Sistema de procesamiento digital de fotogramas para detección de palabras mediante reconocimiento de movimientos labiales

Trabajo Terminal No. _ _ _ _-_ _ _ _

*Alumnos: Gómez Mérida Brandon, *Hernández Ceciliano Luis Ángel*

Directores: Kolesnikova Olga, Serrano Talamantes José Félix

**email: lhernandezc1304@alumno.ipn.mx*

Resumen:

La situación actual en la que vivimos nos hace sentir inseguros en la mayoría de los lugares que frecuentamos, con este proyecto pretendemos brindar un elemento de seguridad en al menos uno de esos lugares, los cajeros automáticos. Intentaremos lograr esto creando un sistema que sea capaz de imitar la habilidad humana de interpretar el habla leyendo los labios del emisor, haciendo uso de cámaras a las que un usuario pueda hablar sin necesidad de emitir sonidos y realizando reconocimiento de patrones, sobre los labios de los usuarios, emitir mensajes de alertas a administradores del sistema.

Palabras clave: Lectura de labios, Machine learning, Reconocimiento de patrones, Sistema

1. Introducción

La percepción del habla es el proceso mediante el cual se extrae información de las señales visuales y/o auditivas de un lenguaje, permitiendo entenderlas e interpretarlas [1]. Si bien la percepción del habla suele ser relacionada principalmente con el campo de la fonología no es una habilidad meramente auditiva, sino que es multimodal, pues, en la gran mayoría de los casos se requiere que el hablante realice movimientos específicos de sus dientes, labios y lengua para producir los sonidos característicos de un lenguaje. La información proporcionada por los labios y el rostro del hablante complementan la comprensión auditiva [2].

La lectura de labios es una forma de comprensión del lenguaje en la que una persona intenta interpretar el habla mediante el movimiento de los labios del emisor. Esta técnica es utilizada mayormente con persona con discapacidad auditiva. En países como España y Estados Unidos de América se ha hecho o se hace uso de técnicas de lectura de labios durante procesos judiciales principalmente con el objetivo de extraer conversaciones que quedan grabadas en videos sin sonido [3] [4].

A continuación, se muestran diversos sistemas que se han desarrollado para realizar lectura de labios por medio de videos.

Nombre	Características					Precio (MXN)
	Tecnología utilizada	Idioma	Reconocimiento de imágenes	Manejo de Lenguaje Natural	Eficacia	
Text Extraction through Video Lip Reading Using Deep Learning [5]	Deep Learning	Inglés	✓		85%	No tiene (Investigación)
LipNet (DeepMind) [6]	Deep Neural Network	Inglés	✓		95.2%	No tiene (Investigación)

Automatic Lip-Reading System Based on Deep Convolutional Neural Network and Attention-Based Long Short-Term Memory [7]	Red neuronal híbrida de CNN y LSTM	Inglés	✓		88.2%	No tiene (Investigación)
Lectura de labios mediante técnicas de Machine Learning [8]	Machine Learning	Español	✓	✓	82.6%	No tiene (Tesis Máster)

Tabla 1. Resumen de productos similares.

Es importante destacar que todos los sistemas listados anteriormente fueron creados con fines de investigación y no tienen una aplicación práctica definida.

2. Objetivo

Construir un sistema que permita la detección de palabras clave que indiquen que un usuario frente a una cámara se encuentra en peligro, permitiendo enviar alertas de seguridad a un administrador.

Objetivos específicos

1. Reconocer al menos 15 de los 22 fonemas presentes en el idioma español mediante los movimientos labiales que los realizan
2. Obtención de los fotogramas capturados en vivo desde una cámara con una tasa de 30 cuadros por segundo mediante la implementación de una red neuronal convolucional
3. Implementación de un método de extracción de características basado en reconocimiento de patrones para detección y clasificación de movimientos labiales
4. Detectar un mínimo de 20 palabras de auxilio
5. Generar una alerta en menos de dos minutos a partir de que un usuario pide ayuda frente a una cámara

3. Justificación

En la Ciudad de México actualmente operan 8,516 cajeros automáticos [9], los cuales cuentan con al menos una cámara de vigilancia enfocada hacia el rostro del usuario [10]. De acuerdo con una encuesta nacional de seguridad pública realizada por el INEGI en 2019 el 82.1% de los entrevistados dijeron sentirse inseguros utilizando cajeros automáticos localizados en la vía pública [11].

De lo mencionado en el párrafo anterior se destaca la inseguridad que sienten los usuarios al usar cajeros automáticos, pues no siempre cuentan con vigilancia. Lo anterior puede ser solucionado implementando una herramienta que los usuarios puedan utilizar para solicitar ayuda si se encuentran en peligro o bajo alguna amenaza.

No es difícil reconocer cuando alguien se encuentra en una situación de peligro, si se pudiera tener personas revisando todas y cada una de las cámaras de seguridad de cajeros automáticos sería sencillo reducir los índices delictivos con los que suceden. Por evidentes razones esto no es factible, costaría demasiado dinero y tiempo de entrenamiento tener a personas revisando cámaras de seguridad, es por esta razón que se propone como solución un sistema que sea capaz de realizar lectura de labios mediante video sin sonido haciendo uso de machine learning mediante las técnicas de reconocimiento de patrones y extracción de características. Se pretende utilizar una red neuronal convolucional supervisada con memoria larga de corto plazo basada en la atención (CNN - LSTM) de manera que nos permita realizar la detección de palabras utilizando una serie de fotogramas consecutivos y tener así una predicción de las palabras presentes en los fotogramas. [12]

Se presenta esta solución como una alternativa a sistemas de reconocimiento de voz, que presentan el problema de no poder trabajar correctamente si hay un elevado índice de ruido de fondo o cuando no existe sonido alguno [13]. También es una alternativa sobre el enseñar a una persona a leer labios ya que es difícil medir su progreso en esta habilidad [14] por lo que resulta poco eficiente buscar entrenar personas para realizar lectura de labios.

Si bien esta solución no puede abordar todas las situaciones de riesgo en que los usuarios de cajeros automáticos se puedan encontrar, ayudaría con varias de ellas. Para este proyecto nos enfocaremos en situaciones de secuestro exprés y asaltos, que son situaciones en las que el usuario de un cajero automático es forzado a retirar dinero de su cuenta [15], se tiene planeado para este sistema detecte señales vocales proporcionadas por el usuario del cajero automático que refieran a palabras de ayuda o que brinden información sobre que está siendo víctima de una actividad delictiva.

Existen diversas limitaciones asociadas con la lectura de labios, algunas de ellas son que la velocidad con la que se habla un lenguaje es muy rápida como para realizar la lectura labial fácilmente, otra dificultad es que diversos movimientos labiales representan más de un sonido lo que puede causar confusión a una persona que esté tratando de leer labios [16]. La segunda limitación presentada puede ser solucionada implementando herramientas de procesamiento de lenguaje natural.

4. Productos o resultados esperados

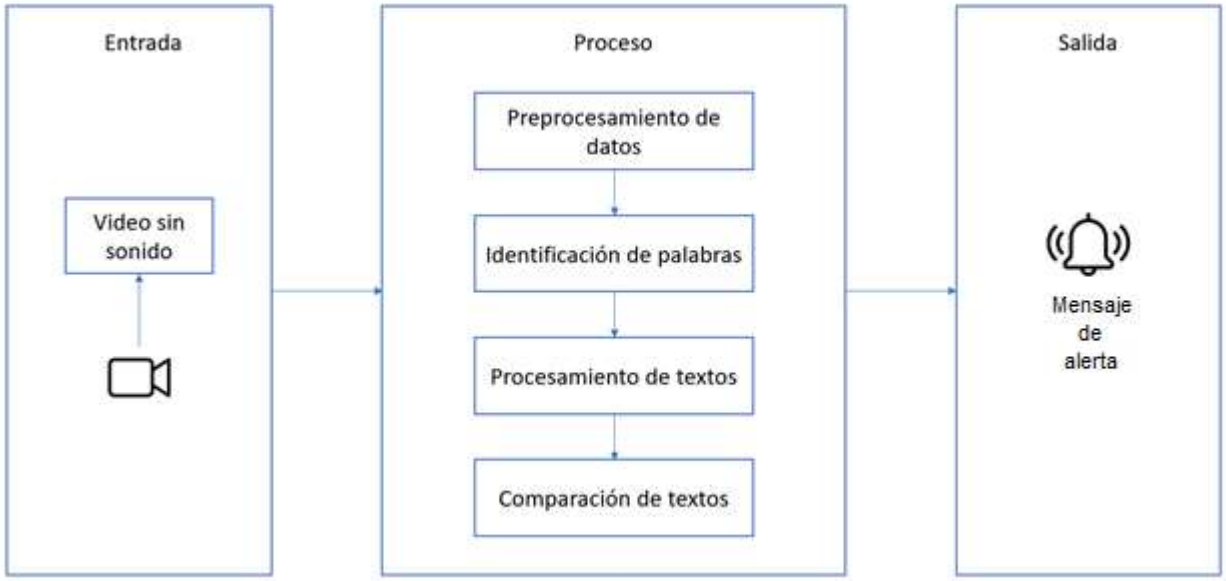

Figura 1. Arquitectura del sistema

- Manual de usuario para el sistema.
- Documento técnico.
- Sistema desarrollado.

5. Metodología

Se propone utilizar el modelo incremental como metodología para el desarrollo de este sistema, pues consideramos que se adapta de manera correcta al proyecto debido a que permite realizar cambios a cada parte del sistema, pues, los sistemas que utilizan machine learning y procesamiento de video sin audio suelen ser complejos y necesitar de ajustes en cada momento del desarrollo con tal de llegar al resultado esperado [5].

Iteración Inicial

● Plática con directores

● Planeación inicial
● Delimitación de los requerimientos del sistema
● Planificación de actividades

1° Iteración (Entrada)
● Prueba de conexiones de cámaras
● Investigación

2° Iteración (Desarrollo del bloque de preprocesamiento de datos)
● Investigación
● Análisis
● Diseño
● Desarrollo del bloque de preprocesamiento de datos
● Pruebas

3° Iteración (Desarrollo del bloque de identificación de palabras)
● Investigación
● Análisis
● Diseño
● Desarrollo del bloque de identificación de palabras desde videos
● Pruebas

4° Iteración (Desarrollo del bloque de procesamiento de textos)
● Investigación
● Análisis
● Diseño
● Desarrollo de analizador léxico-semántico para textos obtenidos
● Pruebas

5° Iteración (Desarrollo del bloque de comparación de textos)
● Análisis
● Diseño
● Desarrollo del bloque de comparación de textos obtenidos con diccionario de riesgo
● Pruebas

6° Iteración (Desarrollo del bloque de generación de alertas)
● Análisis
● Diseño
● Desarrollo del bloque de notificaciones
● Pruebas

7° Iteración (Integración de los bloques)
● Desarrollo y establecimiento de comunicación entre los bloques desarrollados
● Pruebas finales del sistema
● Entrega final del sistema

6. Cronograma

Cronograma del alumno Brandon Gómez Mérida

Actividad	AGO	SEP	OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN
Iteración inicial											
Plática con directores											
Planeación inicial											
Delimitación de los requerimientos del sistema											
Planificación de actividades											
1° Iteración (Entrada)											
Investigacion de fonemas con respecto a las caracteristicas de los labios											
2° Iteración (Desarrollo del bloque de preprocesamiento de datos)											
Investigacion de Conjunto de árboles de regresión.											
Obtencion de dataset de rostros											
Desarrollo del bloque de preprocesamiento de datos											
3° Iteración (Desarrollo del bloque de identificación de palabras)											
Investigacion de palabras de auxilio.											
Investigacion de una red neuronal convolucional											
Investigacion de LSTM											
Desarrollo del bloque											
Pruebas											
Evaluación de TT1											
4° Iteración (Desarrollo del bloque de procesamiento de textos)											
Investigación del método de análisis semántico											
Diseño											
Desarrollo de analizador léxico para textos obtenidos											
Pruebas											
5° Iteración (Desarrollo del bloque de comparación de textos)											
Análisis											
Diseño											
Desarrollo del bloque de comparación de textos obtenidos con diccionario de riesgo											
6° Iteración (Desarrollo del bloque de generación de alertas)											
Análisis											
Diseño											
Desarrollo del bloque de notificaciones											
Pruebas											
7° Iteración (Integración de los bloques)											
Desarrollo y establecimiento de comunicación entre los bloques desarrollados											
Pruebas finales del sistema											
Entrega final del sistema											
Evaluación de TT2											

Cronograma del alumno Luis Ángel Hernández Ceciliano

Actividad	AGO	SEP	OCT	NOV	DIC	ENE	FEB	MAR	ABR	MAY	JUN
Iteración inicial											
Plática con directores											
Planeación inicial											
Delimitación de los requerimientos del sistema											
Planificación de actividades											
1° Iteración (Entrada)											
Prueba de conexiones de cámaras											
2° Iteración (Desarrollo del bloque de preprocesamiento de datos)											
Análisis											
Diseño											
Desarrollo del bloque de preprocesamiento de datos											
Pruebas											
3° Iteración (Desarrollo del bloque de identificación de palabras)											
Análisis											
Diseño											
Desarrollo del bloque											
Pruebas											
Evaluación de TT1											
4° Iteración (Desarrollo del bloque de procesamiento de textos)											
Análisis											
Diseño											
Desarrollo de analizador léxico para textos obtenidos											
Pruebas											
5° Iteración (Desarrollo del bloque de comparación de textos)											
Análisis											
Diseño											
Desarrollo del bloque de comparación de textos obtenidos con diccionario de riesgo											
Pruebas											
6° Iteración (Desarrollo del bloque de generación de alertas)											
Análisis											
Diseño											
Desarrollo del bloque de notificaciones											
Pruebas											
7° Iteración (Integración de los bloques)											
Desarrollo y establecimiento de comunicación entre los bloques desarrollados											
Pruebas finales del sistema											
Entrega final del sistema											
Evaluación de TT2											

7. Referencias

[1] Vniversitat de València, «Percepción del lenguaje,» [En línea]. Available: https://www.uv.es/gotor/Transparencias/tr4PERCEPC.HABLA.pdf. [Último acceso: 15 Marzo 2021].

[2] N. P. Erber, «Interaction of Audition and Vision in the Recognition of Oral Speech Stimuli,» *Journal of Speech and Hearing Research,* vol. 12, nº 2, pp. 423-425, 1969.

[3] M. B. Rey, «La lectura labio-facial (llf) en la investigación de procesos judiciales,» *Tonos digital: Revista de estudios filológicos,* nº 30, pp. 1-15, 2016.

[4] L. D. Rosenblum, «Lipreading for the FBI,» Psychology Today, 16 Marzo 2010. [En línea]. Available: https://www.psychologytoday.com/us/blog/sensory-superpowers/201003/lipreading-the-fbi. [Último acceso: 20 Marzo 2021].

[5] S. M. Mazharul Hoque Chowdhury, M. T. O. M. Rahman y A. Hasan, «Text Extraction through Video Lip Reading,» de *8th International Conference System Modeling and Advancement in Research Trends (SMART)*, Moradabad, 2019.

[6] Y. M. Assael, B. Shillingford, S. Whiteson y N. de Freitas, «LipNet: End-to-End Sentence-level Lipreading,» *arXiv:1611.01599 [cs.LG],* 2016.

[7] Y. Lu y H. Li, «Automatic Lip-Reading System Based on Deep Convolutional Neural Network and Attention-Based Long Short-Term Memory,» *Applied Sciences,* vol. 9, nº 8, pp. 1599-1611, 2019.

[8] D. Gimeno Gómez, «Lectura de labios mediante técnicas de Machine Learning,» *Universitat Politècnica de València,* 2020.

[9] Sistema de Información Económica, «Número de cajeros automáticos por entidad federativa - (CF266),» Banco de México, 2020. [En línea]. Available: https://www.banxico.org.mx/SieInternet/consultarDirectorioInternetAction.do?sector=5&accion=consultarCuadro&idCuadro=CF266&locale=es. [Último acceso: 19 Marzo 2020].

[10] M. A. Villegas Toinga, «Sistema de comunicaciones y monitoreo de un cajero automático de la cooperativa de ahorro y crédito policía nacional a ubicarse en el cantón chimbo – provincia de bolívar,» *Universidad Técnica de Ambato,* p. 143, 2018.

[11] Instituto Nacional de Estadística y Geografía, «ENCUESTA NACIONAL DE SEGURIDAD PÚBLICA URBANA,» Aguascalientes, 2019.

[12] S. Saha, «A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way,» towards data science, 15 Diciembre 2018. [En línea]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53. [Último acceso: 11 Agosto 2021].

[13] L. El Asri, «Cuidado con lo que dices: las máquinas están aprendiendo a leerte los labios,» El Diario, 14 Octure 2014. [En línea]. Available: https://www.eldiario.es/hojaderouter/tecnologia/ordenadores-leer-labios-vigilancia_1_4589135.html. [Último acceso: 16 Marzo 2021].

[14] N. A. Altieri, «Some normative data on lip-reading skills,» *Journal of the Acoustical Society of America,* vol. 130, nº 1, pp. 1-4, 2011.

[15] Universidad Nacional Autónoma de México, «Secuestros en México,» Universidad Nacional Autónoma de México, 18 Marzo 2021. [En línea]. Available: https://www.unam.mx/medidas-de-emergencia/secuestros-en-mexico. [Último acceso: 19 Marzo 2021].

[16] BetterHealth Channel, «Hearing loss - lipreading,» Department of Health, State Government of Victoria, Australia, 11 Abril 2017. [En línea]. Available: https://www.betterhealth.vic.gov.au/health/ConditionsAndTreatments/hearing-loss-lipreading#limitations-of-lipreading. [Último acceso: 19 Marzo 2021].

[17] S. Sehlhorst, «Foundation Series: Software Process (Waterfall Process versus Incremental Process),» 3 Enero 2006. [En línea]. Available: http://tynerblain.com/blog/2006/01/03/foundation-series-software-process-waterfall-process-versus-incremental-process/. [Último acceso: 16 Marzo 2020].

[18] M. G. Piattini Velthuis, Calidad de Sistemas de Información. 4ª edición, Madrid: Ra-Ma, 2018.

[19] R. S. Pressman, Ingeniería del Software: Un Enfoque Práctico, McGraw-Hill, 2006.

[20] J. S. Chong, A. Senior, O. Vinyals y A. Zisserman, «Lip Reading Sentences in the Wild,» 2017.

8. Alumnos y directores

CARÁCTER: Confidencial
FUNDAMENTO LEGAL: Artículo 11 Fracc. V y Artículos 108, 113 y 117 de la Ley Federal de Transparencia y Acceso a la Información Pública.
PARTES CONFIDENCIALES: Número de boleta y teléfono.

Brandon Gómez Mérida. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2013021330, Tel. 5579776587, email: bgomezm1500@alumno.ipn.mx
Firma: _____

Luis Ángel Hernández Ceciliano. - Alumno de la carrera de Ing. en Sistemas Computacionales en ESCOM, Especialidad Sistemas, Boleta: 2014090311, Tel. 5540966561, email: lhernandezc1304@alumno.ipn.mx
Firma: _____

Olga Kolesnikova. - Profesora Titular de ESCOM, miembro del Sistema Nacional de Investigadores, miembro de la Sociedad Mexicana de Inteligencia Artificial. Áreas de investigación: Procesamiento de lenguaje natural, Aprendizaje automático, Redes neuronales artificiales, email: kolesolga@gmail.com
Firma: _____

José Félix Serrano Talamantes, Dr. en C. de la Computación graduado en CIC-IPN en 2011, Adscrito a CIDETEC/IPN, Profesor de ESCOM/IPN impartiendo Image Analysis en el ciclo 2021-2, Áreas de Interés: Visión por computadora, Cómputo Inteligente, Machine Learning, Deep Learning, Inteligencia Artificial, Tel: 57296000, Ext 52533, email: jfserranotal@gmail.com
Firma: _____