

REDUCCIÓN DEL VECTOR DE CARACTERÍSTICAS EN RECONOCIMIENTO FACIAL

Carlos M. Travieso⁽¹⁾, Jordi Solé-Casals⁽²⁾, Vladimir Zaiats⁽²⁾, Jesús B. Alonso⁽¹⁾, Miguel A. Ferrer⁽¹⁾
ctravieso@dsc.ulpgc.es, {jordi.sole, vladimir.zaiats}@uvic.cat, {jalonso,mferrer}@dsc.ulpgc.es

⁽¹⁾ Dpto. de Señales y Comunicaciones. Centro Tecnológico para la Innovación en Comunicaciones (CeTIC).
Universidad de Las Palmas de Gran Canaria. Campus de Tafira s/n, 35017, Las Palmas de Gran Canaria.

⁽²⁾ Grup de Processament de Senyal. Universitat de Vic. c/ de la Laura, 13. 08500. Vic. Barcelona.

Abstract- In this present work, we are proposing a characteristics reduction system for a facial biometric identification system, using transformed domains such as discrete cosine transformed (DCT) and discrete wavelets transformed (DWT) as parameterization; and Support Vector Machines (SVM) and Neural Network (NN) as classifiers. The size reduction has been done with Principal Component Analysis (PCA) and with Independent Component Analysis (ICA). This system presents a similar success results for both DWT-SVM system and DWT-PCA-SVM system, about 98%. The computational load is improved on training mode due to the decreasing of input's size and less complexity of the classifier.

I. INTRODUCCIÓN

Actualmente, en los tiempos que corren y después de los fatídicos atentados del 11 de Septiembre en Estados Unidos, han proliferado de manera exponencial todo tipo de normativas y legislación en cuanto a seguridad se refiere. Por ello, han aumentado la cantidad de entornos controlados que se pueden encontrar actualmente, como son aeropuertos, estaciones de tren y metro, pasos fronterizos entre países, edificios estatales, tanto públicos como privados, etc. Para controlar estos espacios se utilizan numerosos y distintos tipos de sistemas de reconocimiento biométrico [1] [2].

Uno de estos sistemas biométricos empleados es con caras. El problema del reconocimiento de caras ha llegado a ser uno de los mayores retos en el mundo del desarrollo tecnológico, debido principalmente a la importancia real que están alcanzando estas aplicaciones de reconocimiento, como son, vigilancia permanente, accesos de seguridad, sistemas de telecomunicaciones, medicina, etc.

Éste es un campo que se ha desarrollado muy rápidamente pero que no tiene límites en cuanto a la capacidad de éxito que poseen los diferentes sistemas, ya que la combinación de variables que se introducen en éstos puede llegar a ser muy grande. Por ello, se intentan mejorar cada vez más los sistemas existentes introduciendo nuevas características, e introduciendo nuevas líneas de trabajo que resulten válidas para el desarrollo de este tipo de sistemas [3].

El reconocimiento facial es un método no invasivo, lo que supone para estos sistemas una enorme ventaja con respecto a otros sistemas que requieren de la colaboración

guiada de los sujetos que forman parte de la base de datos y del momento de la captura de los datos.

En este trabajo hemos desarrollado un sistema de reconocimiento facial basado en dominios transformados de las imágenes, ya que caracteriza con igualdad las muestras intra-clase y con discriminación las inter-clase frente a las variaciones de posición y expresión de la cara. Debido al número elevado de características, se han empleado herramientas para reducirlas, con la finalidad de mantener las tasas de éxito del sistema, disminuyendo así, la complejidad del sistema.

II. BASE DE DATOS

La base de datos usada, consta de 400 imágenes, 10 imágenes por cada uno de los 40 usuarios. Las imágenes están tomadas con un fondo oscuro, posición frontal y con variaciones en la orientación de las mismas. La muestra total de la misma puede ser observada en la figura 1.

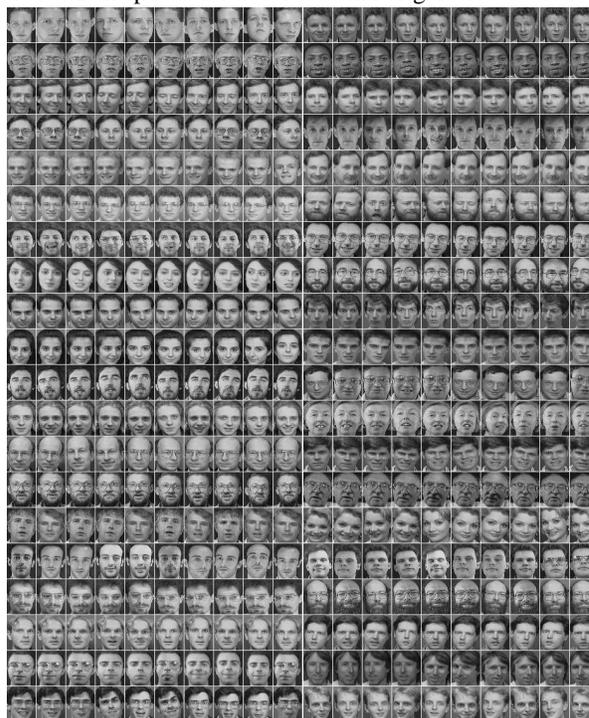


Fig. 1. Base de datos ORL (Olivetti Research Laboratory).

Esta base de datos presenta imágenes con variaciones gestuales como ojos abiertos-cerrados, sonrisa-sin sonrisa, gafas-no gafas y variaciones de iluminación. Estas variaciones de iluminación no están definidas. Todas las imágenes son en escala de gris de 256 valores y tienen un tamaño de 92 x 112 píxeles [4].

III. SISTEMAS DE PARAMETRIZACIÓN

Normalmente las imágenes se representan en escala de intensidad de grises. Pero la representación de las imágenes faciales puede ser más efectiva y eficiente aplicando transformaciones digitales a éstas [1][3], tales como la Transformada Rápida de Fourier (FFT), la Transformada Discreta del Coseno (DCT), la Transformada Discreta Wavelet (DWT), etc. En particular en este trabajo, se va trabajar con la Transformada Discreta Coseno, y con la Transformado Discreta Wavelet.

A. Transformada discreta coseno

La Transformada Discreta del Coseno es un caso especial de la Transformada Discreta de Fourier (DFT) en la que se eliminan los coeficientes senoidales. Ésta toma un conjunto de puntos de un dominio espacial y los transforma en una representación equivalente en el dominio de frecuencias. La DCT está bastante relacionada con la DFT, con la diferencia de que es una transformada real, debido a que los vectores base se componen exclusivamente de funciones coseno muestreadas. Además, la DCT minimiza algunos de los problemas que surgen con la aplicación de la DFT a series de datos. Una DCT produce tantos coeficientes útiles como muestras de entrada [5].

La ecuación de la Transformada Discreta del Coseno en dos dimensiones para una imagen $I(x,y)$ de tamaño $W \times H$ viene dada en la ecuación 1. Un ejemplo para nuestra base de datos se puede ver en la figura 2.

$$F(u,v) = \alpha_u(u)\alpha_v(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (1)$$

y su transformada inversa es:

$$f(x,y) = \alpha_u(u)\alpha_v(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} F(u,v) \cos\left[\frac{\pi(2x+1)u}{2M}\right] \cos\left[\frac{\pi(2y+1)v}{2N}\right] \quad (2)$$

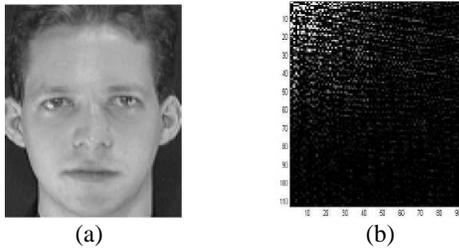


Fig. 2. a) Imagen original (112x92); b) Coeficientes DCT (112x92)

B. Transformada Wavelet

La transformada Wavelet (DWT) es una técnica de análisis de señal que descompone una señal en una familia de funciones con propiedades de localidad, tanto en el dominio del espacio (tiempo en 2D), como en el de la frecuencia (véase la figura 3). La transformada Wavelet es una herramienta de creciente popularidad en los campos de procesado de imágenes y visión computacional entre otros.

Muchas aplicaciones como compresión, detección, reconocimiento, reconstrucción de imágenes, etc. desarrollan estudios sobre esta transformada. La DWT tiene unas buenas características en cuanto a localización espacio-frecuencia y multiresolución se refiere. Las principales razones de la popularidad de las Wavelets son su completo marco teórico, su gran capacidad para elegir diferentes bases, y la baja complejidad computacional.

Las wavelets vienen siendo motivo de estudio para múltiples aplicaciones [6]. La transformada de Wavelet para señales discretas está definida como:

$$F_{m,n}(a,b) = a_0^{-m/2} \cdot \int f(t) \cdot \psi(a_0^{-m} \cdot t + nb_0) \quad (3)$$

Puede verse claramente que el comportamiento de esta función está basada en dilataciones y traslaciones a partir de una Wavelet madre. En ambos casos, esa Wavelet madre, debe satisfacer la propiedad;

$$\int \psi(t) \cdot dt = 0 \quad (4)$$

Observando la ecuación 3, se puede ver que la transformada de Wavelet depende de dos parámetros a y b, que corresponden a las informaciones de escala y tiempo respectivamente. Para la transformada Wavelet discreta, se distinguen dos enfoques: sistemas redundantes discretos (frames) y bases ortonormales de wavelets. El segundo enfoque considera la estrategia de análisis de multiresolución, la cual consiste en la descomposición de una señal (imagen) en distintas escalas. Una descomposición multiresolución permite obtener una interpretación invariante de escala de la imagen. La escala de una imagen cambia en razón de la distancia entre la escena y el centro óptico de la cámara. Cuando la escala de la imagen es modificada, la interpretación de la imagen no debería cambiar.



Fig. 3. Imagen de descomposición Wavelet en dos dimensiones.

IV. REDUCCIÓN DE CARACTERÍSTICAS

Debido al gran número de variables que se tendrá de cada imagen (unidad de millar) en nuestro sistema de reconocimiento facial, es preciso intentar reducir su número a pocas decenas de variables a fin de poder diseñar el sistema de reconocimiento con el menor número de entradas posibles. Esto nos va a permitir ganar en velocidad y en simplicidad.

A. Reducción mediante PCA

Dado que existe una alta correlación entre las variables de entrada, una posibilidad es utilizar el Análisis en

Componentes Principales (PCA) para llevar a cabo esta reducción. Se va a transformar el conjunto original de variables en otro conjunto de nuevas variables incorreladas entre sí (que no tenga repetición o redundancia en la información) llamado conjunto de componentes principales. De modo ideal, se buscan $m < n$ variables que sean combinaciones lineales de las n originales, recogiendo la mayor parte de la información o variabilidad de los datos. Para un mayor detalle se puede consultar [7], [8] y [9].

B. Reducción mediante ICA

Se puede suponer que se observan n mezclas lineales x_1, \dots, x_n de n componentes independientes. Usando notación matricial, el modelo se escribe como:

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (5)$$

donde \mathbf{x} es un vector columna de dimensión n formado por las observaciones x_1, \dots, x_n , \mathbf{s} es un vector columna de dimensión n formado por las componentes independientes s_1, \dots, s_n y \mathbf{A} es la llamada matriz de mezcla de dimensión $n \times n$. El modelo estadístico presentado en la ecuación 5 se llama análisis en componentes independientes, o modelo ICA (*Independent Component Analysis*) [10]. Lo que se observa es el vector aleatorio \mathbf{x} y debemos estimar \mathbf{A} y \mathbf{s} .

El punto de partida para ICA es la asunción que las componentes s_i son estadísticamente independientes. Entonces, después de estimar la matriz \mathbf{A} podemos calcular su inversa, llamada \mathbf{W} , y obtener las componentes independientes como sigue:

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (6)$$

Utilizando la misma idea que en PCA, se va a buscar m componentes independientes para reducir la dimensión del espacio de características. Esto nos va a dar un conjunto de componentes (variables) s_1, \dots, s_m independientes. Para un mayor detalle se puede consultar [10] y [11], donde se presentan diferentes métodos para el cálculo de ICA.

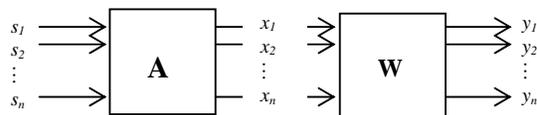


Fig. 4. Modelo de mezcla y separación. Las componentes recuperadas y_1, \dots, y_n se corresponden con las fuentes s_1, \dots, s_n .

V. SISTEMAS DE CLASIFICACIÓN

Para este trabajo se han utilizado dos tipos diferentes de sistemas de clasificación, las redes neuronales artificiales (RNA) y las máquinas de vectores soporte (SVM).

A. Redes Neuronales

La RNA puede ser definida como una estructura distribuida, de procesamiento paralelo, formada de neuronas artificiales, interconectadas por un gran número de conexiones (sinapsis).

La definición de arquitectura en nuestro caso está formada por múltiples capas, que pueden ser clasificadas en tres grupos: capa de entrada, capas intermediarias u ocultas y capa de salida (véase la figura 5).

Para este trabajo se han utilizado las RNA *FeedForward*, que consisten en capas de neuronas donde la salida de una neurona de una capa, alimenta todas las neuronas de la capa

siguiente. El aspecto fundamental de esta estructura es que no existen las uniones de retroalimentación. La red Perceptron Multicapa (MLP) es un tipo de red *Feedforward*, donde la función de umbral es una diferenciabilidad no lineal. Esta diferencia no lineal es necesaria para asegurar que el gradiente se pueda calcular. La característica crítica en un MLP es la no linealidad en las capas ocultas.

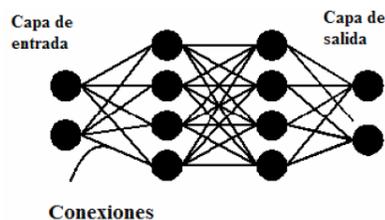


Fig. 5. Esquema de las capas de una Red Neuronal.

B. Máquinas de Vectores Soporte

Las Máquinas de Vectores Soporte están basadas en el concepto de los planos de decisión los cuales están definidos por los límites de decisión. En la figura 6, se pueden observar los objetos originales (parte izquierda de la figura) mapeados, usando un conjunto de funciones matemáticas, conocidas como kernels.

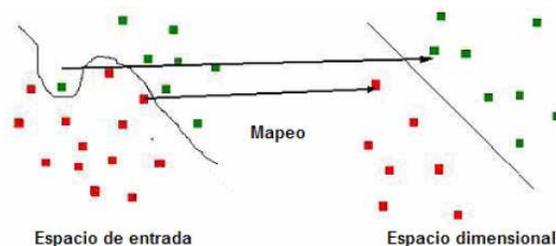


Fig. 6. Mapeo de las muestras en un espacio dimensional de mayor orden.

Las Máquinas de Vectores soporte (SVM) son primeramente un método de clasificación que cumple tareas de clasificación mediante la construcción de hiperplanos en un espacio multidimensional que separa casos de muestras de etiquetadas de diferentes clases. En particular, se va a implementar un SVM multiclase con la técnica de “uno contra todos”. Se aplica para una clasificación entre M tipos de clases, se crean M clasificadores binarios. Esto es, cada clasificador se entrena para discriminar una clase de las $M-1$ clases restantes.

VI. EXPERIMENTOS Y RESULTADOS

Los primeros experimentos realizados han sido la clasificación mediante SVM y RNA, parametrizando con la DCT y la DWT. Para ello, se empleó el 50% de la base de datos para entrenamiento y el resto para la evaluación. Estos experimentos fueron repetidos en 10 ocasiones. El tamaño para la DWT depende de las veces que se ejecuta o nivel aplicado; y la familia usada ha sido la “Biorthogonal 4.4”. El número de elementos de la DCT corresponde a la selección de los coeficientes de baja frecuencia, siendo el valor del lado de un cuadrado. Los resultados se muestran para SVM con un kernel lineal y RBF, ya que para RNA se obtuvieron menores valores de éxito. Estas tasas de éxito se presentan en media y desviación típica (dv) en las tablas 1 y 2.

Tabla 1. Resultados para DCT según el número de coeficientes

Kernel	lineal	RBF
Tamaño	Media % ± dv	Media % ± dv
10×10	89.65% ± 2.85	90.60% ± 2.27
12×11	92.90% ± 2.92	95.20% ± 2.07
15×14	96.10% ± 1.45	97.40% ± 1.08
21×19	96.90% ± 1.02	97.90% ± 1.08
34×29	98.10% ± 0.81	98.90% ± 0.42
60×50	97.05% ± 1.34	97.70% ± 1.44

Tabla 2. Resultados de la DWT según el nivel aplicado.

Kernel	lineal	RBF
# elementos	Media % ± dv	Media % ± dv
5	95.00% ± 1.52	97.10% ± 1.39
10	96.25% ± 1.45	96.20% ± 1.86
15	97.70% ± 1.21	98.00% ± 0.61
20	97.65% ± 1.37	98.20% ± 0.45
25	96.90% ± 1.87	97.10% ± 2.04

Una vez caracterizado el sistema, se pasó a la reducción de las características del mismo. Se utilizó el mejor resultado obtenido, parametrizando con la DWT y clasificando con SVM para un kernel RBF.

Si aplicamos PCA a nuestras imágenes de entrada y nos quedamos con las primeras m componentes en lugar de todas las n componentes iniciales, lo que habremos hecho es reducir la dimensión en una relación $(1-m/n) \times 100$ %. Clásicamente es lo que se ha utilizado más a menudo. El problema radica en saber si estas componentes van a ser las que más discriminen en el momento de clasificar.

El uso de ICA como etapa de preprocesado permite simplificar la estructura del clasificador y por lo tanto incrementar sus propiedades de generalización [12]. La hipótesis implícita del preprocesado con ICA es que el análisis con ICA transformará el espacio de características en un espacio donde las componentes serán independientes y alineadas a los ejes, y por lo tanto estarán mejor adaptadas a la forma en que el clasificador funciona. De la misma manera la inferencia de los pesos del clasificador será mucho más fácil ya que la búsqueda del gradiente en el espacio de características seguirá trayectorias independientes. El resultado será un clasificador menos complejo y en algunos casos con un error menor.

Así, aplicando ICA a las n observaciones para recuperar m componentes independientes vamos a conseguir el mismo efecto de reducción en una relación $(1-m/n) \times 100$ % pero con unas componentes independientes y no solamente incorreladas.

El número de variables o coeficientes usados ha sido de 986 elementos. De esta forma, vamos a proyectar el vector de entradas de dimensión 986 a un espacio de dimensión menor (entre 10 y 40 en nuestra aplicación), de manera que tendremos como nuevo vector de entrada a las componentes (ejes) de esta proyección. La reducción será, pues, entre un 99% (10 componentes) y un 96% (40 componentes).

Los resultados obtenidos al clasificar con las imágenes reducidas se han realizado con SVM, tanto para kernel lineal como para RBF (cambiando la varianza de la gaussiana - γ). Los experimentos también se han repetido en 10 ocasiones, y los resultados obtenidos se muestran en las tablas 3 y 4.

Se puede observar como los resultados para PCA con 30 elementos prácticamente se mantiene a los transformados con

DWT, de 986 coeficientes, pero con una reducción de dimensión del espacio de características del 97%.

Tabla 3. Resultados con el clasificador SVM para una reducción ICA.

Número de Componentes	Tasa de éxito (Media % ± dv)	Tipo de kernel	γ
ICA 40	94.00% ± 2.78	Lineal	---
ICA 30	93.67% ± 1.04		
ICA 20	92.83% ± 1.25		
ICA 10	76.50% ± 2.29	RBF	9×10^{-2}
ICA 40	96.33% ± 2.08		1×10^{-3}
ICA 30	97.00% ± 1.32		2×10^{-2}
ICA 20	96.33% ± 0.57		5×10^{-2}
ICA 10	89.83% ± 3.75		

Tabla 4. Resultados con el clasificador SVM para una reducción PCA

Número de Componentes	Tasa de éxito (Media % ± dv)	Tipo de kernel	γ
PCA40	96.50% ± 0.87	Lineal	---
PCA30	97.33% ± 0.58		
PCA20	93.67% ± 1.53		
PCA10	91.45% ± 3.17		
PCA40	97.33% ± 0.76	RBF	9×10^{-8}
PCA30	98.17% ± 0.76		8×10^{-8}
PCA20	96.17% ± 1.89		6×10^{-8}
PCA10	94.17% ± 3.05		1×10^{-7}

VII. CONCLUSIONES

En el presente trabajo, se ha realizado un estudio sobre la reducción de características en reconocimiento facial, después de haber parametrizado con dominios transformados, y siendo clasificados de forma más eficiente con SVM. Se puede observar en las tablas anteriores, como se mantiene la tasa de éxito del sistema posterior a su reducción.

AGRADECIMIENTOS

Jordi Solé-Casals y Vladimir Zaiats agradecen el soporte recibido del Ministerio de Educación y Ciencia (proyecto TEC2007-61535/TCM) y de la Universitat de Vic (proyecto R0912).

REFERENCES

- [1] J. D. Woodward, N.M. Orlans, P.T. Higgins, *Biometrics*, Ed. McGraw-Hill, 2003.
- [2] Q. Xiao, "Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions", *IEEE Computational Intelligence Magazine*, Vol. 2, N° 2, pp 5-25, 2007.
- [3] C. Iancu, P. Corcoran, G. Costache, "A Review of Face Recognition Techniques for In-Camera Applications; International Symposium on Signals, Circuits and Systems, Vol. 1, pp. 1-4, 2007.
- [4] Base de datos de caras ORL, (última visita: 31-03-2008) <http://www.uk.research.att.com/facedatabase.html>
- [5] Z. Pan, R. Adams, and H. Bolouri, "Image redundancy reduction for neural network classification using discrete cosine transforms", in *Proc. Int. Joint Conf. Neural Networks*, vol. 3, Como, Italy, pp. 149-154, 2000.
- [6] J.C. CoswamiI, A.K. Chan, *Fundamentals of Wavelets*, Ed. Wiley Inter-science, 1999.
- [7] L. David, *Linear Algebra and It's Applications*. New York, USA: Addison-Wesley, 2000.
- [8] Jolliffe I.T. *Principal Component Analysis*, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus.
- [9] Bishop, Christopher, *Neural Networks for Pattern Recognition*. Clarendon, UK: Oxford University Press, 1996
- [10] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis* New York, USA: John Wiley & Sons, 2001.
- [11] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, USA: John Wiley & Sons, 2002
- [12] V. Sanchez Poblador, E. Monte Moreno, J. Solé-Casals, "ICA as a preprocessing technique for Classification", ICA 2004, Granada, Spain., Lecture Notes in Computer Science, Springer-Verlag Volume 3195/2004.