



Selección efectiva de características wavelet en la identificación de bioseñales 1-D y 2-D usando algoritmos genéticos

Mauricio Orozco Alzate

Director: Prof. Germán Castellanos Domínguez

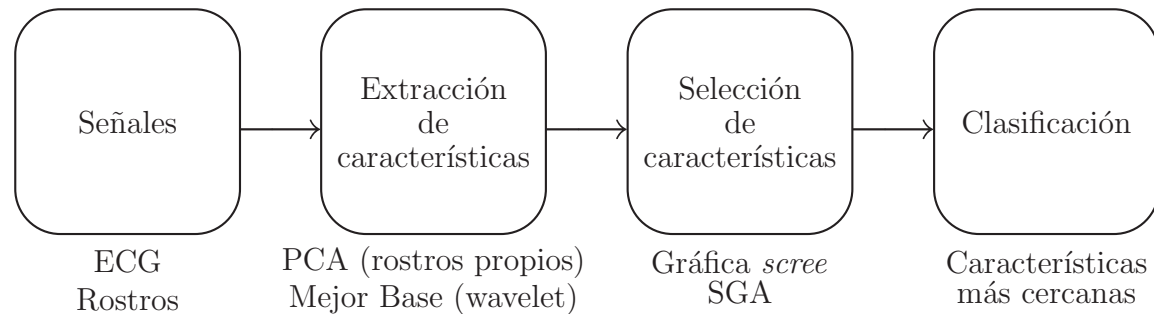
Fecha: Junio 9, 2005

MAESTRÍA EN AUTOMATIZACIÓN INDUSTRIAL
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA Y COMPUTACIÓN
UNIVERSIDAD NACIONAL DE COLOMBIA
SEDE MANIZALES



Introducción

Sistema de clasificación:



Planteamiento del problema:

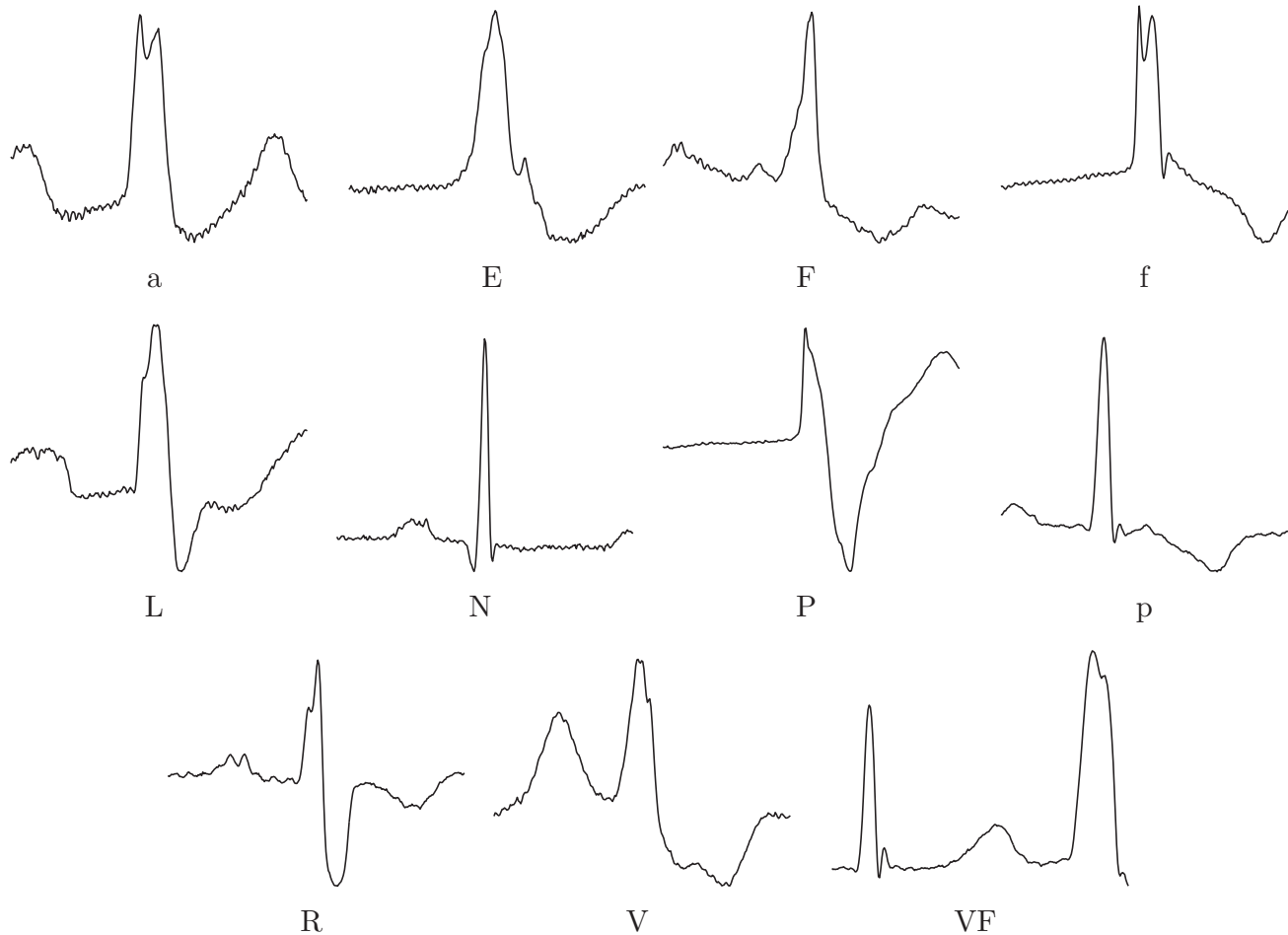
Se deben estudiar alternativas para mejorar el desempeño del sistema de clasificación en cada una de sus etapas:

- *Extracción de características*: se deben estudiar otros métodos, por ejemplo wavelets y compararlos con PCA.
- *Selección de características* se requiere un método automático para encontrar las características más discriminantes.
- *Algoritmos genéticos*: Se requiere un método automático para sintonizar los parámetros de control.
- *Clasificación*: Se requieren clasificadores de naturaleza no paramétrica y gran efectividad con el fin de aplicarlos a un amplio rango de señales.



Señales 1-D y 2-D

Arritmias (1-D): Subconjunto de la base de datos de arritmias MIT-BIH con 100 latidos de cada una de las siguientes clases:





Señales 1-D y 2-D (cont.)

Rostros (2-D): Base de datos ORL. Posee imágenes de 40 individuos, cada uno con 10 fotografías. Las imágenes tienen variaciones de iluminación, expresiones y detalles faciales.





Extracción de características

Análisis de Componentes Principales (PCA): Se denomina también *transformada KL*. Extracción de características para la base de datos de arritmias.

Rostros propios: Modificación de PCA. La diferencia radica en la construcción de la matriz de covarianza; se usa $\Sigma'_{\bar{X}} = \bar{X}^T \bar{X}$ en lugar de $\Sigma_{\bar{X}} = \bar{X} \bar{X}^T$, donde $\bar{X} = [\mathbf{x}_1 - \mu_{x_1} \cdots \mathbf{x}_N - \mu_{x_N}]$.



Extracción de características (cont.)

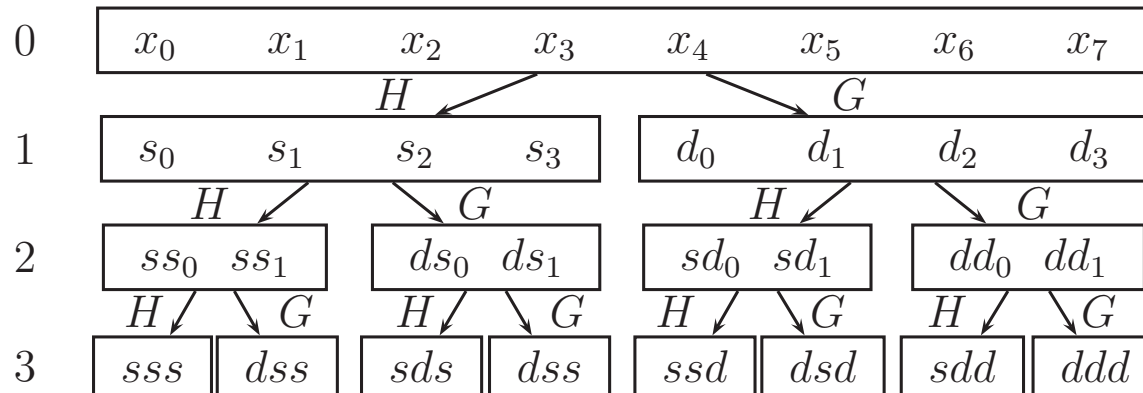
Análisis Wavelet Packet Discreto:

- Wavelet Packets:

$$\psi_{sfp} \stackrel{\text{def}}{=} s^{-s/2} \psi_f(s^{-s}t - p); \quad s : \text{escala}, f : \text{frecuencia}, p : \text{posición.}$$

- Descomposición:

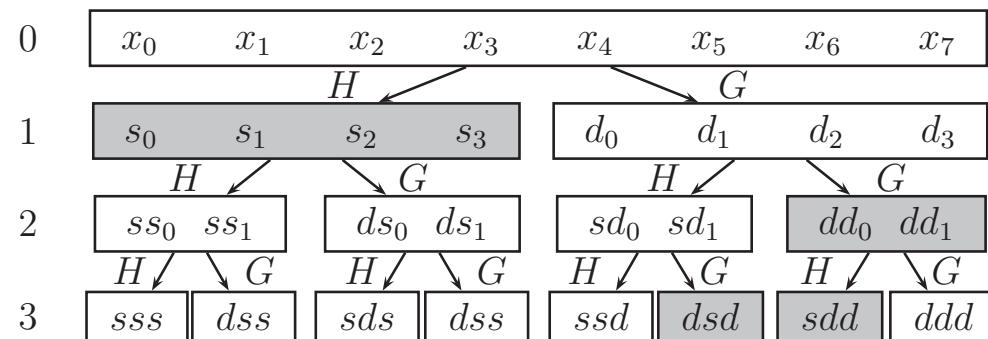
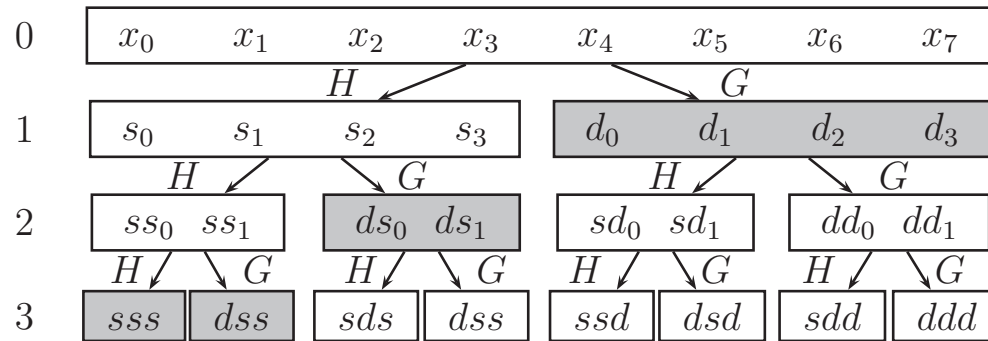
$$\lambda_{sf}(p) \stackrel{\text{def}}{=} \langle x, \psi_{sfp}(-t) \rangle = \int_{\mathbb{R}} x^*(t) 2^{-s/2} \psi_f(p - 2^{-s}t) dt.$$



Extracción de características (cont.)

Búsqueda de la Mejor Base: de acuerdo con un funcional de costo, por ejemplo:

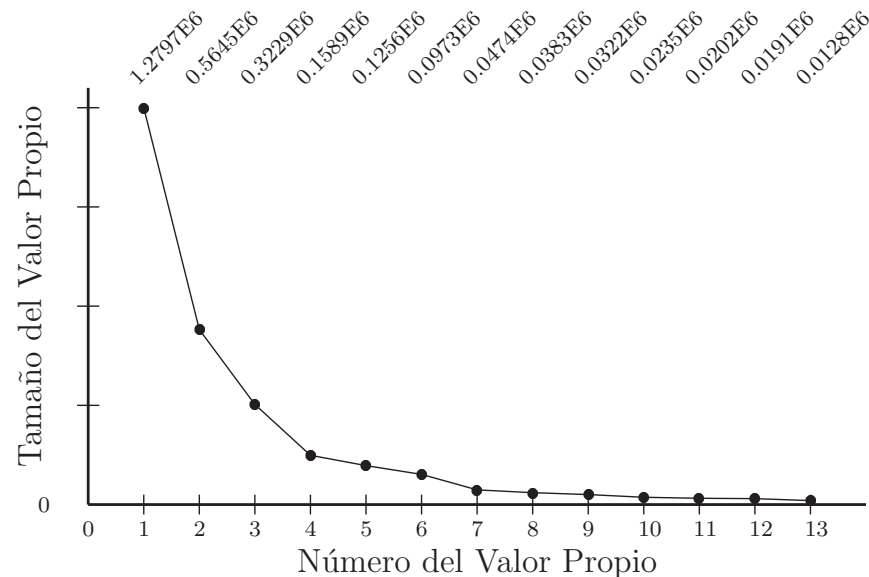
- Valor sobre un umbral.
- Concentración en ℓ^p .
- Entropía.
- Logaritmo de la energía.



Selección de características (cont.)

Retención de componentes principales:

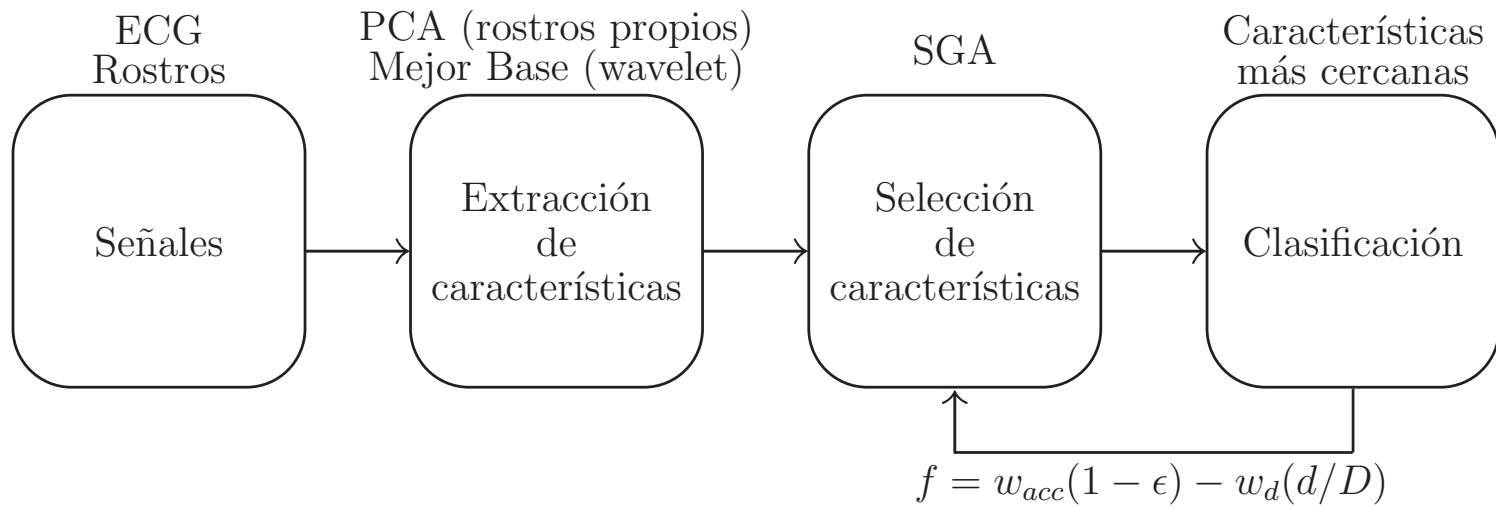
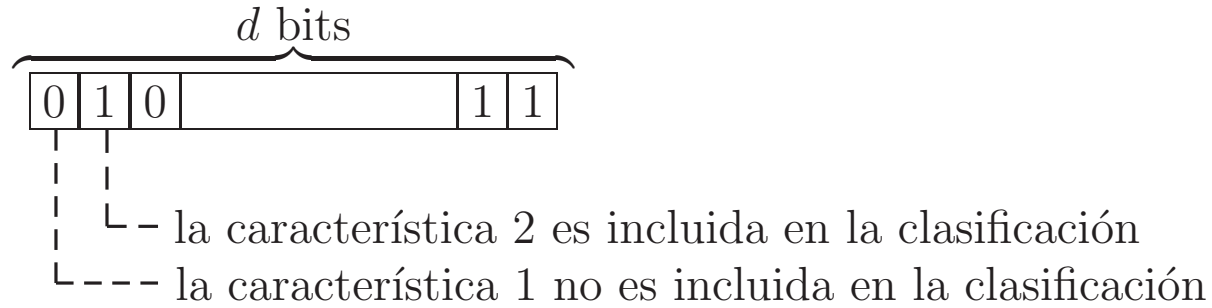
1. Porcentaje especificado del total de la varianza; e.g. 80%.
2. Componentes cuyos valores propios sean mayores que el promedio de todos los valores propios.
3. Buscar el punto de quiebre entre los valores propios grandes y los valores propios pequeños en la gráfica *scree*.
4. Evaluar la significación de las componentes más grandes.
- ✓ 5. Incluir secuencialmente cada componente principal y encontrar el mejor desempeño del clasificador.





Selección de características (cont.)

Algoritmo Genético Simple (SGA):





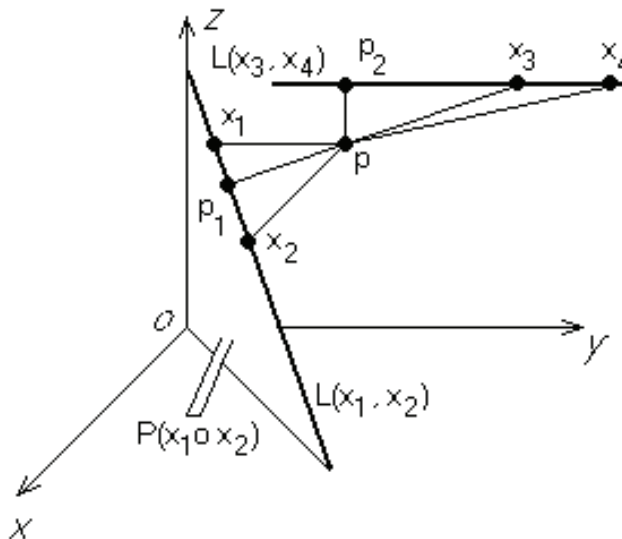
Clasificadores de características más cercanas


Naturaleza no paramétrica y gran efectividad. Se usaron los siguientes:

k vecinos más cercanos (k -NN): [Cover and Hart, 1967].

k líneas más cercanas (k -NFL): Extensión del clasificador de la línea de características más cercanas (NFL) propuesto en [Li and Lu, 1999] para clasificación por voto mayoritario.

k planos más cercanos (k -NFP): Extensión del clasificador del plano de características más cercano (NFP) [Chien and Wu, 2002] para clasificación por voto mayoritario.





Comparación experimental de los clasificadores de características más cercanas

Diferencias teóricas: Se espera que el clasificador k -NFL capture más variaciones que el clasificador k -NN y que el clasificador k -NFP capture más variaciones que el clasificador k -NFL, debido que:

$$\begin{aligned}d(\mathbf{z}, F_{ijm}^c) &\leq \min(d(\mathbf{z}, L_{ijm}^c), d(\mathbf{z}, L_{jm}^c), d(\mathbf{z}, L_{mi}^c)) \\ &\leq \min(d(\mathbf{z}, \mathbf{z}_{ci}), d(\mathbf{z}, \mathbf{z}_{cj}), d(\mathbf{z}, \mathbf{z}_{cm})).\end{aligned}$$

Condiciones de prueba: Se propone

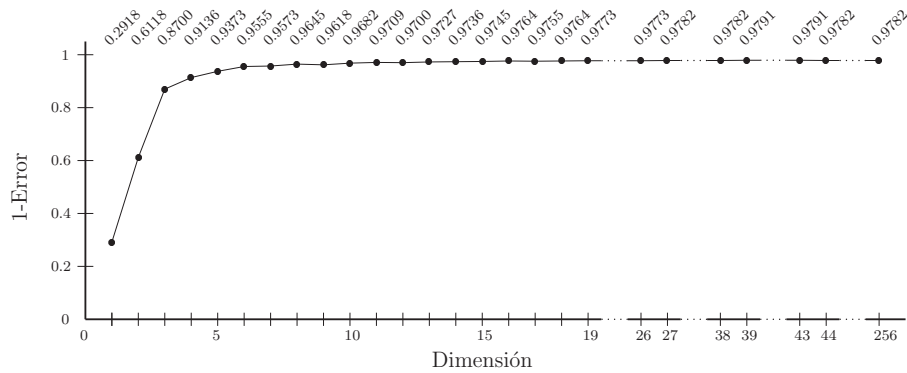
- Validación cruzada de 10 particiones.
- $k = 5$ para todos los clasificadores.
- Extracción de características con PCA (rostros propios).
- Prueba binomial al 5% de significación:

$$H_0 : p = q = 0.5; \quad A \text{ y } B \text{ tienen la misma precisión,}$$

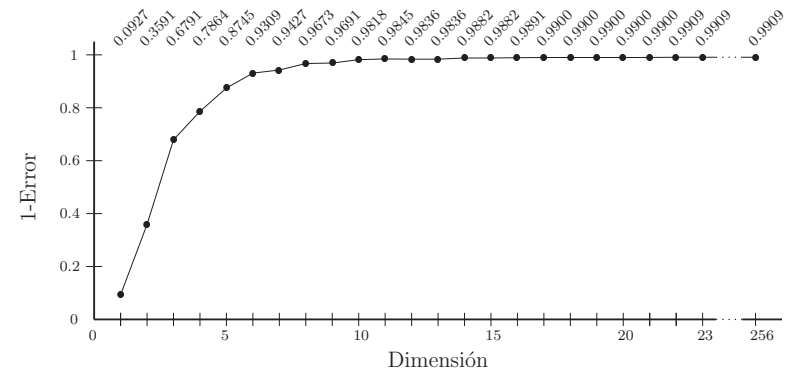
$$H_1 : p > 0.5; \quad A \text{ es mejor que } B.$$

Comparación experimental de los clasificadores de características más cercanas (cont.)

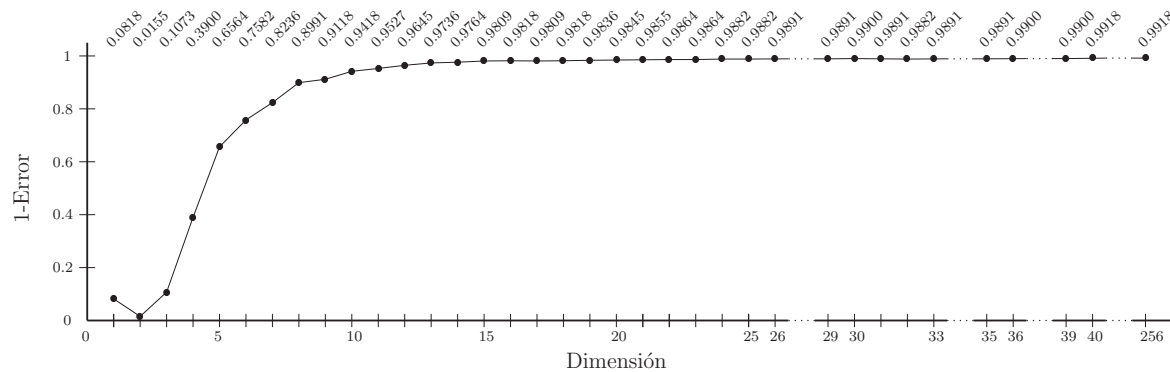
Base de datos de arritmias: comparación de precisión por dimensión.



(c) k -NN



(d) k -NFL



(e) k -NFP

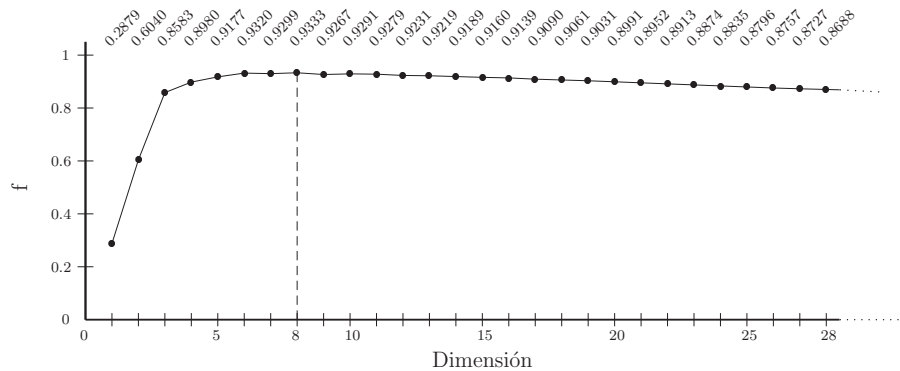


Comparación experimental de los clasificadores de características más cercanas (cont.)

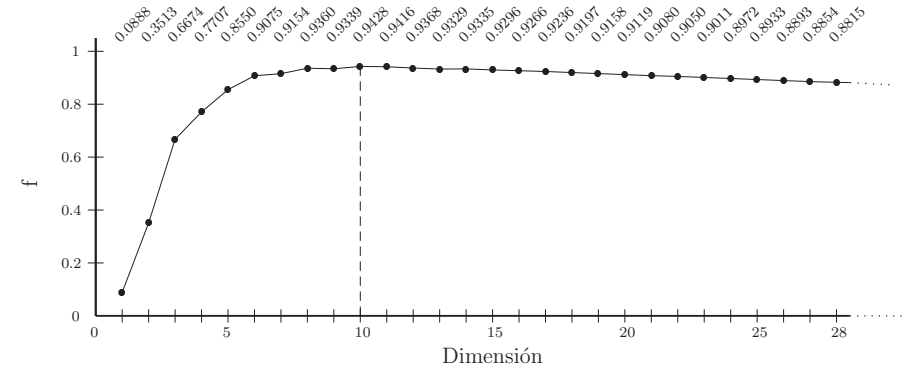
Base de datos de arritmias: comparación del compromiso entre la precisión y la dimensión.

Se propone la función:

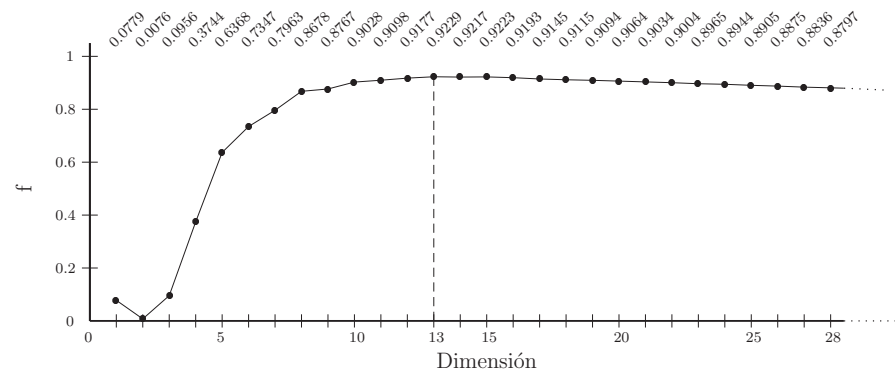
$$f = w_{acc}(1 - \epsilon) - w_d(d/D); \quad w_{acc}, w_d = 1$$



(a) k -NN



(b) k -NFL



(c) k -NFP



Comparación experimental de los clasificadores de características más cercanas (cont.)

Base de datos de arritmias: comparación mediante la prueba binomial:

Dimensión	n	s	Valor p (unilateral)	Rechaza H_0
1	375	78	1.0000	No
2	420	71	1.0000	No
3	276	33	1.0000	No
4	208	34	1.0000	No
5	131	31	1.0000	No
6	75	24	0.9995	No
7	58	21	0.9876	No
8	37	20	0.3714	No
9	42	25	0.1400	No
10	27	21	0.0030	Si
11	23	19	0.0013	Si
12	23	19	0.0013	Si
13	24	18	0.0113	Si

(a) k -NFL vs k -NN

Dimensión	n	s	Valor p (unilateral)	Rechaza H_0
1	158	73	0.8495	No
2	404	13	1.0000	No
3	705	38	1.0000	No
4	514	39	1.0000	No
5	304	32	1.0000	No
6	236	23	1.0000	No
7	159	14	1.0000	No
8	103	14	1.0000	No
9	91	14	1.0000	No
10	58	7	1.0000	No
11	49	7	1.0000	No
12	37	8	0.9999	No
13	27	8	0.9904	No
:	:	:	:	:
39	11	5	0.7256	No
40	9	5	0.5000	No
41	9	5	0.5000	No
:	:	:	:	:

(b) k -NFP vs k -NFL

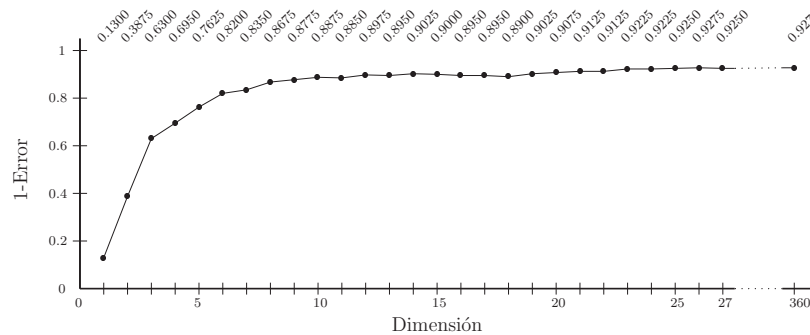
Dimensión	n	s	Valor p (unilateral)	Rechaza H_0
1	359	64	1.0000	No
2	674	9	1.0000	No
3	891	26	1.0000	No
4	608	16	1.0000	No
5	341	16	1.0000	No
6	253	18	1.0000	No
7	183	18	1.0000	No
8	116	22	1.0000	No
9	95	20	1.0000	No
10	65	18	0.9999	No
11	52	16	0.9984	No
12	50	22	0.8389	No
13	39	20	0.5000	No
14	35	19	0.3679	No
15	29	18	0.1325	No
16	26	16	0.1635	No
17	28	17	0.1725	No
18	26	16	0.1635	No
19	25	16	0.1148	No
20	24	16	0.0758	No
21	25	17	0.0539	No
22	24	17	0.0320	Si
23	24	17	0.0320	Si
24	24	18	0.0113	Si
25	24	18	0.0113	Si

(c) k -NFP vs k -NN

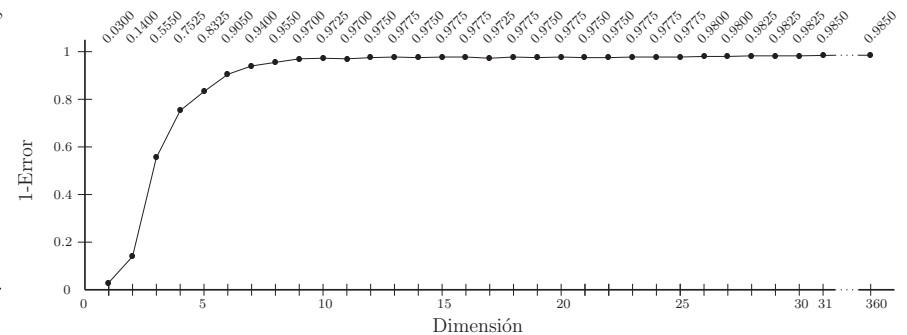


Comparación experimental de los clasificadores de características más cercanas (cont.)

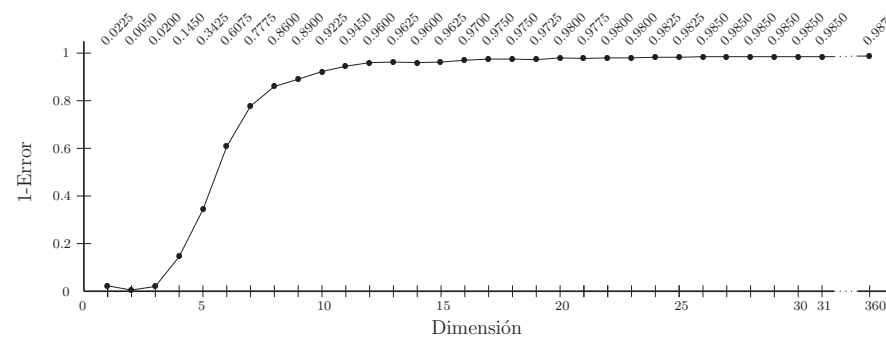
Base de datos de rostros: comparación de precisión por dimensión.



(d) k -NN



(e) k -NFL



(f) k -NFP

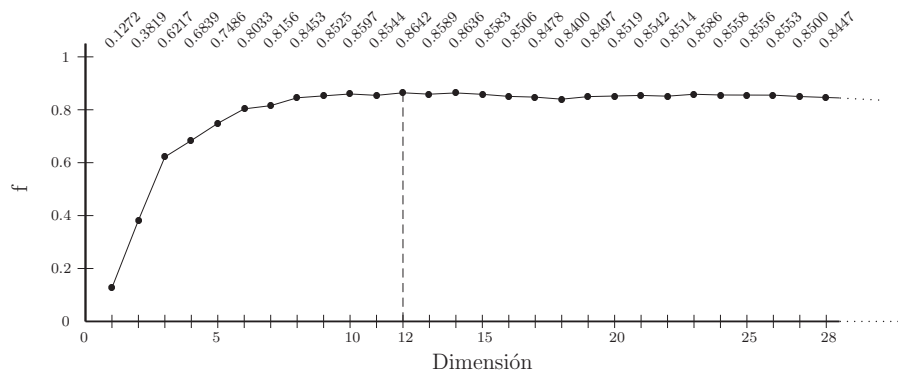


Comparación experimental de los clasificadores de características más cercanas (cont.)

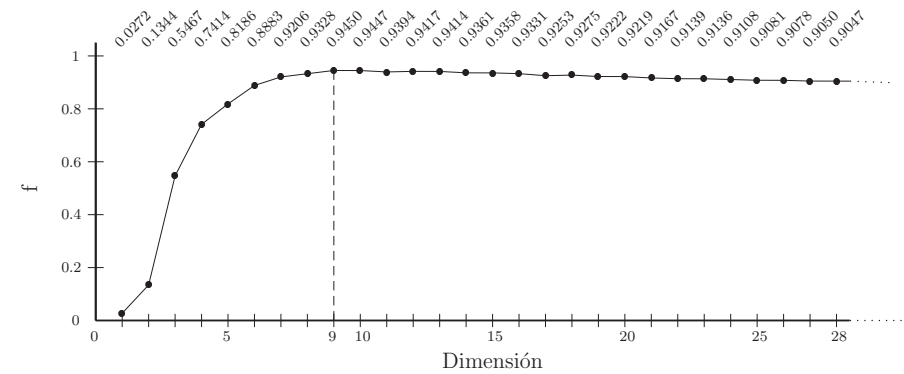
Base de datos de rostros: comparación del compromiso entre la precisión y la dimensión.

Se propone la función:

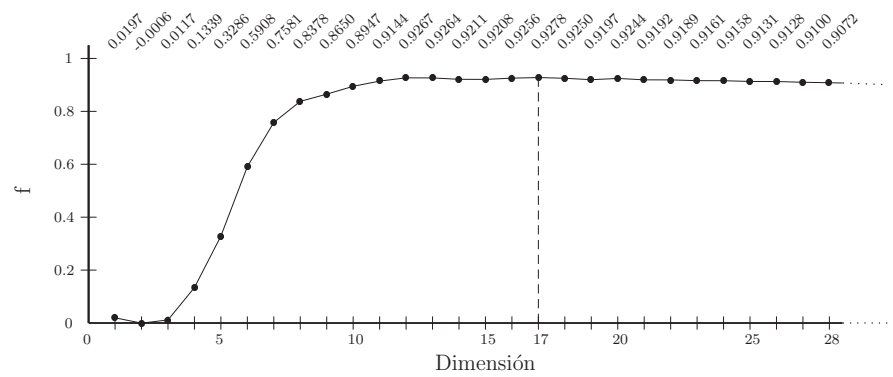
$$f = w_{acc}(1 - \epsilon) - w_d(d/D); \quad w_{acc}, w_d = 1$$



(a) k -NN



(b) k -NFL



(c) k -NFP



Comparación experimental de los clasificadores de características más cercanas (cont.)

Base de datos de rostros: comparación mediante la prueba binomial:

Dimensión	n	s	Valor p (unilateral)	Rechaza H_0
1	19	8	0.8204	No
2	56	1	1.0000	No
3	222	4	1.0000	No
4	255	6	1.0000	No
5	206	5	1.0000	No
6	129	5	1.0000	No
7	81	8	1.0000	No
8	50	6	1.0000	No
9	34	1	1.0000	No
10	24	2	1.0000	No
11	16	3	0.9979	No
12	10	2	0.9893	No
13	8	1	0.9961	No
14	8	1	0.9961	No
15	8	1	0.9961	No
16	7	2	0.9375	No
17	7	4	0.5000	No

Dimensión	n	s	Valor p (unilateral)	Rechaza H_0
1	59	8	1.0000	No
2	153	0	1.0000	No
3	256	6	1.0000	No
4	238	9	1.0000	No
5	202	17	1.0000	No
6	135	25	1.0000	No
7	89	33	0.9947	No
8	67	32	0.6873	No
9	67	36	0.3127	No
10	54	34	0.0380	Si
11	46	35	0.2678E-3	Si
12	41	33	0.0561E-3	Si
13	39	33	0.0071E-3	Si

Dimensión	n	s	Valor p (unilateral)	Rechaza H_0
1	56	8	1.0000	No
2	135	18	1.0000	No
3	130	50	0.9968	No
4	97	60	0.0125	Si
5	84	56	0.0015	Si
6	62	48	0.0871E-4	Si
7	58	50	0.0001E-4	Si

(a) k -NFL vs k -NN

(b) k -NFP vs k -NFL

(c) k -NFP vs k -NN



Sintonización automática de los parámetros de control del SGA

Parámetros:

- Tamaño de la población (r)
- Probabilidad o tasa de mutación (μ)
- Probabilidad o tasa de cruzamiento (χ)

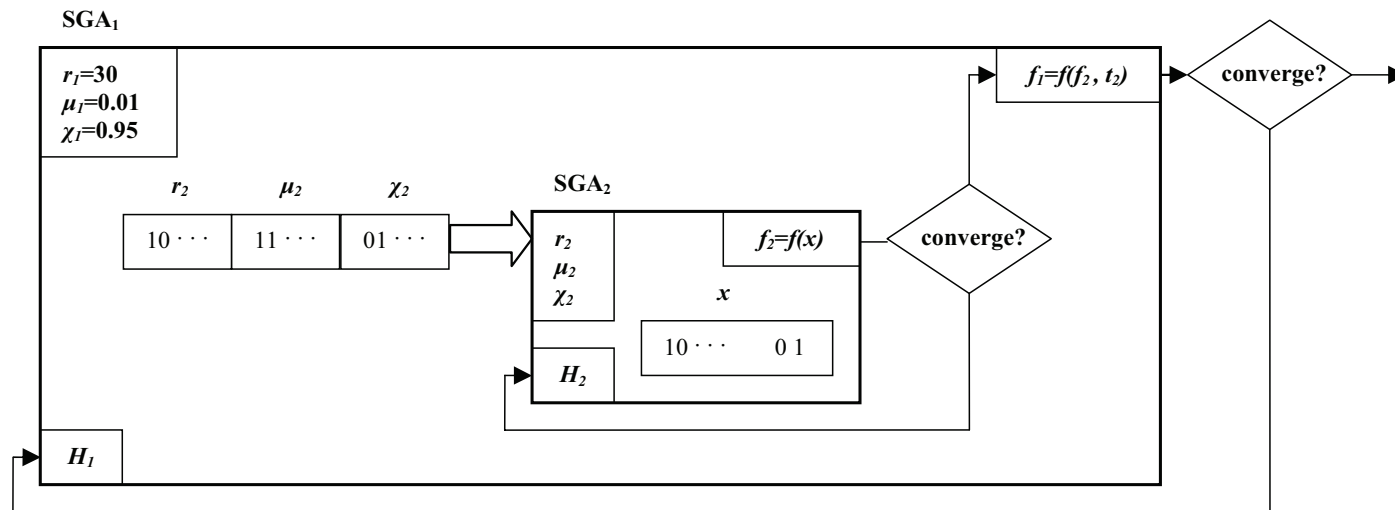
Funciones de evaluación comparativa:

Nombre	Función	Límites
F1	$f_1 = \sum_{i=1}^2 x_i^2$	$-5.12 \leq x_i \leq 5.12$
F2	$f_2 = 100(x_1^2 - x_2)^2 + (1 - x_1)^2$	$-2.048 \leq x_i \leq 2.048$
F3	$f_3 = \sum_{i=1}^5 \text{int.}(x_i)$	$-5.12 \leq x_i \leq 5.12$
F4	$f_4 = \sum_{i=1}^{30} (ix_i^4 + \text{Gauss}(0, 1))$	$-1.28 \leq x_i \leq 1.28$
F5	$f_5(x_i) = 0.002 + \sum_{j=1}^{25} \left(\frac{1}{j} + \sum_{i=1}^2 (x_i - a_{ij})^6 \right)$	$-65536 \leq x_i \leq 65536$
F6	$f_6 = 10V + \sum_{i=1}^{10} (-x_i \sin \sqrt{ x_i }), \quad V = 4189.829101$	$-500 \leq x_i \leq 500$
F7	$f_7 = 20A + \sum_{i=1}^{20} (x_i^2 - 10 \cos(2\pi x_i)), \quad A = 10$	$-5.12 \leq x_i \leq 5.12$
F8	$f_8 = 1 + \sum_{i=1}^{10} \left(\frac{x_i^2}{4000} - \prod_{i=1}^{10} \left(\cos \left(\frac{x_i}{\sqrt{i}} \right) \right) \right)$	$-600 \leq x_i \leq 600$

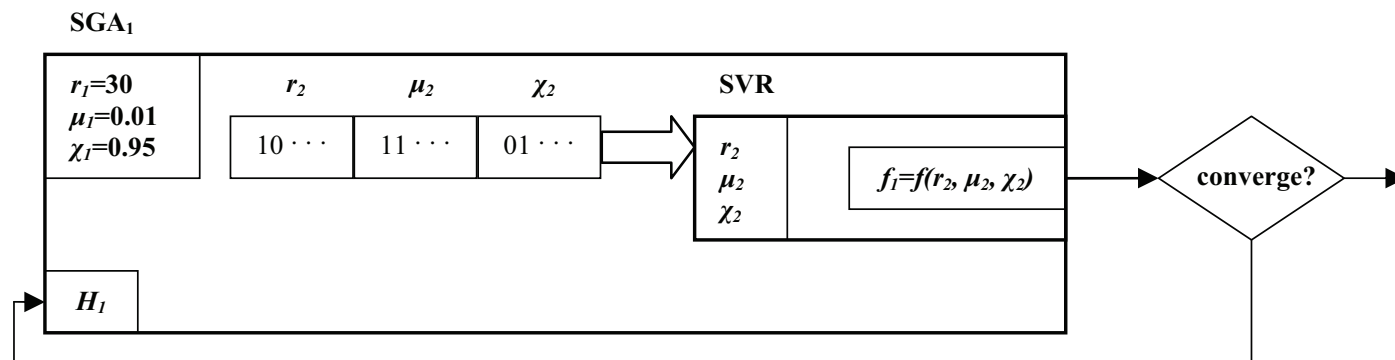


Sintonización automática de los parámetros de control del SGA (cont.)

Meta-algoritmo genético: El cromosoma de un SGA codifica los parámetros de otro. El segundo ejecuta la aplicación específica.



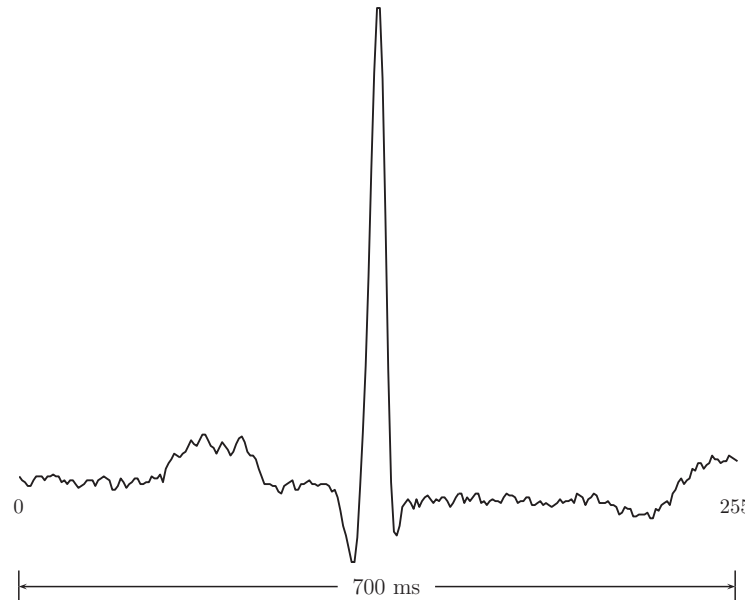
La SVR realiza una regresión para reemplazar el segundo SGA.





Clasificación de arritmias usando algoritmo de la mejor base + SGA

Extracción de características wavelet:



S=0:	0 F=0 255							
S=1:	0 F=0 127				128 F=1 255			
S=2:	0 F=0 63		64 F=1 127		128 F=2 191		192 F=3 255	
	⋮							
S=3:	F=0	F=1	F=2	F=3	F=4	F=5	F=6	F=7

- Función de costo de entropía (logaritmo de la entropía)
- Se propone realizar una búsqueda dentro del diccionario completo de descomposición (modificación de la metodología de rastreo de base)

Clasificación de arritmias usando algoritmo de la mejor base + SGA (cont.)

Selección de características:

- Parámetros de control del SGA: $r = 16$, $\mu = 0.01$, $\chi = 0.9$ y 50 generaciones
- Inicialización de población aleatoria, selección de ruleta, cruzamiento simple, mutación por inversión de bit y elitismo.
- Función objetivo:

$$f = w_{acc}(1 - \varepsilon) - w_d(d/D), \quad w_{acc}, w_d = 1$$

Filtro	Dimensión	Precisión	f	Tiempo
C06	4	0.884304	0.868679	8 h 47 min 44 s
C12	5	0.863153	0.843622	8 h 53 min 54 s
C18	7	0.905313	0.877969	8 h 50 min 40 s
C24	6	0.882613	0.859175	9 h 3 min 36 s
C30	6	0.865512	0.842074	8 h 27 min 48 s
D02	5	0.896591	0.877060	8 h 53 min 6 s
D04	4	0.893395	0.877770	8 h 53 min 41 s
D06	4	0.903395	0.887770	8 h 45 min 18 s
D08	4	0.902486	0.886861	9 h 4 min 32 s
D10	4	0.904304	0.888679	8 h 52 min 31 s
D12	4	0.902486	0.886861	8 h 46 min 43 s
D14	4	0.896122	0.880497	8 h 37 min 43 s
D16	4	0.890668	0.875043	9 h 4 min 51 s
D18	4	0.894304	0.878679	8 h 48 min 55 s
D20	4	0.902486	0.886861	11 h 14 min 20 s

NOTA: C: Coifman, D: Daubechies.



Reconocimiento de rostros usando algoritmo de la mejor base + SGA

Extracción de características wavelet: Si se realiza la descomposición wavelet packet completa, el espacio de búsqueda es demasiado grande: $2^{128} \times 2^{128} = 16384$. En consecuencia, para reducir el espacio de búsqueda del SGA se propone utilizar el algoritmo de la mejor base para calcular la transformada K-L aproximada \implies *cálculo rápido de rostros propios*, reduciendo la dimensión hasta 360 características.

- Complejidad de SVD: $O(d^3)$
- Complejidad del método de mejor base (K-L) aproximado: $O(d^2 \log d)$

Selección de características:

- Parámetros de control del SGA: $r = 16$, $\mu = 0.01$, $\chi = 0.9$ y 50 generaciones
- Inicialización de población aleatoria, selección de ruleta, cruzamiento simple, mutación por inversión de bit y elitismo.

21



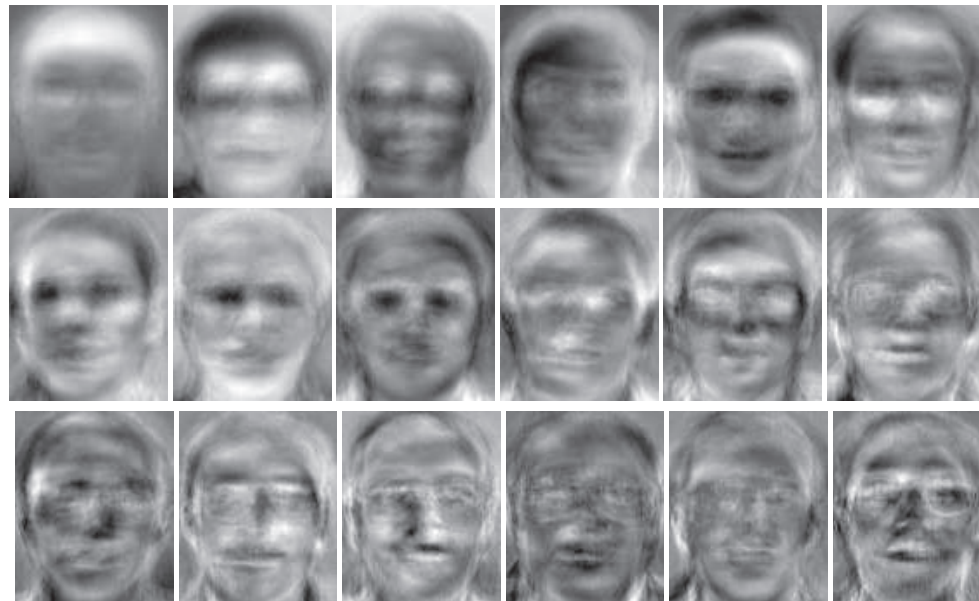
Reconocimiento de rostros usando algoritmo de la mejor base + SGA (cont.)

Selección de características:

- Función objetivo:

$$f = w_{acc}(1 - \varepsilon) - w_d(d/D), \quad w_{acc}, w_d = 1$$

- Mejor individuo: 1111011110111000011010000111010...0, con $f = 0.872361$ y precisión de 0.9225





Reconocimiento de rostros usando algoritmo de la mejor base + SGA (cont.)

Tanto la precisión como el valor de la función f fueron superiores a los obtenidos con la selección secuencial utilizada en el cálculo convencional de rostros propios.

Asimismo, los resultados para esta misma cadena binaria fueron superiores utilizando los otros dos clasificadores:

Clasificador	f	Precisión
k -NN	0.8724	0.9225
k -NFL	0.9451	0.9952
k -NFP	0.9320	0.9821

Estos resultados demuestran que el SGA es una herramienta adecuada para realizar la selección de características y que las componentes principales en su orden natural no corresponden necesariamente a las componentes más discriminantes.



Conclusiones

- La extracción de características wavelet mediante la técnica modificada de rastreo de base mejora la representación de la señal en el sentido de la clasificación. Esto se debe a que al buscar en el árbol completo de descomposición no se eliminan bloques que pueden contener información discriminante.
 - El algoritmo de la mejor base resulta útil para el reconocimiento de rostros como un método para el cálculo rápido de rostros propios.
 - Los resultados obtenidos con los filtros Coifman y Daubechies presentaron resultados inferiores a los obtenidos con análisis de componentes principales. En consecuencia, el desarrollo de un filtro ortogonal adecuado a cada tipo de señal es muy importante para lograr mejores resultados en la extracción de características wavelet.



Conclusiones (cont.)

- La selección de características con SGA permite encontrar las características más discriminantes; sin embargo, es un método computacionalmente costoso.
 - Este método permite caracterizar e interpretar las componentes principales. Es posible, por ejemplo, darle una interpretación subjetiva al contenido codificado en cada uno de los rostros propios asociados a la cadena binaria seleccionada por el SGA.
 - Se planteó un método de sintonización automática de parámetros de control del SGA, usando SVR y funciones de evaluación comparativa.



Conclusiones (cont.)

- Además, se extendieron los clasificadores NFL y NFP a k -NFL y k -NFP respectivamente.
 - Se demostró experimentalmente que el clasificador NFL es el más adecuado, considerando la precisión, la cantidad de información (varianza) necesaria y el número de cálculos de distancia.
 - Aunque las relaciones teóricas anticipan un mejor funcionamiento del clasificador NFP, las pruebas binomiales demostraron que la superioridad del clasificador NFP no es estadísticamente significativa. En consecuencia, el aumento de la cantidad de cálculos de distancia no se ve compensado con un aumento significativo de la precisión.



Aportes

- Extensión de los clasificadores NFL y NFP a k -NFL y k -NFP para clasificación por voto mayoritario.
- Se propuso una función para ponderar la precisión de clasificación con la dimensión del subconjunto de características.
- Se propuso una modificación del método de rastreo de base para la extracción de características wavelet.
- Se propuso un método de sintonización automática de los parámetros de control del SGA.



Discusión académica

Resultados parciales de esta tesis de maestría se presentaron en:

- **IX Simposio de Tratamiento de Señales, Imágenes y Visión Artificial:**

M. Orozco, “Optimización de los parámetros de control del algoritmo genético simple usando regresión de vectores soporte,” en *IX Simposio de Tratamiento de Señales, Imágenes y Visión Artificial*, Manizales, Colombia, Septiembre 2004, ISBN 958-9322-91-3.

- **The 9th IASTED International Conference on Artificial Intelligence & Soft Computing —ASC 2005—:**

M. Orozco and G. Castellanos, “Comparison of the nearest feature classifiers for face recognition,” in *The 9th IASTED International Conference on Artificial Intelligence & Soft Computing (ASC 2005)*, Benidorm, Spain, September 2005. **(ACEPTADO)**



Trabajo futuro

- Estudio de los más recientes clasificadores de características más cercanas: clasificador del espacio de características más cercano (NFS) [Chien and Wu, 2002] y el clasificador del espacio intra-clase más cercano [Liu et al., 2004].
- Metodología de sintonización automática de parámetros de control del SGA (usando funciones de evaluación comparativa): Jorge Alberto Jaramillo y Mauricio Orozco Alzate.
- Aplicación del sistema con las imágenes de la base de datos de patologías de LPH (labio y paladar hendido).
- Construcción de wavelets adaptativas para extracción de características faciales.

Referencias

- [Chien and Wu, 2002] Chien, J.-T. and Wu, C.-C. (2002). Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(12):1644–1649.
- [Cover and Hart, 1967] Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory*, 13(1):21–27.
- [Li and Lu, 1999] Li, S. Z. and Lu, J. (1999). Face recognition using the nearest feature line method. *IEEE Trans. Neural Networks*, 10(2):439–443.
- [Liu et al., 2004] Liu, W., Wang, Y., Li, S. Z., and Tan, T. (2004). Nearest intra-class space classifier for face recognition. In *Proc.*

*of the 17th International Conference on Pattern Recognition
(ICPR'04).*