Reconocimiento de caracteres del alfabeto dactilológico mediante redes neuronales artificiales: Un enfoque experimental

Diego Auquilla[1], Kenneth Palacio-Baus[1], Víctor Saquicela[2]

[1] Departamento de Ingeniería Eléctrica, Electrónica y Telecomunicaciones, Universidad de Cuenca, Av. 12 de Abril y Agustín Cueva, Cuenca, Ecuador, 010201.

[2] Departamento de Ciencias de la Computación, Universidad de Cuenca, Av. 12 de Abril y Agustín Cueva, Cuenca, Ecuador, 010201.

Autores para correspondencia: diego.auquillaa@ucuenca.ec, kenneth.palacio@ucuenca.edu.ec, victor.saquicela@ucuenca.edu.ec

Fecha de recepción: 28 de septiembre 2015 - Fecha de aceptación: 12 de octubre 2015

RESUMEN

En este artículo se presenta el desarrollo de un sistema orientado a facilitar la comunicación de aquellas personas con discapacidad auditiva, del habla o ambas; y que se ven obligados a utilizar otras formas de comunicación como el uso del lenguaje de señas. Para solventar el problema de comprensión de este lenguaje, se propone un enfoque experimental de reconocimiento de caracteres del alfabeto dactilológico mediante la adquisición y procesamiento de imágenes digitales, y la aplicación de un clasificador basado Redes Neuronales Artificiales (RNA).

Palabras clave: Dactilología, lenguaje de señas, Redes Neuronales Artificiales, clasificador.

ABSTRACT

This article presents the development of a system aimed to aid people with speech and/or communication disabilities, who must use the sign language and the dactilologic alphabet in order to transmit their ideas. To assist others in understanding such language, we present an experimental approach for sign language characters recognition through digital image acquisition and processing techniques and the use of a Neural Network based classifier.

Keywords***:*** Dactylology, sign language, Artificial Neural Networks, classifier.

1. INTRODUCCIÓN

Hoy en día, los avances tecnológicos están presentes en muchas actividades de la vida cotidiana y con ello, los esfuerzos de integrar a personas con diversos tipos de discapacidades, como aquellos con dificultades del habla y de escucha, para quienes el lenguaje de señas se ha convertido en la más versátil forma de comunicarse. Este lenguaje se basa en movimientos corporales y caracteres alfanuméricos representados mediante diferentes posiciones de las manos, y permite escribir palabras que pueden ser leídas rápidamente. Sin embargo, esta forma de comunicación es escasamente conocida por el resto de las personas, por lo que es pertinente presentar propuestas enfocadas a la comprensión e interpretación de mensajes transmitidos mediante lenguaje de señas utilizando técnicas de procesamiento de imágenes y reconocimiento de patrones. En este artículo se presenta un sistema orientado a facilitar la comunicación de aquellas personas con discapacidad auditiva, vocal o ambas; y que utilizan otras formas de comunicación como el uso del alfabeto dactilológico, un sistema de representación simbólica que asocia una señal manual con una letra. Se propone un enfoque experimental en el que las señas correspondientes a los caracteres generados por una persona son adquiridos mediante una cámara digital; posteriormente las imágenes se procesan y segmentan

mediante técnicas de tratamiento digital de imágenes para alimentar un clasificador basado en Redes Neuronales Artificiales (RNA) que permite identificar el caracter mostrado. El modelo neuronal corresponde a una red *feedforward* multicapa desarrollada en Matlab[1] y entrenada mediante el algoritmo de retropropagación (*Backpropagation*), (Zurada, 1992; Hecht-Nielsen, 1989). Como una extensión al reconocimiento caracteres, se presenta un prototipo para la escritura de palabras, en la que los caracteres reconocidos mediante la captura de video son transformados a texto escrito.

El resto de este artículo se organiza de la siguiente manera: La Sección 2 presenta algunos trabajos relacionados y sus resultados. La Sección 3 incluye los detalles de implementación del sistema, los procedimientos de adquisición y procesamiento de las imágenes, y una visión general del sistema. La Sección 4 presenta el procedimiento de evaluación del sistema, la funcionalidad del interfaz de usuario y los resultados de reconocimiento. Finalmente, la Sección 5 presenta las conclusiones derivadas de este trabajo así como las líneas futuras de estudio y desarrollo que han surgido a partir de este prototipo.

2. TRABAJOS RELACIONADOS

Existen diversos estudios en el campo social que pretenden integrar a la sociedad a personas con alguna discapacidad física. Esta sección introduce algunos aportes relevantes enfocados a este problema, haciendo uso de varias técnicas computacionales incluyendo RNA.

En el trabajo documentado en (Priego-Perez, 2012) se presenta un sistema para reconocer la información contenida en imágenes del lenguaje de señas. Su autor establece dos etapas: en la etapa de reconocimiento utiliza la cámara de un dispositivo *Kinect*[2], mediante la cual se adquiere nuevos patrones que posteriormente se comparan con los ya almacenados en la base de conocimiento del sistema durante la etapa de aprendizaje. En general, el enfoque se basa en una estimación de similitud entre la imagen adquirida y la del patrón almacenado, considerando únicamente valores de similitud que superen el 90%.

Barkoky & Charkari (2011) proponen en un método para reconocer los números del lenguaje de signos persa utilizando imágenes segmentadas y el método de *Thinning* (Wah Ng & Ranganath, 2002), que convierte imágenes en líneas. Tras la captura, la imagen de la mano es segmentada para lo cual se le asigna un color blanco mientras el fondo es negro. Luego se aísla la porción de brazo visible de la mano para simplificar el problema de interferencia del uso de mangas en la ropa del sujeto. Para separar la mano del brazo se aprovecha el hecho que el ancho de la mano aumenta conforme el brazo termina, además del cambio brusco en su contorno: Figura 1(a).

Para el reconocimiento de la seña, el algoritmo de *Thinning* hace que los objetos se grafiquen como líneas y adicionalmente se crean puntos en las terminaciones y uniones de los dedos, Figura 1(b). Luego se calcula y almacena la longitud de cada segmento de las líneas obtenidas, de tal modo que el segmento que tenga mayor longitud es etiquetado, con una letra. El resto de segmentos se comparan con el segmento más largo etiquetado para poder descartar segmentos demasiado pequeños e introducidos generalmente por ruido. Para reconocer la señal se cuenta el número de puntos que se crean en los dedos (terminación de los segmentos), logrando un 96% de precisión, incluyendo imágenes con rotación y escalamiento.

En (Incertis *et al*., 2006) se propone una interfaz usando visión por computadora en el que la persona que realiza las señas utiliza un guante de color azul, combinando una serie de normas para evaluar la imagen de la señal reconocida con respecto a los modelos de signos almacenados en una base de datos. Para el reconocimiento se utiliza la detección y extracción del contorno del guante mediante un modelo de color HSV[3] (modelo matemático que permite representar los colores mediante

[1] Matlab: www.mathworks.com
[2] Microsoft Kinect: https://dev.windows.com/en-us/kinect
[3] HSV: hue (tono), saturation (saturación), and value (valor)

números). El sistema calcula cuatro distancias y las compara con patrones ya almacenados, logrando un reconocimiento satisfactorio de 19 señales.

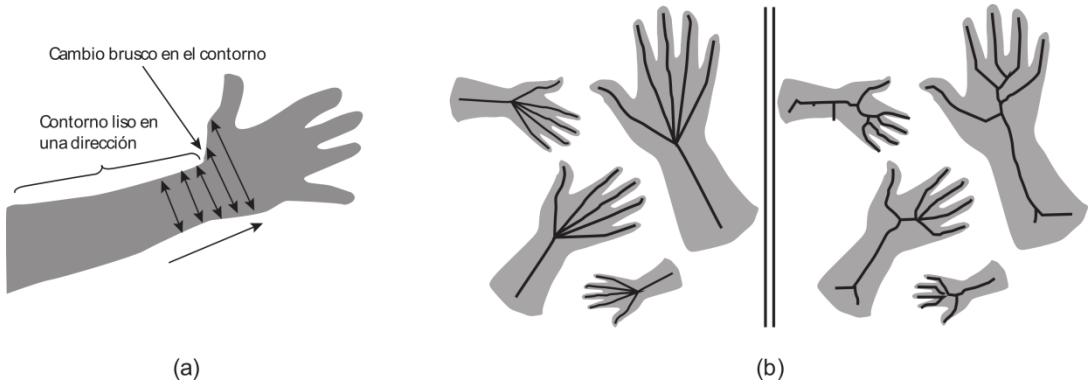


Figura 1. (a) Detección de anchura de la mano para separación del brazo (Barkoky & Charkari, 2011); (b) Representación del algoritmo *Thinning* (Wah Ng & Ranganath, 2002).

El trabajo de (Razo-Gil *et al*., 2009) considera únicamente aquellas letras que se representan en el alfabeto dactilológico sin la necesidad de ejercer movimiento, mediante un método de valor umbral para el procesamiento y segmentación de las imágenes que logra separar unos objetos de otros con el objetivo de identificar las regiones de interés de acuerdo a la postura en la que se encuentre la mano del sujeto. El objetivo de este enfoque es conseguir extraer características relevantes que se puedan medir. Sus autores concluyen que el utilizar los primeros cuatro variantes de Hu sobre imágenes binarizadas mediante el método de Otsu (1979), no es suficiente para conseguir una clasificación satisfactoria.

Heng Du & TszHang (2012) procesamiento de imágenes adquiridas mediante el dispositivo Kinect, dada su capacidad de estimación de profundidad. Primero se segmenta el área de interés separando el fondo de la zona de la mano, seguido de una transformación de la imagen en escala de grises. Después, una etapa de filtrado elimina el ruido de la imagen, para poder extraer el contorno de la mano el cual se aproxima poligonalmente con el fin de encontrar los defectos de convexidad. Estos defectos se filtran para encontrar las puntas de los dedos, similar a Priego-Perez (2012). Para reconocer un caracter determinado, se cuenta el número de dedos identificados, lo que limita el uso del sistema únicamente un conjunto de cinco señales a reconocer.

3. DESCRIPCION DEL SISTEMA

El sistema de reconocimiento de caracteres en el lenguaje de señas mediante RNA, se compone de cuatro etapas según se muestra en la Figura 2.

3.1. Adquisición de la imagen

Para adquirir la imagen se utilizó la cámara de un teléfono celular accesible desde Matlab mediante una dirección IP y para lo cual se requiere de ciertas condiciones favorables de iluminación. El prototipo utiliza una maqueta de adquisición, mostrada en la Figura 3(a), que cuenta con fondo de color negro que mejora las condiciones de contraste y reflejos de luz natural. Se sugiere que el sujeto de prueba utilice una manga de color oscuro que permita únicamente la exposición de la mano. La maqueta de adquisición consta de un soporte para la cámara, cuya altura es ajustable para que las imágenes capturadas contengan a la mano en su totalidad.

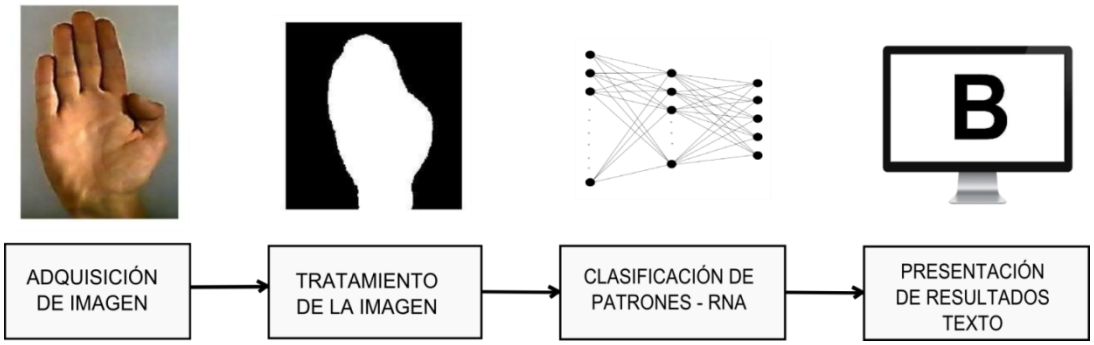


Figura 2. Diagrama de bloques del sistema.

3.2. Tratamiento de la imagen

Una vez que la imagen ha sido capturada se procede a su tratamiento, que consiste en acondicionarla de tal modo que se pueda extraer vectores de características orientados a alimentar un algoritmo de reconocimiento basado en RNA. En la Figura 3(c) se observa las etapas de procesamiento de la imagen, las mismas que han sido implementadas en Matlab. Puesto que se utiliza una cámara IP, la captura de imágenes puede realizarse de manera remota desde cualquier computador dentro de la red del sistema.

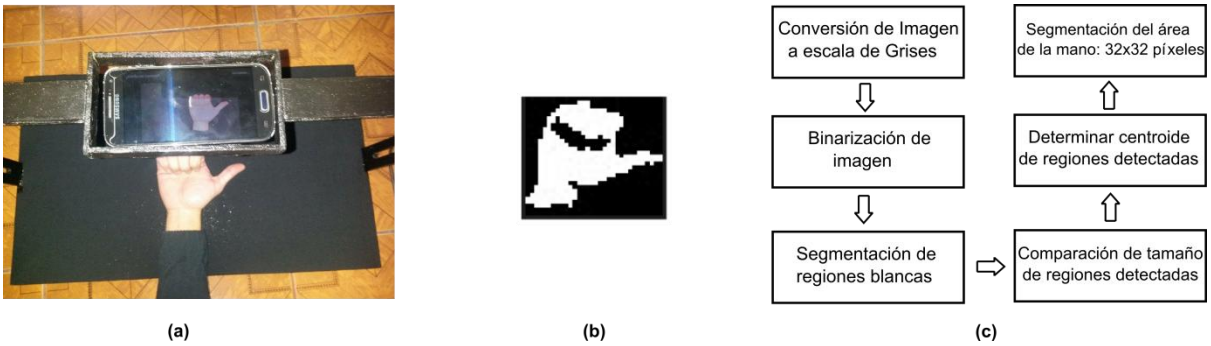


Figura 3. (a) Maqueta de adquisición de señas; (b) Imagen binarizada resultante; (c) Diagrama de bloques para el tratamiento de la imagen.

La imagen adquirida se convierte en escala de grises como paso previo a su binarización, que se facilita dadas las características de la maqueta de adquisición (Sección 3.1). Se establece como una intensidad máxima (color blanco) a todas aquellas áreas de la imagen cuya intensidad se encuentre por encima de un determinado umbral (Otsu, 1979), y como negro al resto. Así, la mano del sujeto de prueba es interpretada como una región de imagen conexa de color blanco; sin embargo, existe la posibilidad de encontrar pequeñas regiones blancas que no corresponden con la mano, introducidas por reflejos de luz indeseados u objetos ajenos al sistema como partículas de polvo. Estas regiones se comparan en función de su tamaño, asumiendo que la más grande es aquella que representa a la mano, lo que permite segmentarla como un espacio limitado que corresponde al patrón a analizar que corresponde a una región cuadrada *32x32* píxeles. La Figura 3(b) ilustra el resultado de este proceso para la letra *A*. Cada letra o seña puede entonces representarse como una matriz de elementos binarios, que representa un espacio multimensional sobre el cual se ha mapeado cada una de las letras a reconocer en el sistema.

3.3. Clasificación de patrones: Red Neuronal Artificial

La implementación del sistema de clasificación, parte del establecimiento vectores asociados a cada patrón de ingreso a la RNA para el aprendizaje y las pruebas. Se utiliza una RNA multicapa entrenada con el algoritmo de retropropagación de error o *backpropagation* (Zurada, 1992), que clasifica los

patrones adquiridos mediante un esquema de aprendizaje supervisado. La arquitectura de la red y sus detalles de implementación se presentan en la Sección 4.2.

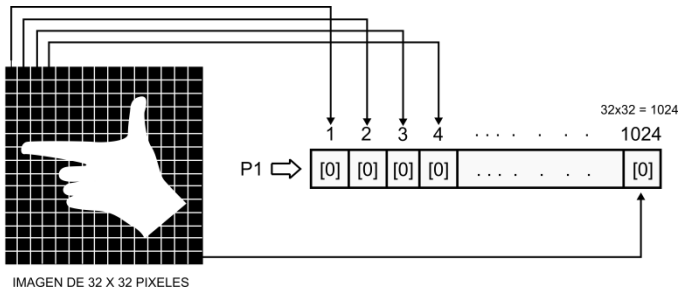


Figura 4. Proceso de Conformación de Patrones para la Letra G.

Como se ilustra en la Figura 4, cada imagen binarizada corresponde a una imagen de tamaño *32x32*, donde cada pixel negro corresponde a un 0 binario y uno blanco a un 1. Así, para conformar los patrones de entrenamiento y prueba, esta matriz se convierte en un vector de 1024 características que representa una letra que se ingresa a la RNA y al mismo tiempo define la dimensionalidad de la capa de entrada.

3.4. Presentación de resultados

Este componente posibilita la presentación en pantalla de los caracteres reconocidos como una cadena a través de concatenar las letras reconocidas del lenguaje de señas para formar palabras y facilitar la comunicación de una persona con discapacidad del habla.

4. EXPERIMENTACIÓN Y EVALUACIÓN DEL SISTEMA

Esta sección presenta la etapa experimental de este trabajo, en la que se detalla el entrenamiento del modelo neuronal a partir de una serie de datos.

4.1. Selección de datos

Para la evaluación del sistema se considera 12 letras que se muestran en la Figura 5(a) y que forman el conjunto de experimentación, en particular, letras que se representan sin movimiento explícito de la mano. Estas letras exhiben claras diferencias visuales entre ellas y por ello menor complejidad al momento de su clasificación. Se obviaron letras que guardan extrema similitud entre sí y letras que requieren de movimiento, como por ejemplo la *J*. La Figura 5(b) presenta un ejemplo de las señales excluidas en este trabajo.

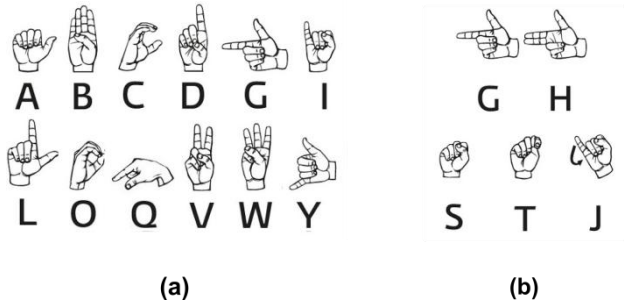


Figura 5. (a) Letras consideradas para el trabajo; (b) Letras no consideradas.

4.2. Conjunto de entrenamiento

El conjunto de entrenamiento para la red neuronal se forma a partir de 10 muestras de cada una de las señas especificadas anteriormente, adquiridas manualmente mediante diferentes fotografías de la mano del sujeto de prueba representando la letra en la maqueta de adquisición. La Figura 6(a) muestra la representación de la letra *G* en el espacio de características establecido e ilustra la conformación de sus 10 vectores destinados a entrenamiento. El set de entrenamiento se conformó a partir un total de 120 imágenes, de las cuales 108 patrones se utilizaron para el entrenamiento y 12 para formar el conjunto de validación.

La arquitectura de la RNA implementada se muestra en la Figura 6(b). Se utiliza un modelo con dos capas ocultas que se entrena mediante el algoritmo de optimización de *BackPropagation*. La primera capa oculta tiene 64 neuronas mientras que la segunda tiene 24 neuronas. Estos valores resultaron de un proceso experimental del cual se obtuvo los mejores resultados. Para establecer el mejor modelo se realizaron 10 diferentes entrenamientos, obteniendo diferentes modelos que se diferencian por el número de neuronas en la primera capa oculta. Para cada entrenamiento se obtuvo un vector de salida de 12 elementos, sobre el cual se determina el resultado de clasificación mediante una salida activa (valor de salida cercano a 1) para la letra ganadora e inactiva para las restantes (valor de salida cercano a 0). La Tabla 1 muestra los valores de salida de cada letra para los 10 modelos, así como el valor promedio de todas las salidas.

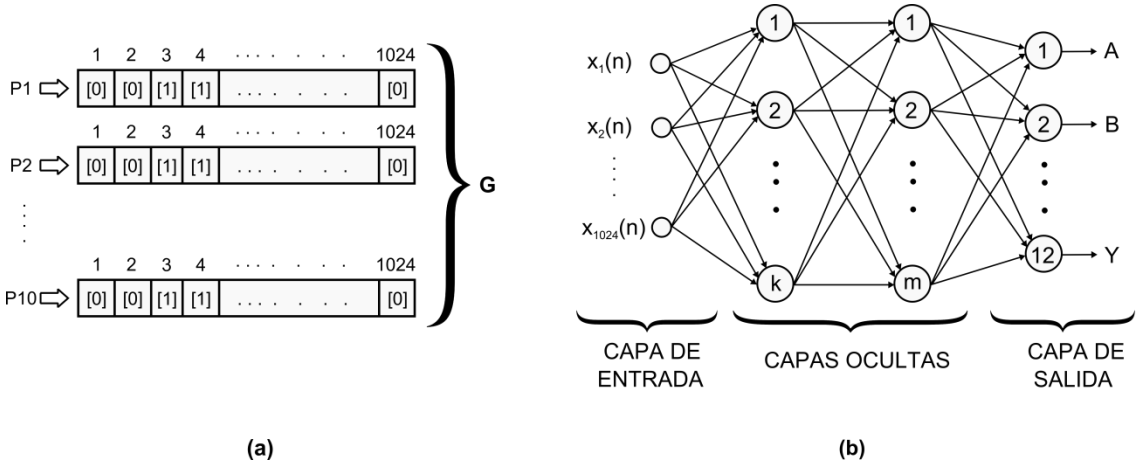


Figura 6. (a) Patrones de entrenamiento para la letra G; (b) Arquitectura de la Red Neuronal.

4.3. Modelo de RNA

El procedimiento para la selección del mejor modelo se basa en la comparación de los valores promedio obtenidos en cada modelo con respecto al reconocimiento de todas las letras. Dado que el valor ideal es 1, se nota que el Modelo 2 es el más adecuado con un valor promedio de 0.9263 (véase Tabla 1).

4.4. Experimentación: Pruebas y Resultados

Una vez que se ha establecido el conjunto de entrenamiento y el modelo de red neuronal, se consideró un total de 100 fotografías de la mano del sujeto realizando la seña de una sola letra, definiendo el conjunto de pruebas. Se consideraron diferentes versiones de cada letra, cambiando sutilmente los parámetros de articulación física de cada una, como por ejemplo la apertura de la mano, posición, ángulo, ubicación de los dedos, etc. Los resultados obtenidos se muestran en la Tabla 2.

Tabla 1. Tabla modelos con sus valores a las salidas para cada letra.

Modelo	Neuronas en capa oculta 1	A	B	C	D	G	I	L	O	Q	V	W	Y	Prom.
1	32	0.8976	0.4953	0.9810	0.7955	0.9877	0.9836	0.9775	0.9350	0.9798	0.9776	0.9713	0.7674	0.8958
2	64	0.9160	0.9200	0.9317	0.9210	0.9598	0.9644	0.9650	0.8890	0.9777	0.9560	0.8732	0.8423	0.9263
3	24	0.6687	0.8556	0.9645	0.9762	0.9852	0.9834	0.9770	0.9711	0.9687	0.9893	0.9000	0.6352	0.9062
4	40	0.9472	0.9578	0.9734	0.9635	0.9678	0.9386	0.9774	0.0543	0.9778	0.9929	0.9134	0.4484	0.8427
5	48	0.8024	0.9344	0.9656	0.9607	0.9858	0.9793	0.9759	0.9665	0.9736	0.9808	0.4714	0.9312	0.9106
6	56													
7	72	0.8988	0.2591	0.9876	0.8882	0.9727	0.9404	0.9761	0.5493	0.8872	0.9780	0.9295	0.8750	0.8452
8	80	0.9677	0.0600	0.9827	0.9099	0.9692	0.9826	0.9544	0.9600	0.9778	0.9877	0.7775	0.9295	0.8716
9	88	0.9403	0.7503	0.9616	0.9634	0.9815	0.4847	0.9797	0.7188	0.8968	0.9825	0.9272	0.9767	0.8803
10	96	0.9452	0.8474	0.8000	0.7800	0.9743	0.8796	0.9659	0.9652	0.9264	0.9713	0.9089	0.7555	0.8933
Prom.		0.7984	0.6080	0.8548	0.8158	0.8784	0.8137	0.8749	0.7009	0.8566	0.8816	0.7672	0.7116	

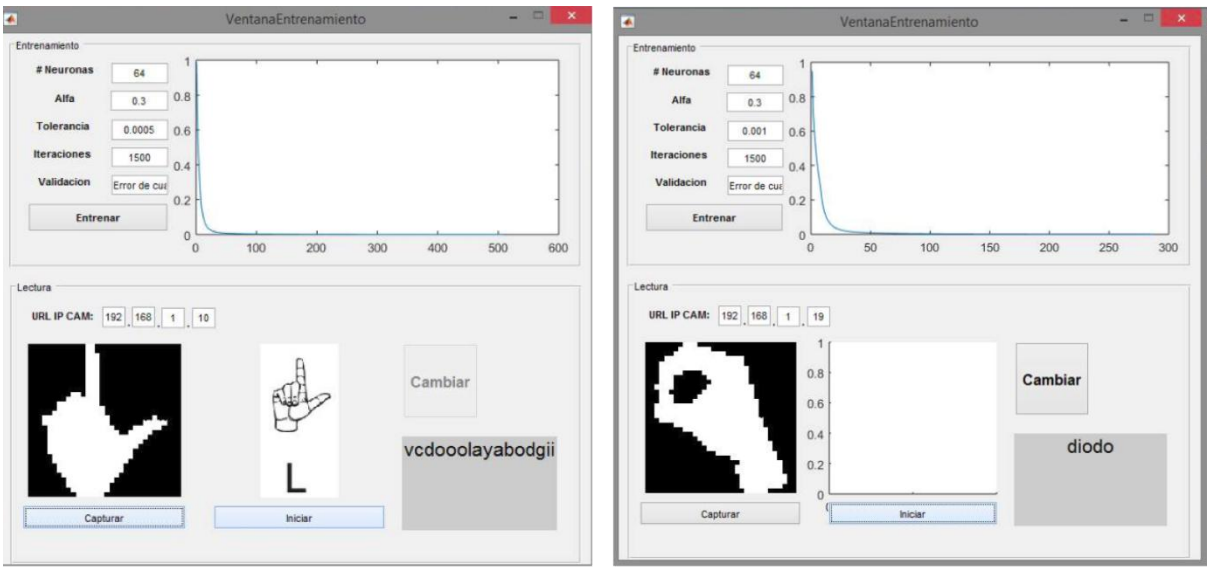
Tabla 2. Resultados de Pruebas Realizadas.

Letra (seña)	Imágenes de Prueba	Aciertos	Errores
A	100	99	1
B	100	100	0
C	100	66	34
D	100	77	23
G	100	89	11
I	100	100	0
L	100	94	6
O	100	96	4
Q	100	97	3
V	100	88	12
W	100	62	38
Y	100	90	10

Las letras más difíciles de clasificar son la *C* y la *W*, mientras que las letras menos propensas a equivocación son, entre otras, la *B* y la *I*. La causa del relativo bajo rendimiento de la red para ciertas letras se debe particularmente a su forma, puesto que durante el proceso de captura de imágenes se notó la presencia de sombras adicionales que no se presentan en otros caracteres y por tanto requieren un tratamiento adicional para compensar diferentes condiciones de iluminación. Experimentalmente se encontró que los resultados mejoran si durante la etapa de pruebas se utiliza una iluminación similar a aquella presente en la etapa de entrenamiento. Los resultados reflejan un porcentaje de acierto superior al 88%.

4.5. Interfaz de Usuario

El proceso de pruebas se realiza mediante un interfaz de usuario que permite observar los resultados de clasificación. La Figura 7(a), muestra el interfaz de usuario del programa, particularmente el instante en el que la letra *L* ha sido ingresada en la maqueta de adquisición y el sistema la reconoce correctamente.



(a) (b)

Figura 7. (a) Reconocimiento de la letra L; (b) Escritura de la palabra *diodo*.

4.6. Módulo de concatenación de caracteres

Este módulo interpreta una cadena de señas realizadas por el sujeto de prueba con el objetivo de escribir en pantalla una palabra resultado del reconocimiento de las señales. El botón *Iniciar* del interfaz de usuario inicia el reconocimiento de la seña presente en el campo de visión de la cámara, y además muestra un elemento visual con la palabra Cambiar que se ilumina por un instante después de un intervalo de tiempo para advertir al sujeto que puede articular la siguiente letra. La Figura 7(b) muestra el funcionamiento del programa cuando el usuario termina de escribir la palabra *diodo*. Durante el tiempo transcurrido t_m entre un determinado *aviso de cambiar* y el siguiente, no es posible determinar el instante preciso en el que el sujeto termina de articular una letra determinada con su mano, por lo que se propone un procedimiento en el que se capturan varias fotografías durante t_m, las mismas que se ingresan a la red neuronal para obtener un resultado de reconocimiento de cada una. Así, se integra un mecanismo de votación basado en los resultados de reconocimiento separados en tiempo. La Figura 8 ilustra el procedimiento en el que varias capturas de la mano del sujeto son reconocidas como la letra *D*, a excepción de una que es reconocida como *B*, con lo que se puede evidenciar la utilidad de un proceso de votación. Se estima que este simple mecanismo facilita la escritura de palabras a través de este sistema, sin embargo existen limitaciones relacionadas al número de caracteres analizados y a la necesidad de incorporar un mecanismo adicional que resuelva el problema de las letras que se producen con movimiento de la mano.

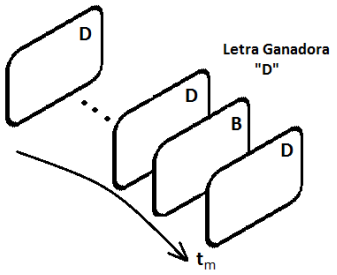


Figura 8. Procedimiento de selección de letra reconocida mediante votación.

5. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se ha presentado un sistema de reconocimiento visual de caracteres dactilológicos basado en RNA, inspirado fundamentalmente en el desarrollo de sistemas de ayuda a personas que no tienen la posibilidad de oír y/o hablar. Se concluye que en el alfabeto dactilológico existen letras que tienen señales muy similares visualmente entre sí lo que dificulta su discriminación y por tanto no fueron consideradas en este prototipo preliminar. Los futuros esfuerzos se centran en la determinación de un conjunto de características que permitan discriminar todas las letras del alfabeto, así como la incorporación de métodos de reducción de la dimensionalidad como el análisis de componentes principales PCA (Principal Component Analysis). Adicionalmente, se ha identificado la necesidad de un procedimiento que posibilite la extracción de características de las letras que se representan mediante movimiento de la mano y con ello, un procedimiento de evaluación y validación adecuado.

REFERENCIAS

Barkoky, A., N. Charkari, 2011. *Static Hand Gesture Recognition of Persian Sign Numbers using Thinning Method. Multimedia Technology (ICMT)*. International Conference on, IEEE, 6548-6551.

Dirección de Investigación de la Universidad de Cuenca, 2014. *Directrices para la elaboración de artículos científicos*. Revista MASKANA. DIUC, Universidad de Cuenca.

Hecht-Nielsen, R., 1989. *Theory of the backpropagation neural network*. Neural Networks, IJCNN, International Joint Conference on, IEEE, 593-605.

Heng, D., T. TszHang, 2012. *Hand Gesture Recognition Using Kinect*. Software Engineering and Service Science (ICSESS), IEEE 3rd International Conference on, 196-199.

Incertis, I., J. Garcia-Bermejo, E. Casanova, 2006. *Hand Gesture Recognition for Deaf People Interfacing*. Pattern Recognition, ICPR, 18th International Conference on, Vol.2, IEEE, 100-103.

Jones, A.S., 1999. On the Complexity of Computing. *Advances in Computer Science*, 555-566.

Knuth, D.E., 1984. *The Text Book*. Publisher: Addison-Wesley.

Otsu, N., 1979. *A Threshold Selection Method from Gray-Level Histograms*. IEEE Transactions on Systems, Man and Cybernetics, 62-66.

Priego-Perez, F., 2012. *Reconocimiento de Imáenes del Lenguaje de Señas Mexicano*. Tesis de Maestría en Ciencias de la Computación, Insituto Politécnico Nacional, Centro de Investigación en Computación, México.

Razo-Gil, L., G. Salvador-Calderon, R. Barrón-Fernández, 2009. *Sistema Traductor Para el Reconocimiento del Alfabeto Dactilológico*. CIC-IPN: Centro de Investigación en Computación, Laboratorio Inteligencia Artificial.

Renault, R.B., 1991. *3D Hierarchies for Animation*. John Wiley & Sons Ltd.

Wah Ng, C., S. Ranganath, 2002. Real-time gesture recognition system and application. *Image and Vision Computing*, 20(13), 993-1007.

Zurada, J.M., 1992. *Artificial Neural Systems*. West Publishing Company.