

Reconstrucción de caracteres para mejora de sistemas OCR con RNAs

Daniela Moctezuma, Oscar S. Siordia

Centro de investigación en Geografía y Geomática Ing. Jorge L. Tamayo AC,
México D.F., México

{dmoctezuma, osanchez}@centrogeo.edu.mx

<http://www.centrogeo.org.mx>

Resumen. Numerosos avances dentro de la Inteligencia Artificial han estado inspirados en las características de los procesos biológicos de las redes neuronales. Estas características, por las ventajas que presentan, han intentado ser emuladas dentro de una computadora para imitar las tareas que son capaces de realizar los seres humanos. Siendo el OCR (Optical Character Recognition) una habilidad humana, la intención de utilizar RNAs (Redes Neuronales Artificiales) para un sistema de reconocimiento de caracteres, va en relación de cómo las RNAs tratan de simular los procesos que ocurren en el cerebro humano. El presente trabajo implementa un sistema OCR mediante RNAs, así como una modificación al mismo, implementando un paradigma para la reconstrucción del patrón de entrada (con una red Hopfield) antepuesto al paradigma utilizado para el reconocimiento (red Perceptrón). Las pruebas realizadas comparan los resultados entre el sistema OCR original y el modificado para este trabajo, utilizando tanto números como letras. Las ventajas de la modificación planteada para este trabajo pueden ser ampliadas a cualquier sistema OCR ya implementado sin importar el método de reconocimiento utilizado. Los resultados muestran una importante mejora en el porcentaje de reconocimiento con el método propuesto respecto al método comparado.

Palabras clave: RNA, OCR, red Hopfield.

1. Introducción

El reconocimiento de caracteres es una habilidad humana cuya emulación en sistemas de cómputo ha recibido mucha atención en los campos de PDI (Procesamiento Digital de Imágenes) y RP (Reconocimiento de Patrones). Como se muestra en la Figura 1, el PDI se hace necesario para el pre-procesamiento de los datos sea cual sea la técnica de reconocimiento a utilizar; sin embargo, esta última puede variar de acuerdo a los requerimientos del problema. Un sistema OCR puede ser definido como el conjunto de procesos mecánicos o electrónicos de conversión de caracteres, escritos o impresos, en un formato que sea editable en una computadora [7]. En un principio eran consideradas como

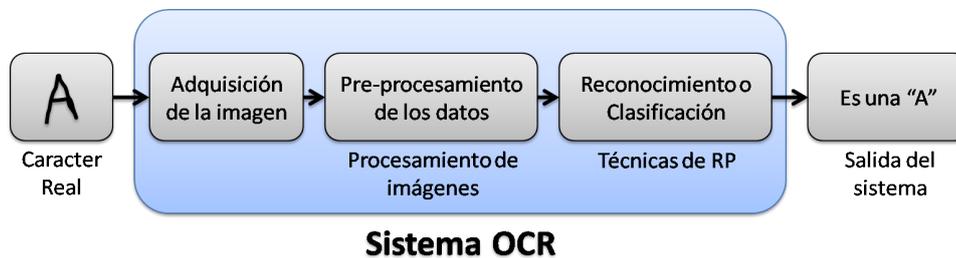


Fig. 1. Diagrama a bloques de un sistema OCR

OCRs únicamente las técnicas de reconocimiento que utilizaban lentes o espejos para realizar su tarea y, por otro lado, aquellas técnicas que utilizaban algoritmos computacionales eran conocidas como DCR (Digital Character Recognition); sin embargo, al existir muy pocas técnicas de OCR reales, esta palabra fue redefinida para incluir ambas técnicas. Los primeros sistemas OCR datan de 1929 y es un campo de investigación muy activo hasta la fecha. Los métodos clásicos de RP no han sido suficientes para el reconocimiento de caracteres debido a las siguientes razones [1]:

- Un mismo caracter difiere en tamaños, formas y estilos de una persona a otra e incluso en la misma persona en diferentes momentos.
- Al tratarse de imágenes, los caracteres pueden contener ruido que dificulte el reconocimiento.
- No hay un conjunto de reglas que definan la apariencia de los caracteres visuales y, por lo tanto, su reconocimiento.

Estos problemas pueden ser resueltos mediante la aplicación de RNAs que, gracias a sus características, permiten resolver problemas de reconocimiento con tolerancia a ruido e invariancia a escalas y rotaciones que pueden ser traducidos a formas y estilos [2]. Estas ventajas han hecho al OCR una de las aplicaciones más comunes en el área de RNAs y ha sido utilizado en los últimos años como el ejemplo práctico por excelencia en el campo académico (por ejemplo los trabajos presentados en [3],[4], [5] y [8]) .

Casi cualquier paradigma de RNAs puede ser aplicado para la solución de este problema. Sin embargo, la mayoría de ellos requieren de un proceso de aprendizaje exhaustivo para garantizar un reconocimiento acertado. En este trabajo se propone una mejora para reducir el tiempo de entrenamiento de un sistema OCR y mejorar la eficacia de los resultados. Lo anterior mediante la implementación de una red de reconstrucción antepuesta a la de clasificación de un sistema ya probado [1]. En la sección 2 de este trabajo, se exponen los detalles de la implementación del sistema OCR mencionado. En la sección 3, se presentan las modificaciones realizadas y la justificación de las mismas. Los resultados obtenidos con esta modificación se muestran en la sección 4 donde se hace también una comparación de ambos sistemas. Finalmente, en la sección 5 se presentan las conclusiones y el trabajo futuro.

2. Sistema de Reconocimiento Original

En [1], se presenta un sistema OCR de fácil implementación que se apoya en una RNA de tipo perceptrón con una sola capa y una neurona por cada caracter que se desea reconocer. En la Figura 2, se muestra un esquema del sistema para el reconocimiento de n caracteres. La entrada I es una imagen que contiene

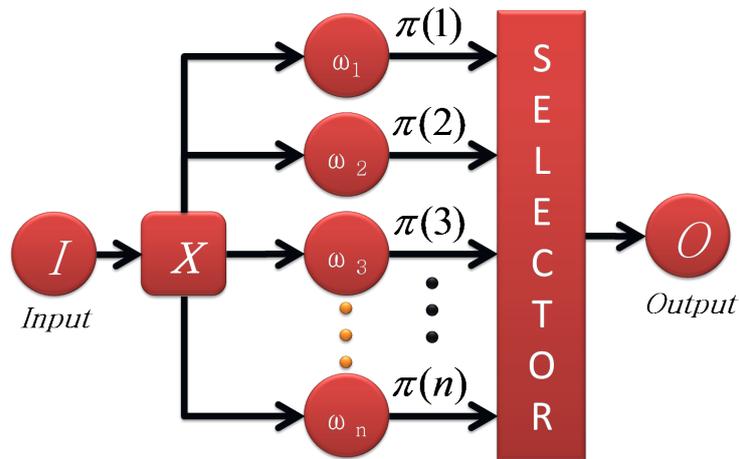


Fig. 2. Esquema del sistema OCR original

el caracter real a reconocer. Esta imagen debe ser convertida a una matriz M binaria de dimensiones $a \times b$ como se muestra en la Figura 3.

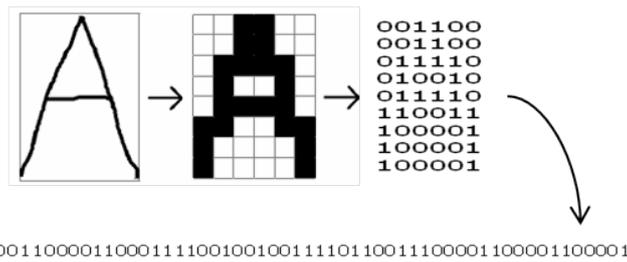


Fig. 3. Proceso de conversión de un carácter real a una matriz binaria

Por cada uno de los caracteres que el sistema sea capaz de reconocer, se requiere una neurona wn . Cada neurona contendrá una matriz de pesos Wn de las mismas dimensiones que M y su salida $\pi(n)$ será el factor de reconocimiento dada una nueva entrada. El conjunto de factores será la entrada de un selector,

que no es más que una red del tipo winner takes all, que indicará en su salida O cuál de las neuronas ha tenido un factor de reconocimiento mayor y, por lo tanto, qué carácter ha sido reconocido por el sistema OCR. Para el aprendizaje, se modificará la matriz de pesos Wn a razón de M con el siguiente método:

```
for all  $i = 1$  to  $x$  do  
  for all  $j = 1$  to  $y$  do  
    if  $M(i, j) == 0$  then then  
       $Wn(i, j) - -$   
    else  
       $Wn(i, j) ++$   
    end if  
  end for  
end for
```

Lo anterior, se realizará siempre que exista una nueva entrada I y se desee entrenar una neurona wn para su reconocimiento. El factor de reconocimiento $\pi(n)$ de cada una de las neuronas, es el cociente de la puntuación del candidato $\psi(n)$ y la puntuación del ideal $\mu(n)$ como se muestra en la ecuación 1.

$$\pi(n) = \frac{\psi(n)}{\mu(n)} \quad (1)$$

La puntuación del candidato, es una medida de qué tanto se parece la entrada M actual a la matriz de pesos Wn de la neurona n y se calcula de la siguiente manera:

```
for all  $i = 1$  to  $x$  do  
  for all  $j = 1$  to  $y$  do  
     $\psi(n) = \psi(n) + Wn(i, j) * M(i, j)$   
  end for  
end for
```

La puntuación del ideal, es una medida de la calificación que tendría el patrón de entrada M si fuera perfecto, es decir si solo tuviera 1's en las posiciones donde la matriz de pesos Wn tiene valores positivos. Este factor para la neurona n se calcula de la siguiente manera:

```
for all  $i = 1$  to  $x$  do  
  for all  $j = 1$  to  $y$  do  
    if  $Wn(i, j) > 0$  then  
       $\mu(n) = \mu(n) + Wn(i, j)$   
    end if  
  end for  
end for
```

El factor de reconocimiento tendrá siempre un rango de $(-\infty, 1]$. Para este caso, el sistema fue implementado para trabajar con matrices de 32×32 ya que permite un reconocimiento más preciso y abre la posibilidad de realizar una mejor comparación entre el sistema original y el que se propone.

3. Desarrollo

En la Figura 4, se muestra el esquema del sistema OCR con la modificación propuesta.

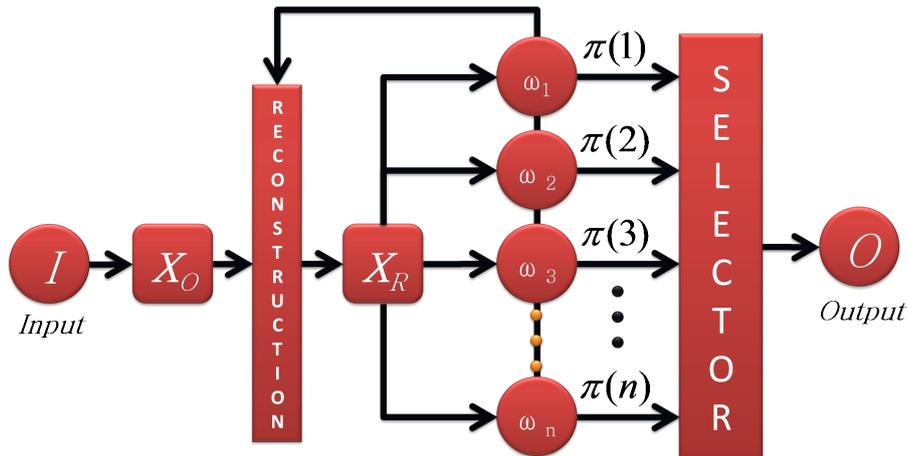


Fig. 4. Esquema del sistema OCR con la modificación propuesta

Esta modificación añade al OCR un módulo de reconstrucción que basa su operación en la información contenida en las neuronas del sistema original por lo que puede ser aplicado en cualquier sistema ya existente como una mejora al mismo. El módulo de reconstrucción se trata de una RNA de tipo *Hopfield* [6] la cual tiene como entrada la matriz binaria M_O y como salida la matriz reconstruida M_R del mismo tipo. El resto del proceso de reconocimiento será idéntico al descrito en la sección 2.

3.1. Aprendizaje del módulo de reconstrucción

Continuando con la terminología utilizada para describir el sistema OCR original, a continuación se describirá el proceso de aprendizaje del módulo de la red *Hopfield*. Para esta etapa, es necesario calcular la matriz de pesos de la nueva red basándose en la información aprendida por la red perceptrón. Para este caso es necesaria una matriz de 1024×1024 que será calculada por medio de la Ecuación 2.

$$W(i, j) = \sum_{s=1}^n W_s(\text{mod}(32, i), \text{div}(32, i) + 1 * W_s(\text{mod}(32, j), \text{div}(32, j) + 1) \quad (2)$$

La matriz de pesos de la red Hopfield se trata de una matriz simétrica cuadrada. Esta matriz deberá ser actualizada cada vez que el sistema OCR original

agregue una nueva neurona a su red perceptrón para reconocer un caracter adicional. Una vez que la matriz de pesos ha sido calculada, es necesaria la etapa de reconocimiento para realizar la reconstrucción de la matriz de entrada, es decir, obtener M_R a partir de M_O . Las funciones *div* y *mod* fueron aplicadas debido a que es necesaria la conversión de una matriz bidimensional a un vector unidimensional.

3.2. Reconocimiento del módulo de reconstrucción

El reconocimiento de una red *Hopfield*, trata de un proceso iterativo que permite la reconstrucción de una matriz de entrada en base a los patrones aprendidos durante el proceso de aprendizaje. Para lograr esta reconstrucción se utiliza la Ecuación 3.

$$M_0(t+1) = fh \left\{ \sum_{j=1}^n w(i,j) M_o(t) \right\} \quad (3)$$

Este proceso se repetirá indefinidamente hasta que $M_O(t) = M_O(t+1)$, en ese momento se tendrá una reconstrucción aceptable de acuerdo a los datos aprendidos por la red. La función *fh* es una función de umbralización definida en la Ecuación 4.

$$fh(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases} \quad (4)$$

La función de umbralización se utiliza para optimizar la calidad de la imagen para su posterior procesamiento, esta función puede ser modificada de acuerdo al problema que se esté resolviendo (calidad de adquisición de la imagen), sin embargo en base a trabajos previos se ha demostrado que estos valores son ideales para trabajar con sistemas OCR.

4. Experimentos y resultados

Las pruebas realizadas para este trabajo incluyeron caracteres tanto de dígitos como de letras. En la Figura 5 se muestra la colección de imágenes utilizadas para el entrenamiento del sistema OCR. En cada prueba se entrenó al sistema OCR una sola vez por cada uno de los patrones de números o letras. El sistema permite realizar un reconocimiento con o sin haber realizado una reconstrucción para fines comparativos. En las Figuras 6, 7, 8, 9, 10, 11, 12 y 13 se muestran diversas pruebas realizadas en el sistema indicando, en cada caso, si se trata de un reconocimiento por medio del sistema OCR original y por el propuesto. Para la primera prueba, fue dibujado a mano el dígito 1. En este caso, el sistema realizó un reconocimiento parcial del número con un 36% de certeza como se muestra en la Figura 6. En la Figura 7, se presenta el reconocimiento utilizando el módulo de reconstrucción. Para este caso, la



Fig. 5. Base de datos de imágenes utilizadas en el entrenamiento

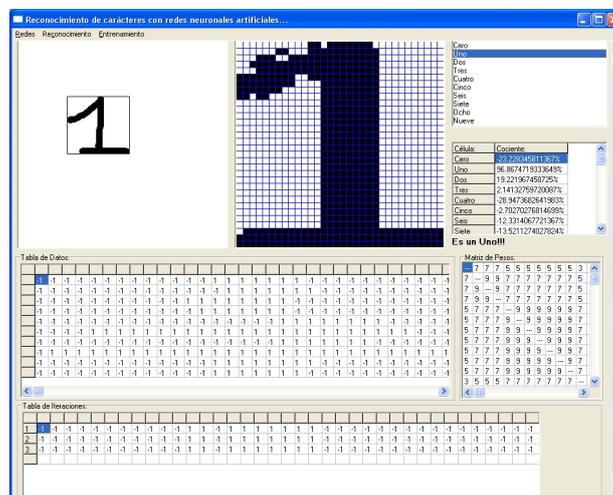


Fig. 6. Prueba con el dígito 1 en el sistema OCR original.

certeza del reconocimiento del número aumentó hasta un 97%. Para la segunda prueba con dígitos numéricos, se intentó realizar un dibujo del número 4 que se diferenciara en mayor proporción al dígito representado en la base de datos de imágenes. En la Figura 8 se muestra el reconocimiento realizado para esta prueba en el sistema OCR original, para este caso se obtuvo un porcentaje de reconocimiento del 19%. En la Figura 9, se presenta la reconstrucción realizada por medio de la red *Hopfield* y el reconocimiento del dígito con una certeza de 100%. Fueron realizadas otras pruebas con dígitos numéricos utilizando bases de datos de imágenes o de números escritos a mano, sin embargo, aunque los dígitos de entrada del sistema son siempre escritos a mano, los resultados de reconocimiento fueron muy superiores cuando la base de datos de entrenamiento se trata de caracteres con una tipografía estándar. También se utilizaron letras para los experimentos y, al igual que los resultados anteriores, los resultados del reconocimiento fueron mejores al utilizar el módulo de reconstrucción propuesto.

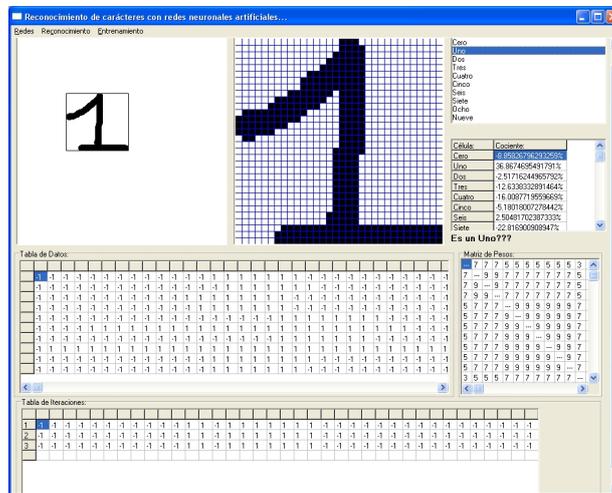


Fig. 7. Prueba con el dígito 1 en el sistema OCR modificado

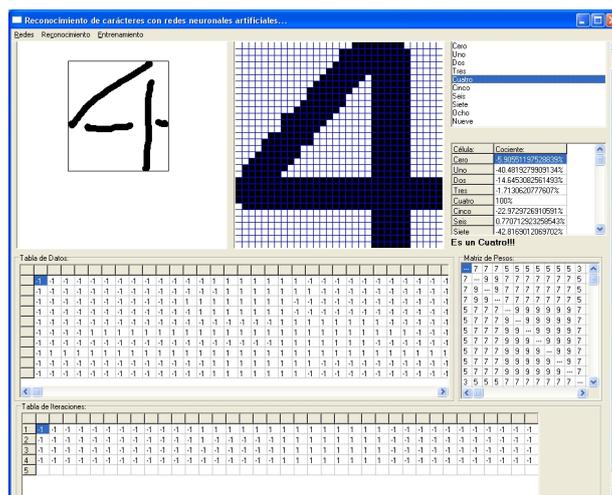


Fig. 8. Prueba con el dígito 4 en el sistema OCR original

En la figura 10 se presenta el reconocimiento de una letra A dibujada de forma asimétrica para dificultar el reconocimiento y reconstrucción del caracter. En el sistema OCR original se logra un reconocimiento con una certeza de 27.8%. En la figura 11, se muestra la reconstrucción realizada para la letra asimétrica y su reconocimiento con el sistema OCR modificado con una certeza del 100%. La última prueba realizada fue con la letra E. Para esta prueba se intentó dibujar una letra E cuyas características principales variaran respecto a la letra utilizada en el entrenamiento. Por ello, se dibujó una E redondeada en todas sus líneas y se intentó obtener un reconocimiento por medio del sistema OCR original. En

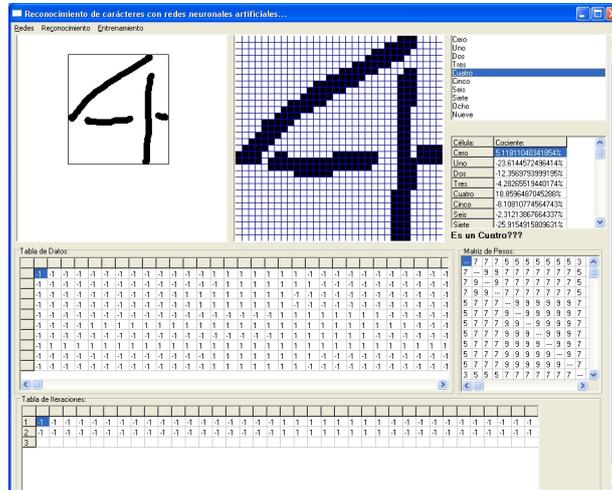


Fig. 9. Prueba con el dígito 4 en el sistema OCR modificado

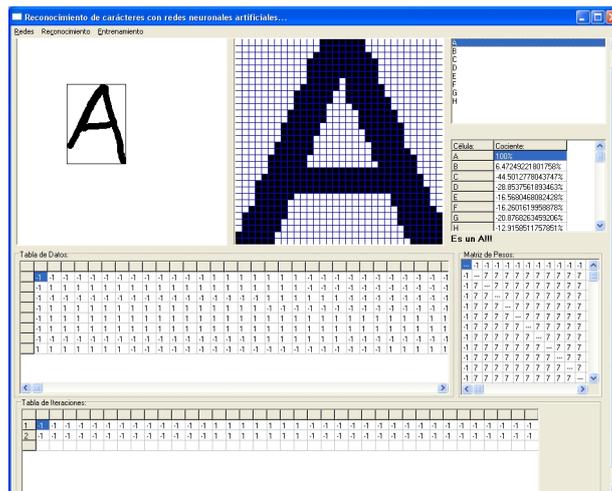


Fig. 10. Prueba con la letra A en el sistema OCR original

la Figura 12, se muestra el reconocimiento parcial de la letra E con una certeza del 31.16%. Finalmente, en la Figura 13 se muestra el reconocimiento logrado después de la reconstrucción (con el método propuesto) donde la certeza fue mejorada hasta un 70.9%

Considerando todos los resultados se puede resumir que el porcentaje de acierto del método propuesto fue de un 97.25% de reconocimiento contra un 30% de reconocimiento del sistema comparado.

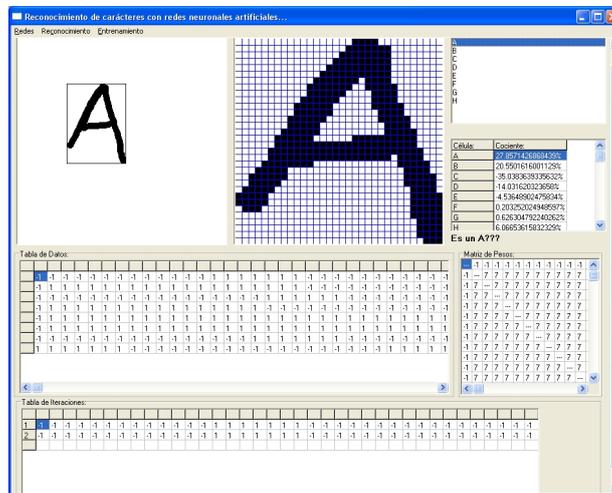


Fig. 11. Prueba con la letra A en el sistema OCR modificado

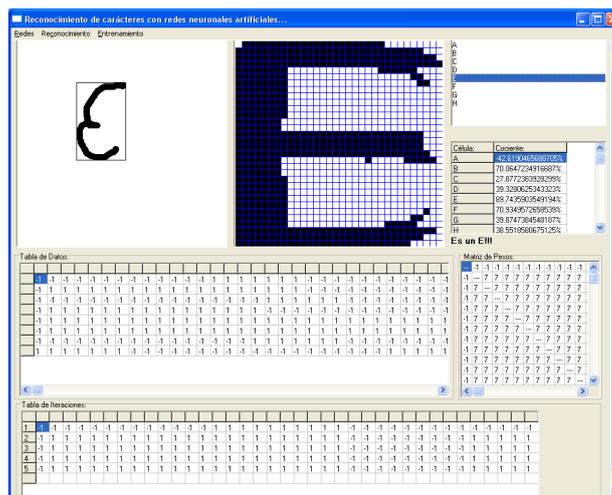


Fig. 12. Prueba con la letra E en el sistema OCR original

5. Conclusiones y trabajos futuros

La modificación propuesta para el sistema OCR es la unión de dos paradigmas que normalmente trabajan separados. Al agregar una red *Hopfield* para la reconstrucción de los datos de entrada, en este caso el caracter a reconocer, aumentó el porcentaje de la eficacia del sistema en la etapa de reconocimiento. El sistema probado fue entrenado únicamente con un patrón de cada una de las letras o números mientras la mayoría de los sistemas OCR requieren de un entrenamiento exhaustivo para lograr resultados aceptables. Los resultados

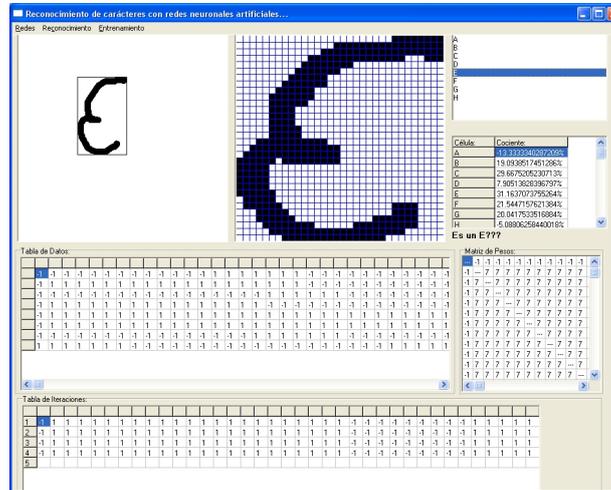


Fig. 13. Prueba con la letra E en el sistema OCR modificado

arrojados por el sistema OCR modificado fueron mejores en la mayoría de las pruebas realizadas; sin embargo, como trabajo futuro se propone ampliar el rango de pruebas para tener resultados más contundentes. Además, se propone realizar una tabla comparativa con el porcentaje de parecido de cada uno de los dígitos y letras que componen al idioma español para poder conocer la probabilidad de confusión entre cada uno de ellos y poder realizar un sistema de reconocimiento que sea capaz de clasificar todas las letras y dígitos en una misma prueba haciendo las modificaciones necesarias al paradigma de acuerdo a los resultados de la tabla mencionada, esto mediante la incorporación de técnicas de aprendizaje incremental.

Referencias

1. Shashank Araokar: Visual Character Recognition using Artificial Neural Networks. MGMs College of Engineering and Technology (2005)
2. Jain, A.K. and Jianchang Mao and Mohiuddin, K.M.: Artificial neural networks: a tutorial. *J. Mol. Biol.*, 147, 31–44 (1996)
3. Alexandria, A.R.; Cortez, P.C.; Felix, J.H.S.; Girao, A.M.; Frota, J.B.B.; Bessa, J.A.: An OCR System for Numerals Applied to Energy Meters. *Latin America Transactions, IEEE (Revista IEEE America Latina)*, 957–964 (2014)
4. Xiaojun Zhai, Bensaali, F., Sotudeh, R.: OCR-based neural network for ANPR. In: *Imaging Systems and Techniques (IST)*, 2012 IEEE International Conference on, pp. 393–397 (2012)
5. Choudhary, A., Ahlawat, S., Rishi, R.: A Neural Approach to Cursive Handwritten Character Recognition Using Features Extracted from Binarization Technique. *Complex System Modelling and Control Through Intelligent Soft Computations*, 319, 745–771 (2015)

Daniela Moctezuma, Oscar S. Siordia

6. Hopfield, J.: Neural networks and physical systems with emergent collective computational abilities. *PNAS*, 79, 2554–2558 (1982)
7. Carolyn Strange, Daniel McNamara, Josh Wodak, Ian Wood: Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers. *Digital Humanities Quarterly*, 8 (2014)
8. Shoichi Taniguchi: Duplicate bibliographic record detection with an OCR-converted source of information. *J. Information Science*, 39, 153–168 (2013)