

CAPÍTULO 4

RESULTADOS Y CONCLUSIONES

Los primeros resultados que obtuvimos, fueron aquellos que generó el entrenamiento de la red neuronal al utilizar como conjunto de entrenamiento, las imágenes completas de los rostros, es decir, sin utilizar una etapa de extracción de características previa al entrenamiento. Como era de esperarse, estos resultados no fueron satisfactorios debido a que como se mencionó previamente, cada una de las imágenes tiene un tamaño (resolución) de 92 x 112 píxeles, por lo cual, al representar la imagen como un vector de n elementos, obtuvimos como resultado, un vector de 10304 elementos, el cual es demasiado grande como para darlo como entrada a la red neuronal. El utilizar un vector de este tamaño, significó tener esa misma cantidad de neuronas en la capa de entrada de la red neuronal, lo cual implicó diferentes consecuencias, entre las cuales podemos mencionar: un tiempo de entrenamiento considerablemente largo, requerimientos de memoria bastante grandes, y posiblemente la no convergencia del algoritmo de entrenamiento.

Parámetros de entrenamiento

A continuación describimos en detalle los resultados que obtuvimos al realizar diferentes pruebas con el sistema propuesto. Para realizar el entrenamiento de la red neuronal, utilizamos un coeficiente de aprendizaje de 0.3, además, los valores de las conexiones entre cada uno de los neuronas (pesos), fueron inicializados con valores aleatorios pequeños, los cuales se encontraban en un rango de -0.9 a 0.9. De manera similar, los umbrales también fueron inicializados con valores aleatorios pequeños en un rango de -0.5 a 0.5. Estos valores fueron seleccionados después de observar el rango de valores en los que quedaban los pesos de la red al finalizar varios entrenamientos.

Los niveles de gris de cada una de las imágenes fueron normalizadas a valores entre 0 y 1 antes de ser utilizadas, es decir, dado que las imágenes están constituidas por un conjunto de valores entre 0 y 255, cada uno de estos valores fue dividido entre 255, obteniendo de esta manera la misma imagen pero con valores entre 0 y 1. Las salidas deseadas de la red neuronal, es decir, las salidas correspondientes a cada una de las clases (sujetos en este

caso), fueron 0's y 1's, donde únicamente uno de los nodos de salida, dependiendo del número de la clase, podía ser igual a 1 y los restantes fueron colocados a 0. La Figura Error! No text of specified style in document.-1 muestra un ejemplo de como se asoció una entrada a la red, con su correspondiente salida, en este caso, supóngase que la imagen que se está dando como entrada al sistema pertenece a la clase 1 (sujeto 1), por lo tanto, el neurón de la capa de salida que debe de ser colocado a 1, es el primero y los restantes colocados a 0.

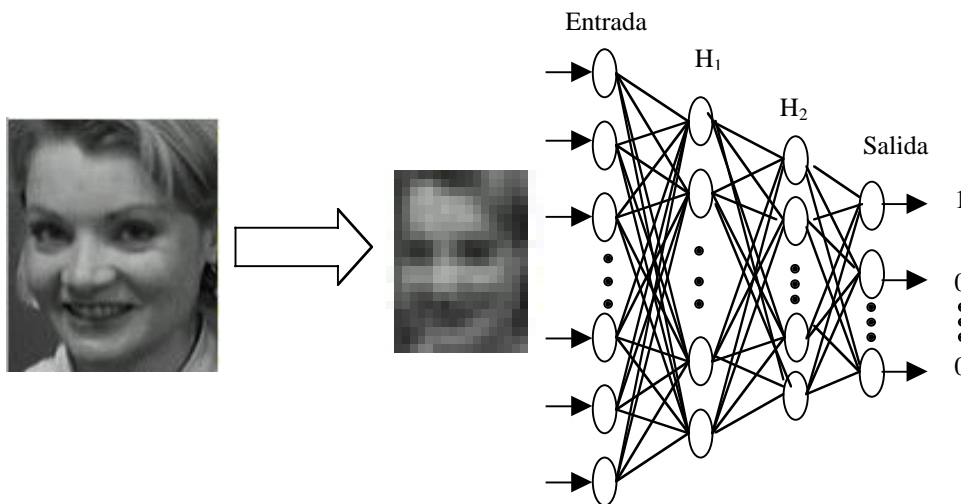
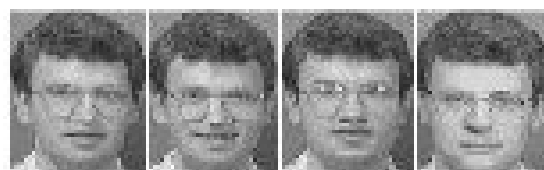


Figura Error! No text of specified style in document.-1. Ejemplo de la asociación entrada-salida de una clase dada.

Para nuestras pruebas utilizamos M nodos de salida, donde M es igual al número de clases que desean ser reconocidas; por ejemplo, M es igual a 40 cuando se utiliza la base de datos completa para entrenar al sistema.

Selección de imágenes para realizar el entrenamiento

Para realizar las pruebas del sistema utilizamos dos conjuntos de entrenamiento. El primer conjunto fue formado con cuatro imágenes por persona, las cuales fueron seleccionadas aleatoriamente de las 10 imágenes disponibles. La Figura Error! No text of specified style in document.-2 a) muestra las cuatro imágenes de una misma persona seleccionadas aleatoriamente y en la Figura Error! No text of specified style in document.-2 b) se muestran, al doble de su tamaño normal, sus 4 vectores de características correspondientes.



a)



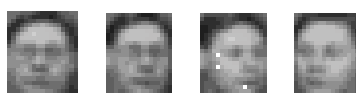
b)

Figura **Error! No text of specified style in document.-2.** a) Las 4 imágenes de entrenamiento de una misma persona seleccionadas aleatoriamente. b) Los 4 vectores de características correspondientes

Por otro lado, el segundo conjunto de entrenamiento también fue formado con cuatro imágenes por persona, pero estas imágenes fueron seleccionadas de tal manera que cumplieran con las siguientes restricciones: que hubiese 2 imágenes de frente tomadas a diferente distancia, una imagen con el rostro orientado hacia la izquierda y una imagen con el rostro orientado hacia la derecha. El objetivo que se pretende lograr al seleccionar de esta manera las imágenes de los rostros, es que exista el mayor grado de ortogonalidad entre cada una de las cuatro imágenes de una misma persona, logrando de esta manera, que el sistema pueda tener un mayor porcentaje de generalización durante la etapa de reconocimiento. En la Figura **Error! No text of specified style in document.-3** a) se muestran las cuatro imágenes de una misma persona seleccionadas de acuerdo a las restricciones anteriores. En la Figura **Error! No text of specified style in document.-3** b) se muestran, al doble de su tamaño normal, los 4 vectores de características correspondientes.



a)



b)

Figura **Error! No text of specified style in document.-3.** a) Las 4 imágenes de entrenamiento de una misma persona seleccionadas de acuerdo a las restricciones. b) Los vectores de características correspondientes.

Una vez seleccionado el conjunto de entrenamiento, realizamos la extracción de características utilizando ventanas de 10 x 10 píxeles, con un traslape de 5 píxeles entre cada una de ellas. (ver Capítulo 3 para una descripción detallada del método de extracción de características).

Casos probados

Se realizaron pruebas con diferentes conjuntos de imágenes con el propósito de observar que era lo que causaba que los porcentajes de reconocimiento y de generalización variaran de un conjunto a otro, de lo cual obtuvimos que esas variaciones eran ocasionadas por algunos rostros que tenían un tamaño un poco diferente a los otros. La Tabla **Error! No text of specified style in document.-1** muestra los resultados obtenidos con diferentes conjuntos de imágenes que fueron seleccionadas aleatoriamente.. En estas pruebas podemos observar que, en el mejor de los casos, el sistema obtuvo un porcentaje de error cercano al 2 por ciento sobre el conjunto de entrenamiento y un porcentaje de error alrededor de un 10 por ciento sobre el conjunto de prueba.

Tabla **Error! No text of specified style in document.-1.** Resultados obtenidos con diferentes conjuntos de imágenes seleccionadas aleatoriamente.

Arquitectura de la Red Neuronal	Número de Clases	Patrones por Clase en el Conj. de Entren.	Coefficiente de Aprendizaje	Error Cuadrático Medio Final	% de Reconocimiento	% de Generalización	ID de sujetos en la BD ORL	Número de Barridas
357-267-89-10	10	4	0.3	0.009	100.0	98.3	2 - 11	29
357-267-89-18	18	4	0.3	0.009	98.6	91.6	6 - 23	68
357-267-89-18	18	4	0.3	0.009	98.6	90.7	11 - 28	65
357-267-89-40	40	4	0.3	0.009	96.8	85.0	1 - 40	105

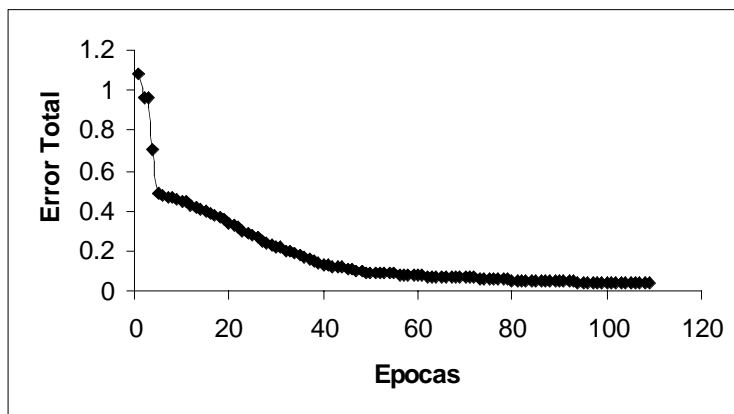
La Tabla **Error! No text of specified style in document.-2** muestra los resultados obtenidos por el sistema utilizando conjuntos de imágenes seleccionadas de acuerdo a las

restricciones que deben de cumplir las cuatro imágenes de entrenamiento de cada uno de los sujetos. En este grupo de pruebas podemos notar que, en el mejor de los casos, el porcentaje de generalización del sistema mejoró considerablemente. Nótese también que aún para 40 clases, lo cual representa la base de datos completa, se obtuvo un porcentaje de error de 3.75 % para el conjunto de prueba.

Tabla Error! No text of specified style in document.-2. Resultados obtenidos con diferentes conjuntos de imágenes seleccionadas en base a las restricciones.

Arquitectura de la Red Neuronal	Número de Clases	Patrones por Clase en el Conj. de Entren.	Coficiente de Aprendizaje	Error Cuadrático Medio Final	% de Reconocimiento	% de Generalización	ID de sujetos en la BD ORL	Número de Barridas
357-267-89-10	10	4	0.3	0.009	97.5	100.0	2 - 11	34
357-267-89-18	18	4	0.3	0.009	98.6	99.2	6 - 23	41
357-267-89-18	18	4	0.3	0.009	98.6	95.7	11 - 28	49
357-267-89-40	40	4	0.3	0.009	98.75	96.25	1 - 40	106

En la Gráfica Error! No text of specified style in document.-1 se muestra el comportamiento del entrenamiento de la red neuronal utilizando como conjunto de entrenamiento la base de datos completa con las imágenes seleccionadas en base a las restricciones. En esta gráfica podemos observar que el error total (Error Cuadrático Medio) obtenido por la red neuronal disminuye rápidamente durante la primera mitad del entrenamiento. Por otro lado, en las últimas épocas del entrenamiento, se observa que el error total obtenido por la red neuronal varía muy poco con respecto al error total obtenido en épocas anteriores, sin embargo, el porcentaje de reconocimiento continúa mejorando a pesar de estas escasas variaciones en el error total.



Gráfica **Error! No text of specified style in document.-1**. Comportamiento del entrenamiento de la red neuronal.

Para poder realizar una comparación del porcentaje de reconocimiento obtenido por el sistema, en la Tabla **Error! No text of specified style in document.-3** se muestran de manera resumida, los resultados que han sido obtenidos por otros autores, los cuales también han utilizado la base de datos de rostros ORL [4]. Para realizar sus pruebas, Zhang *e. al* [6] utilizaron un conjunto de entrenamiento formado de 2 imágenes por persona. Por otro lado, Vel y Aeberhard [8], al igual que Lawrence *et al* [11], utilizaron un conjunto de entrenamiento formado de 5 imágenes por persona.

Tabla **Error! No text of specified style in document.-3**. Resultados obtenidos por otros autores.

Autores	Método	ORL
Samaria [9]	HMM	95.0
Zhang <i>et al</i> [6]	<i>Eigenface</i>	80.0
Zhang <i>et al</i> [6]	Igualamiento elástico	80.0
Lin et al [10]	Redes Neuronales	96.0
Lawrence <i>et al</i> [11]	Redes Neuronales	96.2
Lawrence <i>et al</i> [11]	<i>Eigenface</i>	89.5
Vel y Aeberhard [8]	Segmentos de línea	100.0

Ventajas y Desventajas

El método de extracción de características de imágenes de rostros es tolerante a variaciones en cuanto a las expresiones faciales, y a rotaciones que no sean muy significativas.

Las variaciones en cuanto a la iluminación no deben de ser muy significativas puesto que los vectores de características con los que es entrenada la red neuronal están en términos de los niveles de gris de la imagen original. Por esta razón, si la iluminación de las imágenes de los rostros varía significativamente entre el conjunto de entrenamiento y el conjunto de prueba, entonces el sistema no tendrá un porcentaje de generalización aceptable.

Conclusiones

En esta tesis se ha desarrollado un sistema de reconocimiento de rostros el cual utiliza la metodología de Redes Neuronales Artificiales. Además se propuso e implementó un método para la extracción de características de imágenes de rostros. Este método tiene la ventaja de ser fácil de implementar y de no necesitar de cálculos complejos. Además, al utilizar una red neuronal artificial, el sistema es tolerante a variaciones en las expresiones faciales y a detalles faciales tales como el uso de aretes, lentes, etc. Los resultados obtenidos por el sistema indican que al combinar la Red Neuronal Artificial, con el método de extracción de características propuesto, se obtiene una eficiencia bastante aceptable.

La eficiencia del sistema fue comparada con la de otros sistemas desarrollados por otros autores, quienes también probaron sus sistemas con la base de datos ORL, y pudimos observar que la eficiencia del sistema desarrollado, está dentro de los más altos.

También realizamos una serie de pruebas con diferentes conjuntos de entrenamiento y llegamos a la conclusión de que para mejorar la eficiencia del sistema, las imágenes de una misma persona, utilizadas en el conjunto de entrenamiento, deberían de ser lo más diferente posible entre si. Por esta razón, utilizamos como conjunto de entrenamiento 4 imágenes por persona, de las cuales, 2 tenían el rostro con una orientación de frente y a diferente

distancia, por otro lado, las otras 2 imágenes tenían el rostro con una orientación hacia la izquierda y otra hacia la derecha.

Por último, como trabajo futuro para mejorar la eficiencia del sistema se sugiere el desarrollo de una etapa previa a la extracción de características, la cual normalice el tamaño de los rostros.

