted to a Vucetich based classification system A. Bartesaghi, A. Gómez, A. Fernández - International Conference on Audio-and Video-Based Biometric Person Authentication, Proceedings. Halmstad, Sweden 2001
- 2004: Performance improvement in a fingerprint classification system using anisotropic diffusion G. Vallarino, G. Gianarelli, A. Gómez, A. Fernández, A. Pardo - CIARP 2004. Puebla, México, 2004
- 2007: Aguará: An Improved Face Recognition Algorithm through Gabor Filter Adaptation C. Aguerrebere, G. Capdehourat, M. Delbracio, M. Mateu, A. Fernández, F. Lecumberry - IEEE Workshop on Automatic Identification Advanced Technologies, Alghero, Italy 2007
- 2014: An a-contrario Approach for Face Matching - L. Di Martino, J. Preciozzi, F. Lecumberry and A. Fernández - International Conference on Pattern Recognition Applications and Methods 2014.
- 2014: Evaluation of a face recognition system performance's variation on a citizen passport database G. Lema, L. Di Martino, S. Berchesi, A. Fernández, F. Lecumberry and J. Preciozzi (CLEI- 2014).
- 2014: Face Matching with an A-Contrario False Detection Control - L. Di Martino, J. Preciozzi, F. Lecumberry and A. Fernández - Journal on Neurocomputing, Elsevier.

## Identificación



MICHIGAN

Expires  
10-10-2010

OPERATOR LICENCE  
R 789 965 543 897

ARUN ROSS  
123, McDermott Street  
East Lansing MI 48824

**Online  
Banking**



**Username**

**Password**

[Log In](#)

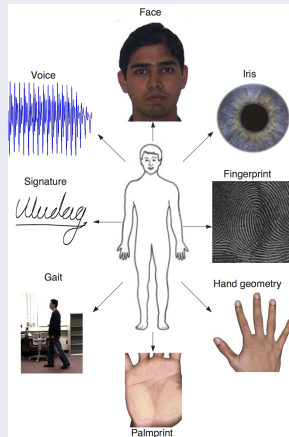
[Forgot  
Password?](#)

[Sign Up](#) | [Learn  
More](#) | [Security](#)



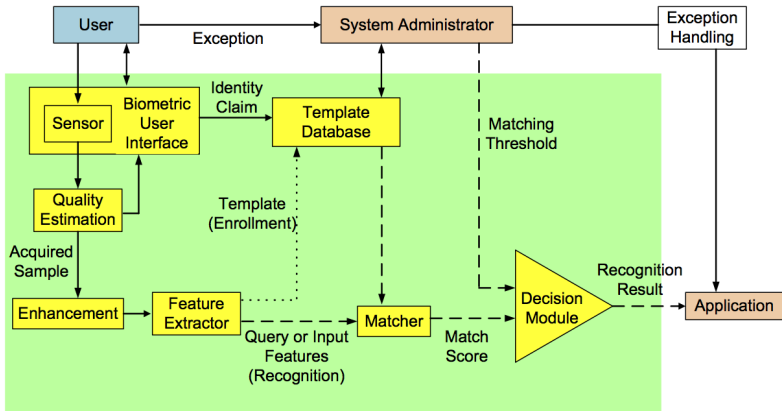
# Sistemas biométricos

- Utilizan rasgos biométricos: huella dactilar, cara, iris, voz, etc.
- Características que se buscan:
  - Universal
  - Único
  - Permanente
  - Medible
  - Rápido
  - Aceptación
  - Robusto



# Introducción a la biometría - Marco teórico

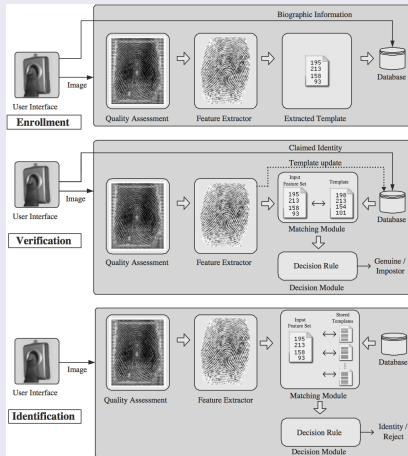
## Diagrama general de un sistema biométrico





# Introducción a la biometría - Marco teórico

## Funcionalidades de un sistema biométrico



## Performance de un sistema biométrico

- **FMR** False Match Rate (o **FAR**: False Acceptance Rate): La probabilidad que dos muestras pertenecientes a dos personas distintas sean clasificadas como pertenecientes a la misma clase.
- **FNMR** False Non-Match Rate (o **FRR**: False Rejection Rate): La probabilidad que dos muestras pertenecientes a la misma persona distintas sea clasificada como que no pertenecen a la misma clase.
- **ROC** Receiver Operating Characteristic: Es la curva que se determina al graficar los verdaderos positivos contra los falsos positivos (FAR)

## Dirección Nacional de Identificación Civil

Tiene por cometido esencial la identificación de las personas físicas que habitan el territorio de la República, otorgando la Cédula de Identidad de acuerdo a la documentación probatoria y a la confrontación dactiloscópica. La DNIC tiene además competencia para la emisión de pasaportes comunes.

## Datos básicos (agosto 2015)

- 35 oficinas distribuidas en todo el país
- Base de datos centralizada
- 5.623.107 registros de persona en la base de datos
- 3.543.383 personas con huellas dactilares en el sistema
- 3.700.966 personas con al menos una fotografías del rostro

## Funciones biométricas

- Enrolamiento: 10 huellas dactilares y una fotografía (en breve, firma)
- Verificación de identidad: basado en huellas dactilares
- Identificación: basado en huellas dactilares

# Enrolamiento

- Enrolamiento: Se toman en papel 10 huellas dactilares



- Se escanean, se cortan por dedo y se guardan en el sistema



## Verificación de Identidad

- Se lleva a cabo en la renovación del documento
- El costo de aceptar un falso positivo es muy alto, por lo tanto se usará un FAR muy bajo

## Módulos

- Sensores
- Extracción de características
  - Análisis de calidad
  - Segmentación de la imagen
- Clasificación (matcheo)
- Decisión

## Sensores utilizados

MorphoSmart 300

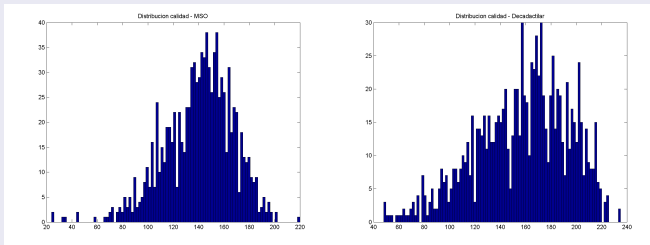


Dermalog ZF1



## Extracción de características

### Análisis de calidad para el dedo índice



| Ficha papel | Media | Varianza |
|-------------|-------|----------|
| Pulgar      | 157   | 36       |
| Índice      | 121   | 35       |
| Medio       | 128   | 42       |
| Anular      | 120   | 43       |
| Meñique     | 94    | 37       |



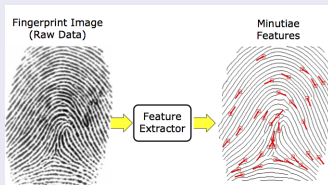
## Extracción de características

### Segmentación



| Ficha papel | Media | Mejora | Varianza | Mejora |
|-------------|-------|--------|----------|--------|
| Pulgar      | 163   | +6     | 31       | -5     |
| Índice      | 139   | +18    | 32       | -3     |

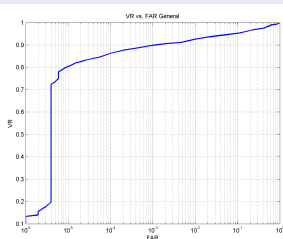
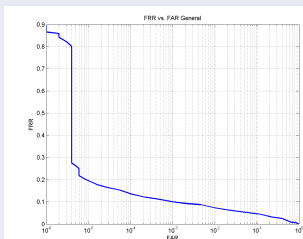
## Extracción de características (Morpho)



## Clasificador (Morpho)

Retorna un score relacionado con el FAR mediante:

$$s = 3500 - 500 \log(FAR/10^{-4})$$



|                                    | FAR=10e-3 | FAR=10e-5 |
|------------------------------------|-----------|-----------|
| Score deriv. fórmula               | 3000      | 4000      |
| Score deriv. análisis en base DNIC | 3125      | 4370      |

## Relación entre matcheo y calidad

Se separan los dedos en tres calidades: buena, regular y mala

| (FAR=10e-3         | VR   | FRR   |
|--------------------|------|-------|
| mala vs mala       | 0.40 | 0.60  |
| regular vs regular | 0.91 | 0.09  |
| buena vs buena     | 0.99 | 0.008 |

## Identificación

Dos contextos bien distintos:

- Enrolamiento: se busca si la persona ya no está en la base de datos
- Identificación durante un procedimiento policial o judicial (indocumentado, NN, etc.)

En cualquier caso, el costo de NO encontrarlo si está es altísimo: se va a permitir un FAR no muy restrictivo

## Desafíos para los próximos años

- Toma de huellas a bebés
- Base de huellas dactilares de 5 años
- Servicio de verificación de identidad Web
  - Problema: No se controla el input y por lo tanto la calidad
- Documento electrónico: Aplicación Match-On-Card
  - Se realiza a nivel del template
  - El FAR se determina al momento de personalizar el documento
  - El FAR actual es de 0.01
- Reconocimiento facial

## Reconocimiento Facial

- Problemas del reconocimiento facial:
  - Gran variabilidad intra-clase
  - Similitud inter-clase



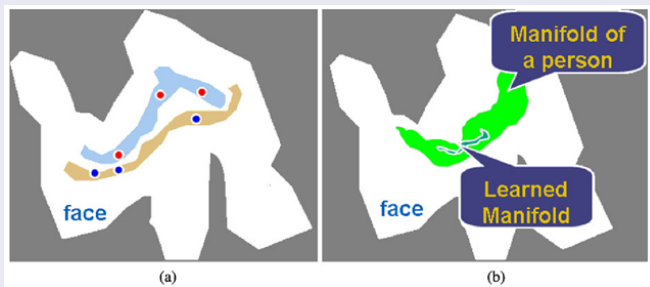
(a)



(b)

## Reconocimiento Facial

- Problemas para caracterizar bien cada clase:





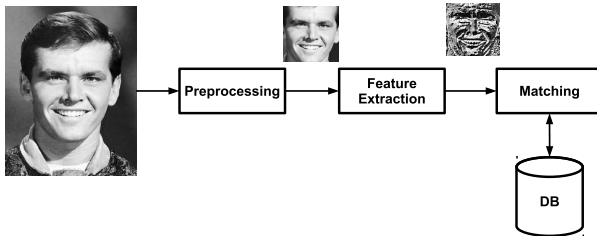
## Reconocimiento Facial

- Muy pocas muestras por clase, la mayoría tiene una única imagen (1.746.189, aprox. 49%)
- Envejecimiento entre toma y toma (10 años en el caso normal)
- Pose e iluminación controlados
- Gran base de datos (agosto 2015):
  - 3.700.966 de registro de personas con al menos una foto
  - 7.352.339 de fotos en total

## Preguntas ?

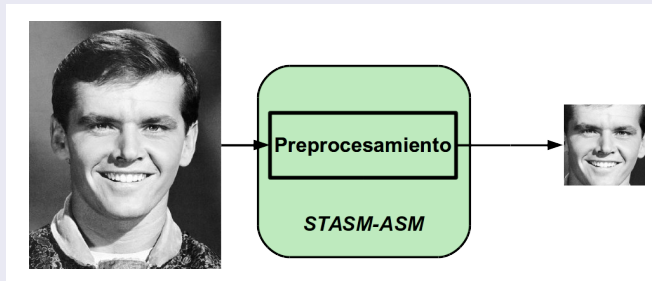
Javier Preciozzi: [jpreciozzi@dnic.gub.uy](mailto:jpreciozzi@dnic.gub.uy)

## Módulos de un sistema de reconocimiento facial

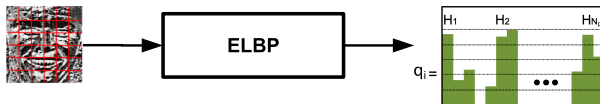
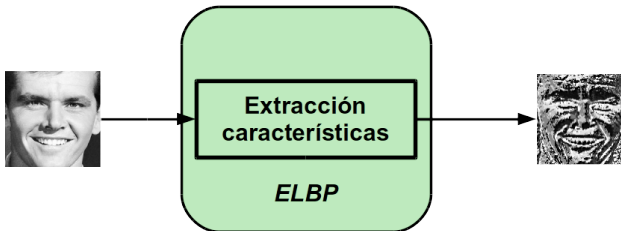


## Pre-procesamiento

- Sensores: Cámaras fotográficas estándar (Olympus SP 350, Canon SX 110 IS, Kodak C533)
- Extracción de características: segmentación



## Extracción de características



## Criterio de decisión

- Criterio basado en un modelo *a contrario*
- Idea: en lugar de buscar si alguien es un usuario genuino, se decide que es altamente probable que NO SEA impostor mediante la hipótesis *a contrario*
- Ventajas: evita la estimación de la distribución de probabilidad del evento esperado

## Ejemplo

- Supongamos una base con 1000 personas
  - La clase impostor tiene 999000 elementos
  - La clase genuinos tiene 1000
- Por lo tanto será siempre mejor trabajar con la clase impostor que con la de genuinos

- Sea  $H_0$  el modelo de fondo.
- $e$  es la observación de una variable aleatoria  $E$ .

## Número de falsas alarmas

El ( $NFA$ ) se define como la esperanza del número de ocurrencias del evento ( $e$ ) bajo el modelo de fondo  $H_0$ .

Un evento se dice que es  $\varepsilon$  – *significativo* si  $NFA(e) < \varepsilon$ .

- El NFA puede ser definida como sigue:

$$NFA'(E) = \mathcal{N} \cdot P(E \geq E|H_0) \quad (1)$$

## Modelo *a contrario*

- Hipótesis  $\mathcal{H}_0$ : Las caras corresponden a diferentes personas
- Persona de consulta  $q$
- Persona en la base de datos  $g_i$
- Distancia asociada:  $D(q, g_i)$
- Probabilidad que la distancia entre  $q, g_i$  sea menor a un  $\delta$  en la hipótesis  $\mathcal{H}_0$ :

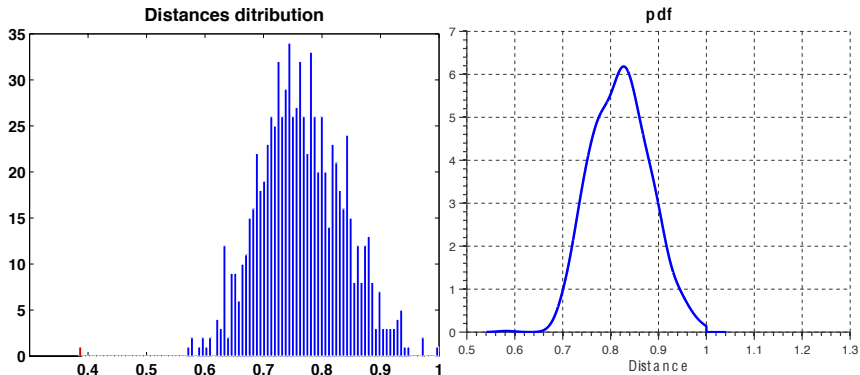
$$P(D(q, g_i) \leq \delta | \mathcal{H}_0) = \int_{-\infty}^{\delta} p_{q_i | \mathcal{H}_0}(x) dx$$

- donde  $p_{q_i | \mathcal{H}_0}$  es la función de densidad de probabilidad de las distancias de la persona  $q$  bajo la hipótesis  $\mathcal{H}_0$



# Introducción a la biometría - Caso de Uso: DNIC

- La  $pdf$   $p_{q_i|H_0}$  se calcula directamente de los datos



En nuestro caso el evento significativo es que dos caras  $(q, g_j)$  corresponden a la misma persona. El  $NFA$  se define entonces como:

$$NFA(q_i, g_j) = N_Q(N_G - 1)P(D(q_i, g_j) \leq \delta | H_0) \quad (2)$$

- donde  $N_Q$  and  $N_G$  son los tamaños del conjunto galería y de consulta respectivamente, y por lo tanto ,  $N_Q(N_G - 1)$  corresponde a todas las posibles configuraciones del evento anterior

## Ventajas de usar NFA

La definición del umbral  $\varepsilon$  en el modelo *a contrario* tiene varias ventajas:

- **Permite controlar la performance del sistema desde el inicio**
- **El umbral se adapta automaticamente a la base de datos**

