

Francisco Morillas-Espejo, Ester Martinez-Martin
 {francisco.morillas,ester}@ua.es

Motivación

La sociedad actual se caracteriza por ser mayoritariamente oyente, tanto a nivel nacional como internacional, lo cual puede generar una barrera comunicativa para una pequeña parte de la población, el colectivo sordo.

Esta barrera comunicativa se puede presentar en lugares cotidianos como pueda ser una consulta médica o el transporte público. En estas situaciones, el uso de las nuevas tecnologías y de la inteligencia artificial pueden ayudar a mitigar esta barrera, facilitando la comunicación entre toda la sociedad, tanto oyente como sorda.

En este ámbito se centra la presente investigación, proponiendo un sistema de interpretación para la Lengua de Signos Española (LSE) haciendo uso de diferentes técnicas de visión por computador y de aprendizaje profundo (*deep learning*).

Descripción del Proyecto

Como punto inicial, la investigación se ha centrado en el reconocimiento del alfabeto dactilológico. Motivado por al combinación de signos estáticos y en movimiento.

También cabe destacar que, con el objetivo de generar un sistema de bajo coste y al alcance de todos, se trabaja con imágenes en color y sin ningún tipo de segmentación por color de piel. Permitiendo su empleo con cualquier tipo de cámara cotidiana y por cualquier persona.

Adicionalmente, el sistema se ha desarrollado teniendo en cuenta que una persona signa con su mano dominante, por lo que el sistema es capaz de identificar el signo tanto para diestros como para zurdos.

Con el fin de llevar a cabo la identificación de cada signo, primero es necesario detectar el esqueleto de la persona para posteriormente identificar las manos y brazos. Con estos puntos detectados la imagen se codifica con color azul para la parte izquierda del cuerpo y rojo para la parte derecha. Estas imágenes son las que constituyen el conjunto de datos.

Resultados

Con el conjunto de datos generado se procede a realizar el entrenamiento de las diferentes arquitecturas de redes neuronales. La investigación se ha centrado particularmente en dos tipos, las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN). De las cuales se han entrenado 4 modelos existentes en el estado del arte y 10 modelos propios.

Tras el análisis de las diferentes arquitecturas se alcanza un máximo de **78.81%** de precisión en el conjunto de test, es decir, en imágenes nunca antes vistas.

Esta tasa de acierto se considera buena ya que las principales confusiones suceden entre pares de letras donde la diferencia es su movimiento como pueden ser los casos: *l-ll*, *n-ñ*, *r-rr* o *u-v*. Además de las letras *f-t* donde, sin el uso de una cámara de profundidad no es posible solventar el error.

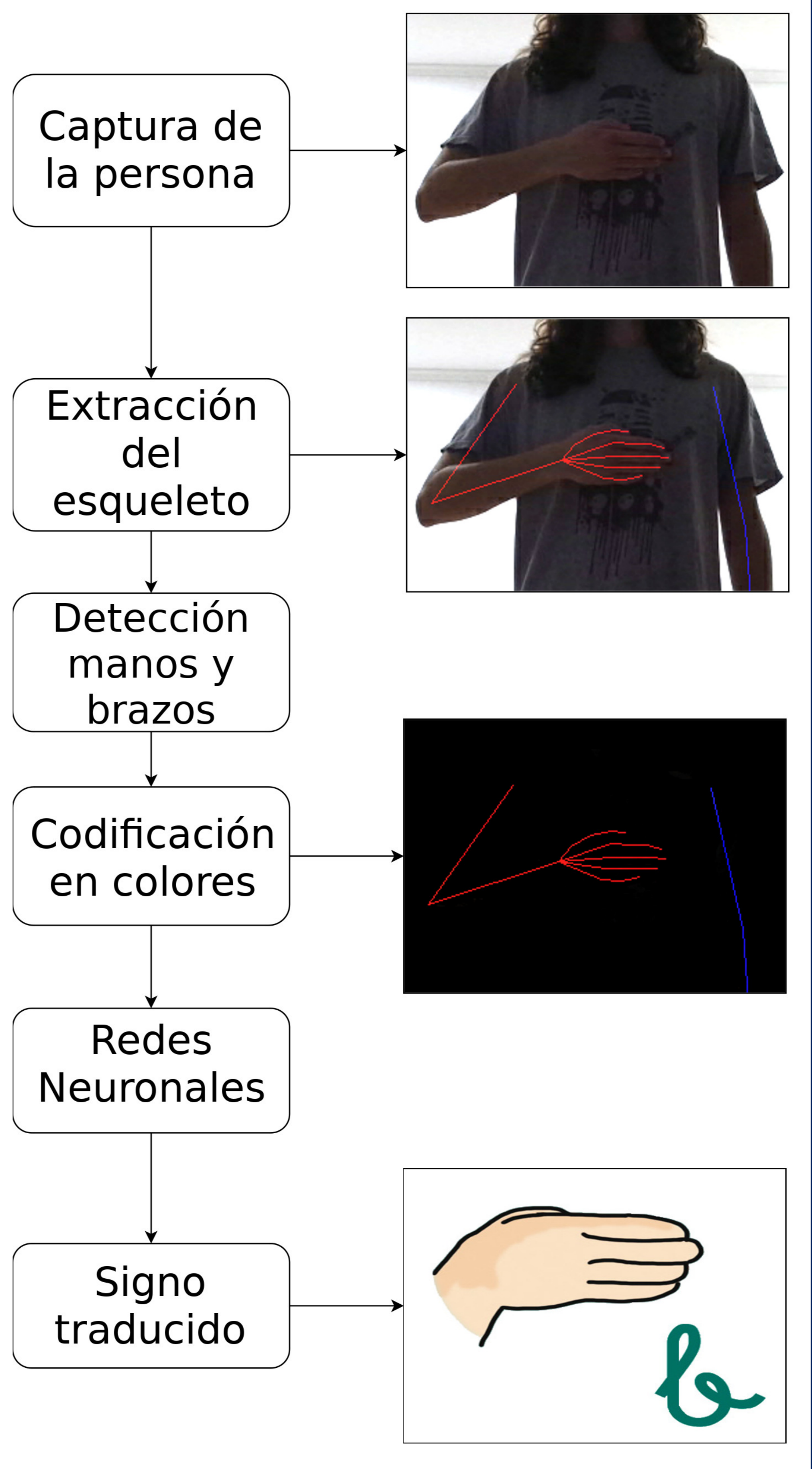
Conclusiones

La literatura existente muestra que no existe, actualmente, ningún sistema similar relacionado con la interpretación de la Lengua de Signos Española. En su lugar, existen aportes en otros idiomas, destacando la lengua de signos americana (ASL).

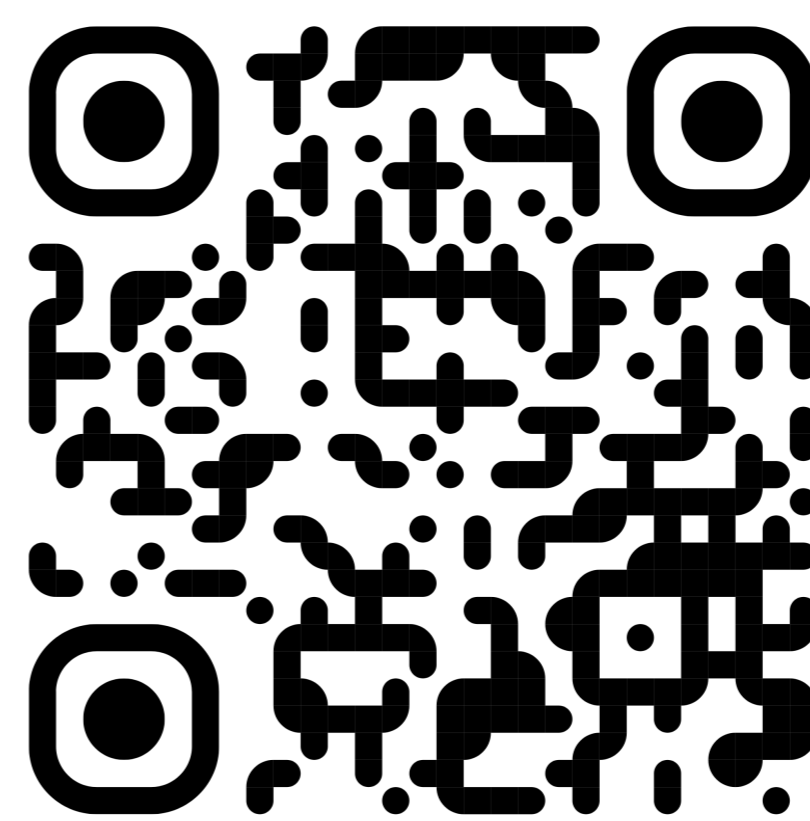
Los resultados experimentales muestran una buena precisión, **78.81%**, la cual posibilita su uso en tiempo real. Aunque este resultado se corresponde con el alfabeto dactilológico, se está trabajando en una versión con un vocabulario específico perteneciente al campo de las tareas cotidianas donde interpretar frases completas.

Adicionalmente se está trabajando en un avatar virtual para poder realizar la traducción en sentido opuesto, esto es, de castellano a Lengua de Signos Española, lo cual constituye un sistema de interpretación completo. Todo ello con el objetivo de alcanzar una comunicación sin barreras.

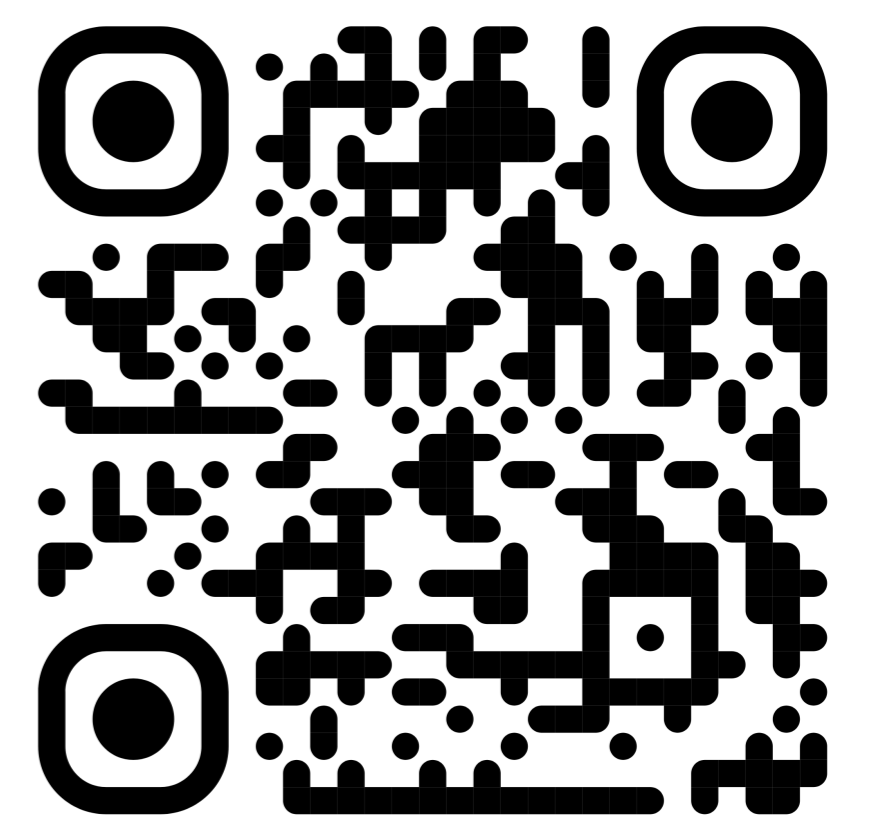
Estructura a seguir



Enlaces



Ejemplo del sistema



Explicación signada