

Reconocimiento del lenguaje de señas mediante aprendizaje automático para niños de primaria

Sign language recognition with machine learning for elementary school children

Richard Bodmer¹, Lucia Liu¹, Winson Liu¹, José Carlos Rangel^{2*}

¹Licenciatura en Ingeniería en Sistemas y Computación – Campus Víctor Levi Sasso– Universidad Tecnológica de Panamá

²ROBOTSIS - Facultad de Ingeniería de Sistemas Computacionales – Universidad Tecnológica de Panamá

Resumen Los gestos de las manos son utilizados para demostrar expresividad, un punto mucho más importante para personas con discapacidad auditiva, quienes utilizan esta representación de símbolos manualmente como lenguaje. Un sistema que reconozca el lenguaje de señas panameño que esté orientado hacia niños con discapacidades auditivas es relevante para un aprendizaje más fluido y efectivo. Para el desarrollo, se deberán usar herramientas de inteligencia artificial, como aprendizaje automático, procesador y clasificador de imágenes, además del uso de una cámara digital para capturar los gestos expresados con las manos. Como objetivo, este sistema solamente reconocerá las vocales del alfabeto (a, e, i, o, u) en texto como etapa de inicio y prueba.

Palabras clave Aprendizaje automático, clasificador de imágenes, gestos, lenguaje de señas.

Abstract For most people, hand gestures are used to demonstrate expressiveness for people with hearing disabilities. This manual representation of symbols constitutes a language. A system that recognizes Panamanian sign language that is oriented towards children with hearing disabilities is relevant to a more fluid and effective learning. For development, artificial intelligence tools should be used, such as machine learning, image processing and classifier, as well as the use of a digital camera to capture gestures expressed by hands. As an objective, this system will only recognize the vowels of the alphabet (a, e, i, o, u) in text as a start and test stage.

Keywords Machine learning, image classifier, gestures, sign language.

* Corresponding author: jose.rangel@utp.ac.pa

1. Introducción

“Las lenguas de señas son lenguas naturales de producción gestual y percepción visual que tienen estructuras gramaticales perfectamente definidas y distintas de lenguas orales con las que cohabitan.

Hasta el presente la intervención tecnológica usada como herramientas para el apoyo de la población que presenta discapacidad auditiva ha tenido un gran efecto, y ha permitido mejorar en la calidad de vida de este sector de la población; con el continuo avance de la se hace posible encontrar más y mejores soluciones que buscan el desarrollo de nuevos dispositivos orientados a esta población.” [1]. Entre las herramientas creadas están los intérpretes virtuales, que realizan la traducción de la lengua de señas británica; traductores de lenguas de señas en línea; también se ha desarrollado un dispositivo japonés que permite que una persona con discapacidad auditiva, que utiliza lenguaje de señas, pueda conversar con otra persona que no está familiarizada con este lenguaje [1].

Anteriormente, se desarrolló una mano robótica con teleoperación 3D. Se captura la imagen de la posición de una mano, este aparato infiere con precisión la posición tridimensional de las articulaciones y refleja la pose [2].

Nuestra propuesta se enfoca en el uso técnicas de Inteligencia Artificial para crear un sistema de reconocimiento de lenguaje de señas para niños de escuelas primarias, que presenten o no discapacidades auditivas, de tal manera que se incentive la enseñanza de las vocales del alfabeto de una manera diferente. En este proceso las herramientas a utilizar serán un clasificador de imágenes, equipo de procesamiento y aprendizaje automático.

En Panamá, al 2010, había 15,191 personas con discapacidades auditivas, de las cuales al menos 6,000 utilizan el lenguaje de señas como forma de comunicación [3]. Este lenguaje conforma el idioma por el cual las personas con deficiencias auditivas pueden comunicarse.

Es muy importante destacar que: “La Escuela Nacional de Sordos, en 1951, crea el Instituto Panameño de Rehabilitación Especial (IPHE), esto con la finalidad de fomentar la

habilitación de los jóvenes panameños que padeciesen algún tipo de retardo mental, sordera y ceguera en las Escuelas de Enseñanza Especial, la Escuela Nacional de Sordos y la Escuela de Ciegos, respectivamente.” [4]. Esto demuestra que la educación a las personas con discapacidades auditivas en Panamá forma un ambiente de exclusión y los aparta del mundo de personas oyentes, al no crear un espacio de inclusión educativa, social y profesional de la comunidad de personas sordas.

“En Panamá faltan estudios e investigaciones sociolingüísticas y pedagógicas que sustenten los necesarios cambios en el marco de la inclusión educativa y la educación especial con enfoque de apoyo.

Otro de los tantos factores que pueden incidir en la persistencia de dicha situación que va en detrimento de la población sorda, es la discriminación y exclusión de los por parte de los grupos que integran la sociedad. Otra dimensión para evaluar y considerar es la persistencia del fantasma de lo clínico que se presenta en las instituciones de educación especial, fantasma que también incluye los enfoques de especialización y habilitación además de un sistema duplicado de ‘educación inclusiva’ especial en el Ministerio de Educación, el cual requiere una evaluación integral.” [4].

Con el desarrollo de la herramienta propuesta, incentivamos que la lengua de señas sea aprendida desde temprana edad, específicamente para niños de primaria, de manera fácil y divertida.

Queremos que las personas sordomudas se sientan muy acogidas por la comunidad de los oyentes y viceversa, inculcando esta habilidad en los niños, ya sea con discapacidad auditiva o no.

1.1 Objetivo general

Desarrollar un sistema de reconocimiento de lenguaje de señas por medio de herramientas de aprendizaje automático, orientado al reforzamiento del aprendizaje para niños de primaria ya sea con o sin discapacidad.

1.2 Objetivos específicos

- Implementar la captura de video en tiempo real para el reconocimiento de imágenes.
- Evaluar algoritmos de predicción y herramientas de reconocimiento de patrones y descriptores de imágenes.
- Entrenar el sistema de aprendizaje automático con las vocales del alfabeto en lenguaje de señas [5].
- Aplicar pruebas de reconocimiento de vocales en lenguaje de señas a personas con o sin discapacidad y obtener resultados con alto grado de precisión y rapidez.
- Crear una herramienta que incentive a las personas a comunicarse en lenguaje de señas.

1.3 Propuesta

La solución propuesta se basa en un sistema que permita el reconocimiento del lenguaje de señas panameñas, el cual logre

reconocer, en primera instancia, las vocales del alfabeto con un alto nivel de precisión y construya una representación textual, siendo extensible a las demás letras del alfabeto y señas que signifiquen palabras o frases. Se pretende orientarlo hacia niños de primaria y que estos puedan ir familiarizándose con el lenguaje de señas y aprenderlas a lo largo del transcurso de su desarrollo académico sin que haya una barrera de comunicación (figura 1).

1.4 Justificación

- Para que los niños de primaria, ya sean con discapacidades auditivas o no, puedan aprender este lenguaje y comunicarse entre sí.
- Reducir la brecha de comunicación que hay entre las personas con discapacidades auditivas y personas oyentes, para así evitar la exclusión en actividades sociales, académicas y laborales.
- Para que las personas sordas no pierdan su esencia de comunicación, que es la lengua de señas, cuando se comunican con personas oyentes.
-

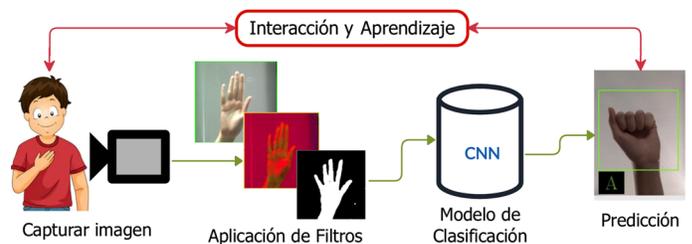


Figura 1. Sistema de funcionamiento planteado.

2. Materiales y métodos

La creación de nuestro sistema consta en reconocer, en primera instancia, las vocales del alfabeto en lenguaje de señas. Con este objetivo, se procedió a investigar diversos planteamientos utilizados para el reconocimiento de objetos con visión artificial para la elección de herramientas y procedimientos.

2.1 Obtención de imágenes

Se capturaron gestos de las vocales del alfabeto en lenguaje de señas con iluminación controlada a través de lámparas que apuntaban directamente a la persona que hacía la seña, fondo blanco, imágenes estáticas para el entrenamiento y reconocimiento de gestos con una cámara web.

Igualmente se pudieron obtener más muestras en línea para así obtener diversidad en diferentes variables de entrenamiento del modelo. Estas muestras se obtuvieron del dataset público.

Sign Language Recognition donde se encuentran imágenes de señas realizadas con las manos (: https://github.com/shadabsk/Sign-Language-Recognition-Using-Hand-Gestures-Keras-PyQT5-OpenCV/tree/master/Source%20Code/Dataset/test_set/A)

Luego de la recopilación de las imágenes, se obtuvo un dataset compuesto por 2000 imágenes, balanceado para cada una de las vocales usadas en la construcción del modelo.

2.2 Procesamiento y corrección de imágenes

Se aplicaron filtros de saturación y color [6], transformación a fondo negro y el gesto de la mano a blanco, recortada la imagen en una resolución 64x64 píxeles. Esto se emplea para las imágenes de entrenamiento, prueba y a la captura de imagen en tiempo real.

En el procesamiento de imágenes, se debe generar un vectorizado de sus características (feature extraction [7],[8]). El uso de los filtros mencionados nos permite que se produzcan valores más uniformes y en un rango similar dentro de tales vectorizados o representaciones; utilizando matices de colores en estos, dichos valores pueden diferir (figura 2).

Las imágenes de cada categoría se agruparon por carpetas (cada vocal). Esta división permite asignar las etiquetas a cada grupo de imágenes (vocales del alfabeto; A, E, I, O, U). Esto quiere decir que, al momento de entrenar o probar el modelo, los grupos de imágenes ya estarán asignadas a su vocal correspondiente.

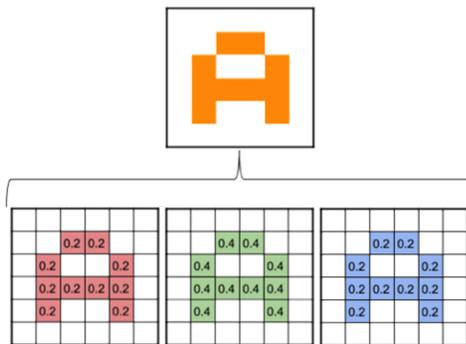


Figura 2. Diferencia entre las matrices de colores y los posibles valores en sus vectores [8].

2.3 Selección de herramientas y el modelo de predicción

Se utilizó la librería OpenCV (<https://opencv.org>) para todo el procesamiento y filtrado de imágenes por captura de video en tiempo real, ventanas para la visualización de la cámara y controles de ajustes (figura 3).

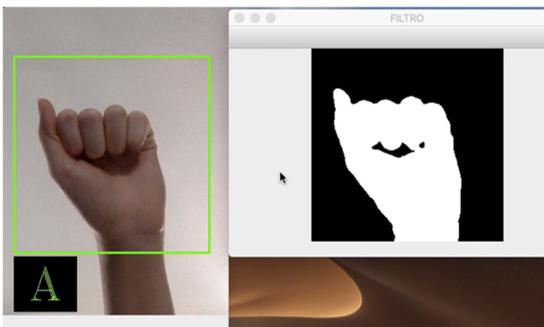


Figura 3. Interface de OpenCV para la captura de video.

Se utilizó un espacio estático delimitado por un cuadro para mantener el tamaño de resolución en 64x64 píxeles.

Se aplican los filtros, ya que la imagen cruda se forma de más matrices de colores. El difuminado gaussiano demostró ser más efectivo para suavizar los bordes de la mano y finalmente, la extracción de colores se aplica para eliminar variaciones en RGB y HSV y procesarlo en escala blanco y negro (figura 4).

Para la predicción de señas, se entrenaron clasificadores utilizando los algoritmos *K-nearest neighbors* (KNN) [9], *Support Vector Machines* (SVM)[10],[11], *Random Forest* [12] y *Naives-Bayes Gaussiano* [13].

Se utilizó también un modelo de red neuronal convolucional (CNN), el cual es comúnmente utilizado para el análisis visual de imágenes.

Finalmente se compararon los modelos con un porcentaje de precisión más alto para probar el sistema de predicción, implementándose el de mayor valor.

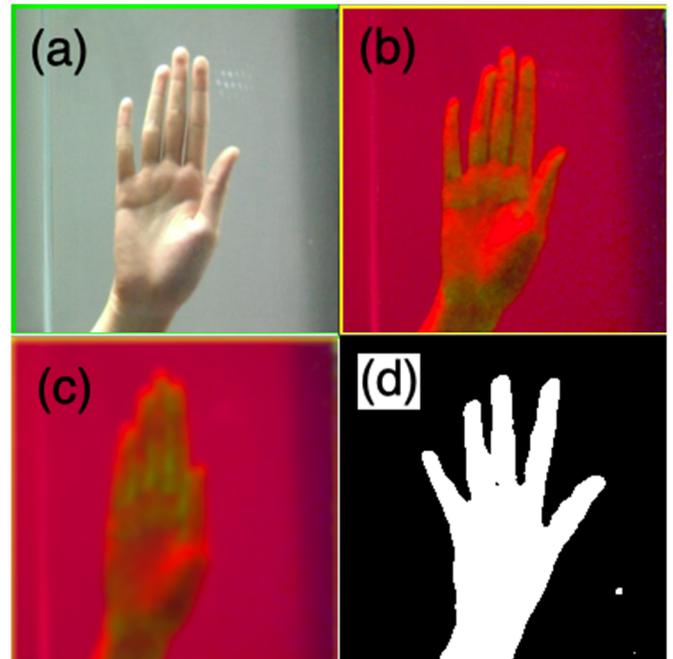


Figura 4. (a) Captura del gesto original. (b) Filtro HSV (saturación - color). (c) Difuminado gaussiano. (d) Extracción de rango de colores.

2.4 Entrenamiento

Para el entrenamiento y prueba de precisión se utilizaron 1750 y 250 imágenes respectivamente. Las mismas se utilizaron para todos los modelos propuestos.

2.5 Pruebas del sistema de predicción

Se realizaron pruebas de los clasificadores propuestos y finalmente con el modelo de red neuronal convolucional. Luego, con este último modelo, sometimos a sujetos de prueba para demostrar la precisión de las predicciones. Estas pruebas consistieron en ubicar al sujeto de prueba frente a un computador que se encontraría ejecutando la aplicación. Al

mismo se le pediría que realizase la seña correspondiente a una vocal y comprobar que el sistema reconocía adecuadamente esta seña, este reconocimiento se realizaría en tiempo real. Esto se repetiría al menos 10 veces utilizando todas las vocales y en un orden aleatorio. Los sujetos de prueba eran personas adultas con discapacidad auditiva y personas sin discapacidad, que ya conocían el lenguaje de señas panameño, de igual manera se probó con niños que cursan la escuela primaria.

3. Resultados y discusión

Los resultados de precisión para los modelos de clasificación se muestran en la tabla 1.

Tabla 1. Precisión de los modelos de clasificación elegidos

Modelo	Precisión (%)
K-Nearest Neighbor (3)	48.124
SVM	43.312
Naive-Bayes	48.369
Gaussiano	
Random Forest	62.724
CNN	99.95

Con el modelo de red neuronal convolucional se obtuvo una pérdida de 0.18 y una precisión de 99.95%.

3.1 Modelos finales propuestos

Para las predicciones con el modelo de *Random Forest* se lograron algunas señas acertadas, sin embargo, no se obtuvo el grado de precisión buscado, al contrario del modelo CNN, que si obtuvo resultados precisos en todas las pruebas (figura 5).

Por tal razón, implementamos el modelo CNN en el sistema de predicción de señas. Analizando los resultados obtenidos por los demás clasificadores, uno de los posibles factores que pudo influir en la precisión obtenida es el proceso seguido para la extracción de las características de las imágenes del *dataset*, por lo cual, la representación o descriptor de la imagen no encapsulaba debidamente la información de esta. De igual manera, la cantidad de imágenes utilizadas en el *dataset*, es comúnmente un factor a tomar en cuenta en la construcción de modelos, ya que, a mayor variabilidad en los datos de entrenamiento, puede hacer que la capacidad de generalización de los modelos sea superior. Un tercer elemento que puede influir en la ejecución de los modelos son los parámetros de entrenamiento para cada modelo, las librerías definen algunos parámetros de forma general, para que sean modificados por los usuarios, de tal manera que estos solo deban ajustar los parámetros básicos requeridos para el algoritmo. Por ende, se requeriría una búsqueda exhaustiva para encontrar, tanto los parámetros a modificar, como los valores que se deben asignar.

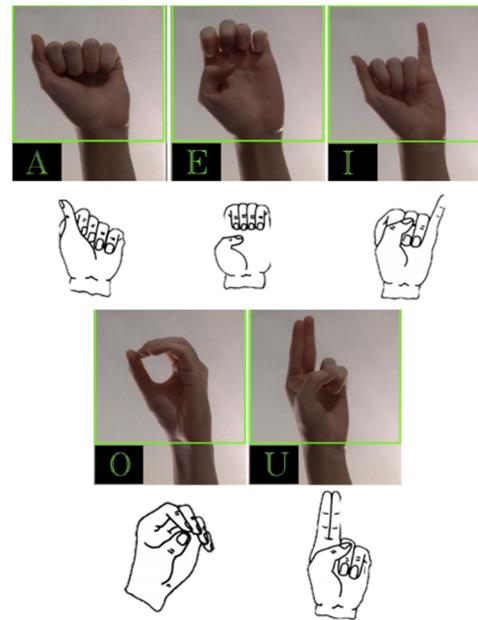


Figura 5. Reconocimiento de las vocales del alfabeto en lenguaje de señas utilizando una CNN.

Generalmente, al tener otros objetos que provocan ruido dentro de la sección delimitada que se utiliza como zona de reconocimiento, el modelo podría tener menos precisión. Sin embargo, se obtenían predicciones exactas al tener un fondo claro, uniforme y con condiciones de luz favorable e igualmente tomando en cuenta que, en la ventana del filtro, estuviera negro en su totalidad y la mano mostrada apareciera en blanco uniformemente.

3.2 Resultados con sujetos de prueba

En todas las pruebas el prototipo logró reconocer las vocales en lengua de señas con un alto grado de precisión. En estas pruebas el clasificador acertaba aproximadamente el 90% de las señales mostradas por los sujetos de prueba.

4. Conclusiones

Como el sistema solamente se limita a reconocer las señas de manos que representan las vocales del alfabeto, tomará menos tiempo para que aprenda y reconozca las mismas. Además, se podrá obtener un grado de precisión alto en menos tiempo.

Las pruebas desarrolladas con el modelo entrenado demuestran que la utilización de las CNN permite construir un sistema robusto y efectivo para la enseñanza/aprendizaje de la lengua de señas.

A pesar de que el sistema solamente reconoce las señas de las vocales, se busca que el sistema pueda reconocer también las consonantes para así tener el alfabeto completo. Luego de esto, se quiere que el sistema pueda reconocer movimientos complejos de las manos, es decir, las señas que involucran movimiento, para interpretar las mismas en palabras.

Además de esto, se puede diseñar una interfaz que sea amigable y más interactiva para lograr el aprendizaje de los niños.

Se requiere crear herramientas o métodos que impulsen una educación de calidad e inclusión hacia la comunidad de personas con discapacidad auditiva, en especial desde temprana edad, para así asegurar su participación íntegra y exitosa en la comunidad social y profesional.

AGRADECIMIENTOS

En este trabajo, José Carlos Rangel ha sido apoyado con fondos del Sistema Nacional de Investigación (SNI) del SENACYT.

REFERENCIAS

- [1] C. Hernández, J. L. Pulido, y J. E. Arias, "Las tecnologías de la información en el aprendizaje de la lengua de señas", Revista de Salud Pública, vol. 17, pp. 61-73, 2015. doi: <http://dx.doi.org/10.15446/rsap.v17n1.36935>
- [2] F. Gomez-Donoso, S. Orts-Escolano, y M. Cazorla, "Accurate and efficient 3D hand pose regression for robot hand teleoperation using a monocular RGB camera", Expert Systems with Applications, vol. 136, pp. 327-337, 2019, doi: <https://doi.org/10.1016/j.eswa.2019.06.055>.
- [3] A. Oviedo, (2015) "Panamá, atlas sordo – Cultura Sorda". [En línea] Disponible: <https://cultura-sorda.org/panama-atlas-sordo/> [Accedido: 23-jun-2019]
- [4] A. Arjona. (2017). "Desafíos de la educación a personas sordas en Panamá". [En línea]. Disponible: <http://laestrella.com.pa/panama/politica/desafios-educacion-personas-sordas-panama/24026257> [Accedido: 15-may-2019]
- [5] D. Pimentel, R. Walker y M. Fajardo. *Lengua de Señas Panameñas*. Panamá: Editora Panamá América, S.A. 2018. Pp 17-18.
- [6] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat, and J. Jatakia (2013). "Human Skin Detection Using RGB, HSV and YCbCr Color Models". Advances in Intelligent Systems Research, vol 137 pp 324-332, 2017.
- [7] A. Rosebrock. (2014). "Charizard Explains How To Describe and Quantify an Image Using Feature Vectors". [En línea]. Disponible: <https://www.pyimagesearch.com/2014/03/03/charizard-explains-describe-quantify-image-using-feature-vectors/> [Accedido: 4-may-2019]
- [8] A. Nikishaev. (2018). "Feature extraction and similar image search with OpenCV for newbies". [En línea]. Disponible: <https://medium.com/machine-learning-world/feature-extraction-and-similar-image-search-with-opencv-for-newbies-3c59796bf774> [Accedido: 14-may-2019]
- [9] A. Rosebrock. (2016). "k-NN classifier for image classification". [En línea]. Disponible: <https://www.pyimagesearch.com/2016/08/08/k-nn-classifier-for-image-classification/> [Accedido: May. 4 2019]
- [10] DataTurks (2018). "Understanding SVMs: For Image Classification". Revista electrónica Medium. [En línea]. Disponible: <https://medium.com/@dataturks/understanding-svm-for-image-classification-cf4f01232700> [6-may-2019]
- [11] C. Cortes y V. Vapnik, "Support-vector networks", Machine Learning, vol. 20, n.º 3, pp. 273-297, sep. 1995, doi: [10.1007/BF00994018](https://doi.org/10.1007/BF00994018)
- [12] L. Breiman, "Random Forests", Machine Learning, vol. 45, n.º 1, pp. 5-32, oct. 2001, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324)
- [13] P. Domingos y M. Pazzani, "On the Optimality of the Simple Bayesian Classifier under Zero-One Loss", Machine Learning, vol. 29, n.º 2, pp. 103-130, nov. 1997, doi: [10.1023/A:1007413511361](https://doi.org/10.1023/A:1007413511361)