INFORME FINAL DEL PROYECTO

Sistema de Reconocimiento de Glosas en Lengua de Señas Colombiana (LSC) usando RNN y Streamlit

Autor: Camilo Augusto Alvarado Rueda, Juan David Rodriguez

Curso: *Ciencia de Datos*

1. Resumen

Este proyecto presenta el desarrollo de un prototipo funcional capaz de reconocer glosas de la Lengua de Señas Colombiana (LSC) a partir de videos cortos mediante técnicas de visión por computador y redes neuronales recurrentes (RNN).
El sistema permite subir un video de aproximadamente 2 segundos a través de una interfaz desarrollada en Streamlit, procesa la secuencia mediante landmarks corporales y de manos usando MediaPipe Holistic, y finalmente predice la glosa más probable basada en un modelo entrenado.

Aunque el modelo actual presenta un nivel de confianza moderado y todavía requiere mayor entrenamiento para mejorar su precisión, este prototipo demuestra la viabilidad de construir herramientas automáticas para apoyo educativo e interpretación básica de señas en LSC.

2. Introducción

La Lengua de Señas Colombiana (LSC) es el principal medio de comunicación para la comunidad sorda en Colombia. Su difusión, estudio y apoyo tecnológico se han convertido en una necesidad creciente, especialmente en contextos educativos y de inclusión social.

El reconocimiento automático de señas es un desafío técnico complejo debido a la variabilidad en movimientos, velocidad, ángulos de grabación, iluminación y diferencias entre personas. Sin embargo, los avances recientes en visión por computador y aprendizaje profundo han permitido desarrollar modelos capaces de interpretar patrones en secuencias de video.

El propósito de este proyecto es crear un prototipo capaz de reconocer glosas aisladas a partir de videos cortos, utilizando una arquitectura simple y accesible. El sistema final

incluye una interfaz amigable que permite cargar videos y obtener una predicción inmediata acompañada de su nivel de confianza.

3. Objetivos

Objetivo general

Desarrollar un sistema prototipo que reconozca glosas de la Lengua de Señas Colombiana (LSC) usando un modelo de aprendizaje profundo y una interfaz web interactiva.

Objetivos específicos

- Procesar videos cortos extrayendo características corporales mediante MediaPipe Holistic.
- Entrenar un modelo basado en redes neuronales recurrentes para clasificación de glosas.
- Implementar una interfaz web sencilla usando Streamlit para cargar videos y mostrar predicciones.
- Evaluar el desempeño básico del sistema y establecer líneas de mejora futura.

4. Marco teórico

4.1 Glosas en LSC

Una *glosa* corresponde a la representación escrita de una seña. Cada glosa describe un movimiento o gesto específico.

4.2 Visión por computador aplicada a secuencias

En lugar de analizar imágenes estáticas, los modelos deben interpretar secuencias de movimiento. Esto requiere técnicas que mantengan la información temporal.

4.3 MediaPipe Holistic

Framework que permite extraer landmarks corporales y de manos con alta velocidad, ideal para proyectos educativos o prototipos.

4.4 Redes Recurrentes (RNN / LSTM / GRU)

Son modelos diseñados para trabajar con datos secuenciales. Captan dependencias temporales entre frames consecutivos del video, permitiendo clasificar patrones de movimiento.

5. Metodología

La metodología se divide en cuatro etapas principales:

5.1 Dataset

Se utilizó el dataset **DataSet_LSC**, compuesto por videos cortos (~2 segundos) correspondientes a distintas glosas.
Se trabajó con un subconjunto equilibrado de clases, dado el alcance del proyecto.

Cada video se almacenó con su glosa correspondiente.

5.2 Preprocesamiento de videos

Para preparar los videos se realizaron los siguientes pasos:

1. **Extracción de frames del video** (16 frames por video).
2. **Aplicación de MediaPipe Holistic**, obteniendo:
    - 33 landmarks de cuerpo (pose)
    - 21 landmarks de mano derecha
    - 21 landmarks de mano izquierda
3. **Transformación a vector numérico** de tamaño fijo (150 características por frame).
4. **Normalización y padding** para garantizar que cada video tenga exactamente 16 frames.

El resultado final para cada video es un tensor:

```
(16 frames, 150 características)
```

5.3 Modelo de reconocimiento

El modelo entrenado está basado en una **Red Neuronal Recurrente GRU**, con las siguientes características:

- Entrada: secuencias de `(16, 150)`
- Una capa GRU
- Capa densa final para clasificación
- Función de pérdida: Categorical Crossentropy
- Optimizador: Adam
- Entrenamiento por 20 épocas

El entrenamiento se realizó en Google Colab.

5.4 Interfaz web en Streamlit

Se desarrolló una interfaz web que permite:

- Subir un video (MP4, MOV, AVI, MKV)
- Reproducirlo en pantalla
- Procesarlo automáticamente
- Obtener la glosa predicha y su nivel de confianza

Incluye un mensaje especial cuando la confianza del modelo es baja.

6. Resultados

Durante el proceso de validación se observó:

- El modelo logra reconocer patrones básicos en las glosas.
- La confianza promedio en videos controlados es moderada.
- En videos reales grabados con diferentes ángulos o iluminaciones, la confianza puede disminuir.
- La interfaz funciona correctamente y permite una interacción sencilla.

Ejemplo de salida ante baja confianza:

> *“No estoy seguro de la glosa. Mejor candidato: año (confianza: 0.09).*
> *El modelo necesita más entrenamiento o un video más claro.”*

Esto demuestra que el sistema es transparente respecto a su nivel de certeza.

7. Interfaz del sistema

La interfaz desarrollada en Streamlit permite:

1. Cargar un video directamente desde el computador.
2. Ver la vista previa del video.
3. Procesar el video y obtener una predicción.
4. Mostrar la glosa con su probabilidad.
5. Advertir al usuario cuando la confianza es baja.

Este diseño sencillo facilita su uso en entornos educativos y de pruebas.

8. Discusión

Fortalezas

- Arquitectura ligera, rápida y fácil de ejecutar.
- Extracción automática de características significativas usando MediaPipe.
- Predicción en tiempo real mediante Streamlit.
- Código modular y organizado para futuras mejoras.

Limitaciones

- Cantidad limitada de videos por glosa.
- Variabilidad entre grabaciones afecta el rendimiento.
- El modelo puede fallar cuando el usuario no está bien centrado en cámara.
- La confianza es baja en algunos videos reales.

Causas de los errores observados

- El dataset es relativamente pequeño.
- Algunos videos no tienen condiciones óptimas de iluminación.
- Las diferencias entre la grabación del dataset y tus videos reales afectan la generalización.

9. Trabajo futuro

Aunque el prototipo funciona, su desempeño puede mejorar significativamente con:

1. **Aumentar el dataset** con más ejemplos por glosa.
2. **Mejor iluminación y calidad en los videos de entrenamiento**.
3. **Incorporar modelos de arquitectura más robusta**, como:
    - CNN + LSTM
    - Transformers para video

4. **Entrenamiento incremental** con ejemplos nuevos de cada usuario.
5. **Normalizar fondo, distancia y postura** antes del reconocimiento.
6. Crear una **API** para conectar la IA a aplicaciones móviles o plataformas educativas.

Estas mejoras garantizan que, a futuro, el sistema logre reconocer glosas de forma más precisa y estable.

10. Conclusiones

El proyecto logró desarrollar un sistema funcional capaz de:

- Procesar videos cortos de la Lengua de Señas Colombiana.
- Extraer características corporales y de manos mediante MediaPipe.
- Aplicar una red neuronal recurrente para predecir glosas.
- Mostrar los resultados de manera clara mediante una interfaz web.

Aunque el modelo todavía presenta limitaciones y baja confianza en algunos casos, se establecieron bases sólidas para continuar entrenándolo y mejorarlo progresivamente. El prototipo demuestra la viabilidad del reconocimiento automático de señas en LSC usando herramientas accesibles y eficientes.

Este trabajo constituye un primer paso hacia la creación de herramientas tecnológicas que apoyen la enseñanza, el aprendizaje y la accesibilidad comunicativa para la comunidad sorda en Colombia.