

## CAPÍTULO 3

### RED NEURONAL PARA EL RECONOCIMIENTO DE ROSTROS

#### Descripción de la base de datos

Como datos de entrenamiento, en este proyecto, se utilizó la base de datos ORL [1], la cual contiene un conjunto de imágenes de rostros con 256 niveles de gris y una resolución de 92x112 píxeles. Estas imágenes fueron tomadas entre abril de 1992 y abril de 1994. En total, la base de datos contiene 10 imágenes diferentes de 40 sujetos distintos. Para algunos de los sujetos, las imágenes fueron tomadas en diferentes momentos. Todas las imágenes fueron tomadas con los sujetos en una posición frontal, con una tolerancia para alguna inclinación y rotación de hasta 20 grados. También existen variaciones en la escala de hasta 10 %. En la Figura Error! No text of specified style in document.-1 se muestra un ejemplo de las 10 imágenes de un mismo sujeto.



Figura Error! No text of specified style in document.-1. Las 10 imágenes de un mismo sujeto en la base de datos ORL.

Por comodidad, para poder realizar la tarea de reconocimiento, cada una de las imágenes es representada como un vector  $\mathbf{F}$  de dimensión  $N \times 1$ , donde  $N$  es número total de píxeles de la imagen. Este vector es obtenido al concatenar cada uno de los renglones de la matriz que representa a un rostro. Cada uno de los elementos de un vector dado, representa las

intensidades de los píxeles de la imagen en niveles de gris. Una vez que ya se cuenta con los vectores de los rostros se procede a realizar la etapa de extracción de características.

### Extracción de características

Para poder realizar la extracción de características de una imagen de un rostro, se utilizó un método que consiste en obtener un conjunto de ventanas aplicando los siguientes pasos: primero se toma una pequeña área de la imagen original a la cual llamaremos ventana. Para nuestro ejemplo utilizamos una ventana de 10 x 10 píxeles como se muestra en la Figura Error! No text of specified style in document.-2. Una vez que se obtienen los píxeles que formarán parte de la ventana, se obtiene el valor promedio de dichos píxeles. Este valor es colocado en una nueva matriz que contendrá todos los promedios de cada una de las ventanas obtenidas de la imagen original.

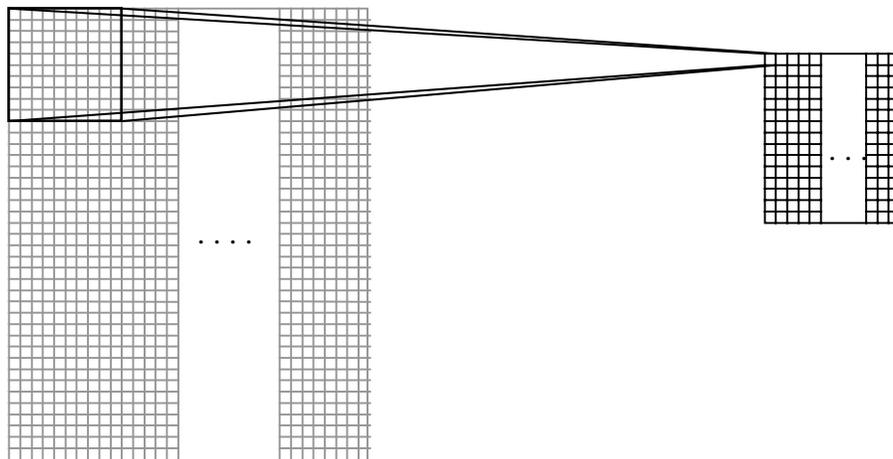


Figura Error! No text of specified style in document.-2. Primera ventana obtenida de la imagen original.

Después, se recorre la ventana un cierto número de píxeles, de tal forma que quede traslapada con la ventana anterior, como se muestra en la Figura Error! No text of specified style in document.-3.

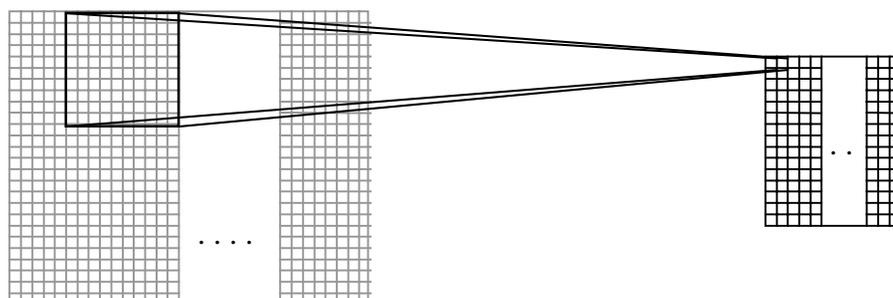


Figura **Error! No text of specified style in document.**-3. Segunda ventana obtenida de la imagen original.

A esta nueva ventana se le aplica el mismo procedimiento de la primera. El procedimiento de ir obteniendo ventanas a lo ancho de la imagen como se muestra en las figuras anteriores, se repite hasta llegar al otro extremo de la imagen.

Una vez que se ha llegado al otro extremo de la imagen, se repite el mismo proceso pero ahora con la ventana desplazada una cierta cantidad de pixeles hacia abajo (5 pixeles para nuestro ejemplo), de tal forma que esta nueva ventana quede traslapada con la ventana que se obtuvo en el proceso anterior como se muestra en la Figura **Error! No text of specified style in document.**-4.

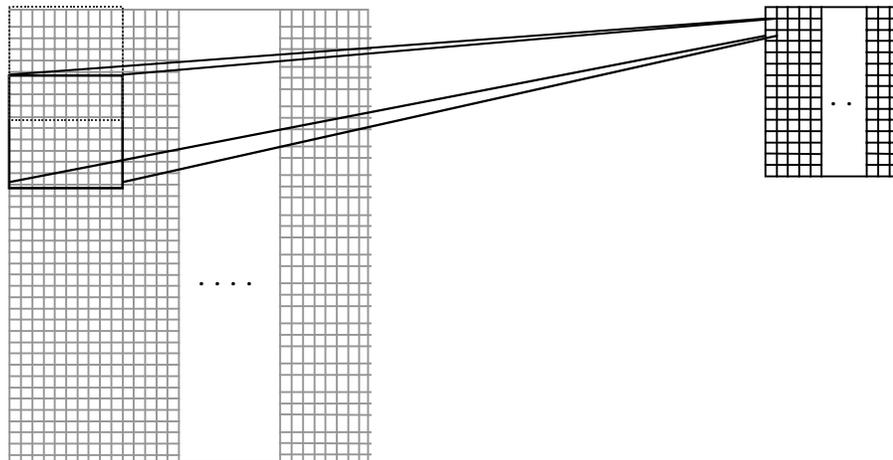


Figura **Error! No text of specified style in document.**-4. Primera ventana obtenida en la segunda etapa.

Nótese que la ventana obtenida al extremo derecho de la imagen puede llegar a ser más pequeña que las ventanas anteriores; esto ya dependerá del tamaño de la ventana y de cuantos pixeles se traslapan entre cada una de las ventanas. Lo mismo puede llegar a

sucedir con las ventanas obtenidas de la parte inferior de la imagen. Estas ventanas de diferente tamaño a las anteriores no deben de llegar a causar muchos problemas ya que de igual manera se obtiene el promedio de los pixeles que las conforman.

Este proceso se repite hasta que se llega a la esquina inferior derecha de la imagen original. La matriz que se obtiene al final del proceso, la cual también podemos considerar como un vector si concatenamos los renglones, es a lo que nosotros llamaremos el vector de características de la imagen original.

### Arquitectura de la red neuronal

Una vez que ya contamos con los vectores de características de los rostros, procedemos a realizar el reconocimiento, para lo cual, utilizamos una red neuronal alimentada hacia adelante de perceptrones multicapa (MLP). Esta red consta de tres componentes principales, una capa de entrada, dos capas ocultas  $H_1$  y  $H_2$ , y una capa de salida. En la Figura Error! No text of specified style in document.-5 se muestra la arquitectura de la red neuronal.

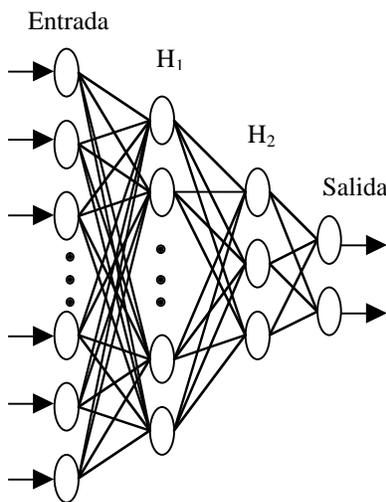


Figura Error! No text of specified style in document.-5. Arquitectura de la red neuronal.

Para todos los resultados presentados en este trabajo, el número de neuronas en la capa de entrada es igual al tamaño del vector de características, que en este caso es igual a 357 neuronas. Para la capa de salida, el número de neuronas es igual al número de clases que

desean ser reconocidas, en este caso es igual a 40 neuronas. Para las capas ocultas, el número de neuronas en cada una de ellas es calculado de la siguiente forma: el número de neuronas en la primera capa oculta ( $H_1$ ) es igual al 75% del número total de neuronas en la capa de entrada, y de forma similar, el número de neuronas en la segunda capa oculta ( $H_2$ ) es igual al 25% del número total de neuronas en la capa de entrada.

Para realizar la etapa de entrenamiento del sistema utilizamos como conjunto de entrenamiento, 4 imágenes de las 10 disponibles por cada sujeto, reservando las 6 imágenes restantes como conjunto de prueba para realizar la etapa de reconocimiento. Como algoritmo de entrenamiento se utilizó el algoritmo de retro-propagación de error (ver Apéndice B).

En la Figura Error! No text of specified style in document.-6 a) se muestran algunas de las imágenes utilizadas para realizar el entrenamiento y en la Figura Error! No text of specified style in document.-6 b) se muestran sus correspondientes vectores de características, al doble de su tamaño original, que fueron obtenidos utilizando el método de extracción de características propuesto. Para obtener estos vectores de características utilizamos ventanas de 10 x 10 con un traslape de 5 píxeles tanto horizontal como verticalmente entre cada una de las ventanas.

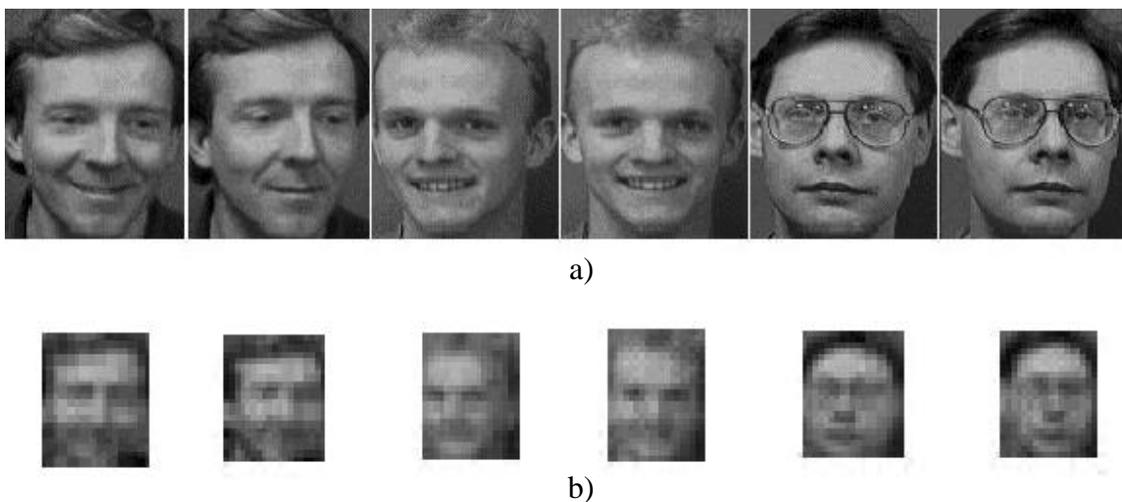


Figura Error! No text of specified style in document.-6. a) Imágenes originales, b) Vectores de características (al doble de su tamaño normal).

Nótese que para realizar el entrenamiento de la red neuronal, los vectores de características fueron normalizados a valores entre 0 y 1 (ver la sección Parámetros de Entrenamiento del Capítulo 4 para más detalles).

Para realizar la clasificación de un rostro durante la etapa de reconocimiento, es decir, decidir de que sujeto se trata, se utilizó el método de la mínima distancia entre vectores, utilizando la salida de la red neuronal como un vector de dimensión 40 para el caso de la base de datos completa y calculando la distancia de éste y los 40 vectores posibles, dando como resultado el vector con la mínima distancia.