

SISTEMA DE DETECCIÓN DE EMOCIONES A TRAVÉS DE EXPRESIONES FACIALES

EMOTION DETECTION SYSTEM THROUGH FACIAL EXPRESSIONS

Enoc Tapia Méndez

Universidad Autónoma de Querétaro, México
etapia17@alumnos.uaq.mx

Álvaro Sánchez Rodríguez

Tecnológico Nacional de México / IT de Celaya, México
alvaro.sanchez@itcelaya.edu.mx

Luis Ballesteros Martínez

Mabe TyP, México
luis.ballesteros@mabe.com.mx

Recepción: 28/septiembre/2021

Aceptación: 30/noviembre/2021

Resumen

En la actualidad la inteligencia artificial en colaboración con nuevas tecnologías contribuye a la mejora de procesos y el desarrollo de innovaciones.

En el presente artículo, se presenta la implementación de un algoritmo programado en lenguaje Python capaz de detectar las 7 emociones humanas básicas a través de expresiones faciales. Además, dicho sistema fue implementado en una serie de experimentos en los cuales se evaluó un nuevo producto que saldría al mercado. Como resultado final, se obtuvo un sistema con una precisión del 95% en la detección de emociones.

Palabras Clave: Emociones, expresiones faciales, visión artificial.

Abstract

At present, artificial intelligence in collaboration with new technologies contributes to the improvement of processes and the development of innovations.

In this article, the implementation of an algorithm programmed in Python language it is capable to detect the 7 basic emotions through facial expressions. In addition, this

system was implemented in a series of experiments in which a new product that would be released on the market was evaluated.

As a final result, a system with a 95% accuracy in emotion detection was obtained.

Keywords: *Artificial vision, emotions, facial expressions.*

1. Introducción

El reconocimiento facial a través de la expresión emocional, puede definirse como la capacidad de todos los individuos de reconocer formas básicas de expresión afectiva, la cual aparece en los rostros de las personas [Ceebrián, 2016].

Las emociones son una parte esencial de quiénes somos y cómo nos comportamos. Son vitales para que podamos funcionar como seres humanos que toman decisiones racionales. La principal preocupación de la psicoterapia es la reparación de los trastornos emocionales. Dicha terapia aspira a que los pacientes tomen conciencia de sus emociones haciendo entablar una conversación con el terapeuta, además de los estados de pensamiento interno y autorreflexión. Esta forma considera que la interacción anima a los pacientes a reconocer patrones de comportamiento y tratar mejor o ayudar a resolver sus propios problemas [C. Heth, 2007].

Según, Paul Ekman, autor en que se basó el siguiente proyecto para la detección de emociones, describe que existen 7 emociones básicas en los humanos que son: Felicidad, sorpresa, ira, neutral, tristeza, miedo y desprecio [Vázquez, 2009].

Para realizar un estudio de mercado y analizar si un producto es innovador, es necesario obtener datos valiosos con respecto a los nuevos productos, por lo que se llevan a cabo distintas etapas, como son la demostración del producto, percepciones de este, análisis y llenado de encuesta de satisfacción.

El proceso involucra mucho tiempo y que los resultados obtenidos no sean los verídicos, debido a que el usuario no responde fielmente con respecto a su percepción personal.

Con la implementación del sistema inteligente, se solucionan los problemas descritos anteriormente y se agiliza todo el proceso para la obtención de datos.

2. Métodos

Para llegar a la solución del problema, que se simplifica en la necesidad de evaluación de nuevos productos de lanzamiento al mercado, se siguió la metodología del modelo de Archer del proceso de diseño, que se muestra en la figura 1.

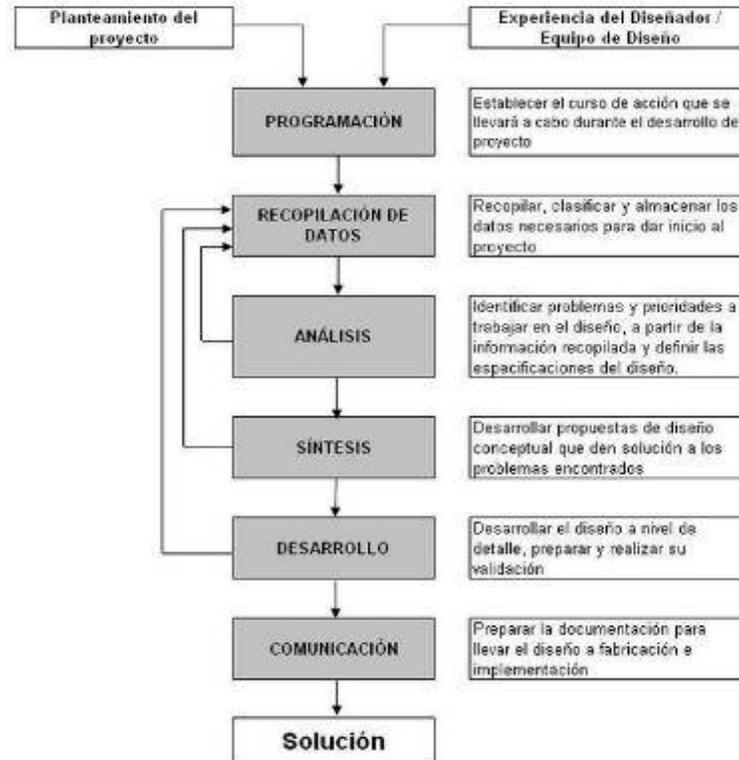


Figura 1 Metodología de diseño de Archer.

Las expresiones faciales se refieren a los movimientos musculares producidos en la cara; tales movimientos son indicativos de emociones en personas. Paul Ekman, un psicólogo que fue pionero en investigaciones de reconocimiento emocional, inicialmente determinando que hay 6 emociones básicas, que él llamó emociones universales y tienen una expresión facial característica, como se muestra en la figura 2 [Ekman, 1981]. La arquitectura de visión artificial se compone por un principio de funcionamiento y un diagrama de bloques del sistema, en los cuales se describe el funcionamiento de la tecnología. Dicho proceso puede observarse en la figura 3.



Figura 2 Expresiones faciales de las emociones.

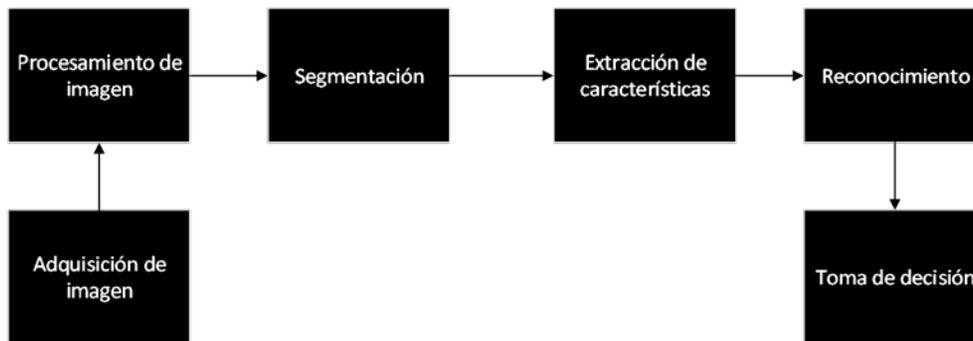


Figura 3 Principio de funcionamiento de visión artificial.

Una red neuronal artificial es una distribución masiva de un procesador paralelo que tiene una facilidad natural para almacenar el conocimiento obtenido de la experiencia para que pueda ser utilizado eventualmente. Las redes neuronales artificiales son similares al cerebro humano en dos aspectos:

- El conocimiento se obtiene a través de un proceso de aprendizaje.
- Las conexiones entre las neuronas se conocen como pesos sinápticos y se utilizan para almacenar el conocimiento [Vallez, 2007].

Para el reconocimiento de imágenes, la conectividad total en las redes perceptrón multicapa son una limitación cuando el tamaño de las imágenes aumenta. Esta red funciona para imágenes pequeñas, pero a medida que aumenta la imagen, el número de parámetros para el entrenamiento aumenta excesivamente haciendo que el proceso de aprendizaje sea lento.

La convolución es una operación del producto entre la imagen de entrada y kernel o filtro, generando un mapa de características correspondiente a cada posible ubicación de la imagen original, por lo que el número de conexiones y el número de parámetros para el entrenamiento son reducido. Si una imagen es de $n \times m$, existen k filtros de $a \times b$ en una capa de convolución, entonces se generan k mapas de características de $k(n - a + 1)(m - b + 1)$ tamaño.

La red está estructurada con tres capas de convolución. Cada capa tiene un tamaño de 64, con un núcleo de 3×3 y ReLu como función de activación, se agrega un abandono al final de la primera capa, que consiste en desconectar un porcentaje de neuronas en las iteraciones de entrenamiento y una función soft-max en la capa totalmente conectada está implementado.

Después de cada capa de convolución, se aplica una capa de agrupación, utilizando la operación de agrupación máxima para la primera capa y la operación de agrupación promedio para las capas restantes. La configuración de la red neuronal se muestra en la tabla 1.

Tabla 1 Resumen de la configuración de la red neuronal.

Capa	Salida
Conv2D	(None, 46, 46, 32)
MaxPooling2D	(None, 22, 22, 32)
Dropout	(None, 22, 22, 32)
Conv2D	(None, 20, 20, 64)
Conv2D	(None, 18, 18, 64)
AveragePooling2D	(None, 8, 8, 64)
Conv2D	(None, 6, 6, 128)
Conv2D	(None, 4, 4, 128)
AveragePooling2D	(None, 1, 1, 128)
Flatten	(None, 128)
Dense	(None, 1024)
Dropout	(None, 1024)
Dense	(None, 1024)
Dropout	(None, 1024)
Dense	(None, 7)

Antes de entrenar a la red, se establece un conjunto de datos de aprendizaje profundo generado, para cada expresión facial (felicidad, tristeza, ira, sorpresa, miedo y neutral), usando la imagen de Google motor de búsqueda y la opción de filtrado de imágenes con derecho a usar, compartir o modificar libremente,

obteniendo imágenes representativas de expresiones faciales. Se realiza una nueva muestra de la imagen, dejando solo la cara y ajustando el tamaño para que sea de 48 x 48 píxeles; una conversión del espacio de color RGB a escala de grises con un rango de valores de 0 a 255 se hace. El hardware utilizado para esta etapa es una computadora portátil con un procesador Intel Core i3 con 2 núcleos, una red estructurada en código de secuencia de comandos de Python, utilizando la biblioteca Tensorflow que puede manejar redes neuronales convolucionales, además de ser una biblioteca de red abierta de software de fuente para cómputo numérico utilizando gráficos de flujo de datos, con una arquitectura flexible que permite su implementación en una variedad de plataformas (CPU, GPU, TPU), servidores, que también incluyen plataformas móviles e integradas. Otras bibliotecas como keras, numpy y matplotlib se utilizan para manejo de datos.

3. Resultados

Los resultados se dividen en dos secciones, donde la primera son pruebas realizadas con imágenes buscadas en la red y mostrándolas al sistema para conocer como respondía a las mismas. En figura 3 se puede observar como el sistema responde ante las mismas.

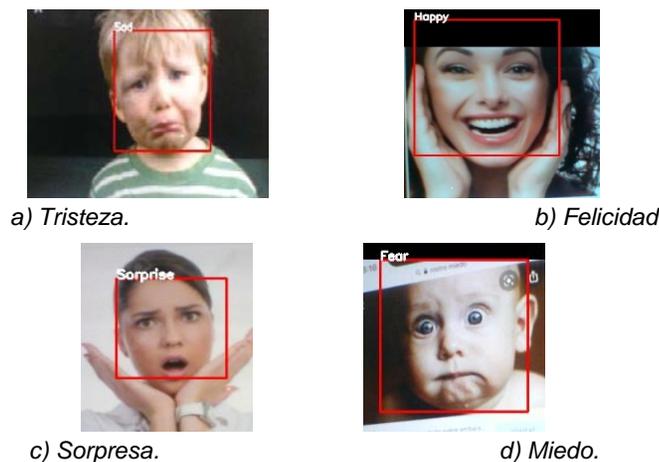


Figura 3 Expresiones faciales.

En la figura 3a puede identificarse la emoción “tristeza” y arroja un resultado positivo, ya que la imagen buscada sí representa dicha emoción. La siguiente

emoción para probar con el sistema fue la emoción felicidad, como se puede observar en la figura 3b, con lo cual, se obtiene otro resultado favorable al reconocer la emoción buscada. Sucesivamente, se realizaron dos pruebas extras para probar la efectividad del sistema y las emociones detectadas fueron sorpresa y miedo, arrojando datos favorables y que el sistema funciona correctamente. Dichas pruebas se pueden observar en las figuras 3c y 3d.

Como segunda parte de los resultados, se elaboró una prueba en conjunto con el centro de investigaciones Mabe TyP, en la que se evaluarían una serie de aspectos entre usuarios que contaban con un sistema de purificación y quienes utilizaban un método de abastecimiento tradicional con garrafón. Cabe señalar que el objetivo de la prueba fue obtener datos sobre el producto que se lanzaría al mercado. Las variables que se evaluaron fueron las siguientes:

- Atracción
- Credibilidad
- Innovación
- Relevancia
- Valor
- Originalidad
- Calidad
- Intención de compra

Las pruebas se realizaron de forma remota, en las que participaron el entrevistador y un usuario elegido al azar. Mientras se realizaba la prueba, el sistema estaba en ejecución y después de generar una pregunta al usuario con la que se evaluaría cada factor, se contabilizaban el número de veces que se producía la emoción en el usuario. El total de las pruebas realizadas fue de cuatro y en la tabla 2 se muestran los resultados. Analizando los datos, se puede asumir que para los usuarios 1 y 2, que ya cuentan con un sistema de purificación, no les genera una emoción positiva el nuevo producto, mientras que a los usuarios 3 y 4, que utilizan un método de abastecimiento tradicional, les genera un tipo de emociones positivas el nuevo producto que es un sistema de purificación de agua.

Tabla 2 Resultados de pruebas a usuarios.

Emociones	Sistemas de purificación		Garrafón	
	Usuario 1	Usuario 2	Usuario 3	Usuario 4
ATRACCIÓN	Negativo	Neutral	Positivo	Neutral
CREDIBILIDAD	Negativo	Neutral	Neutral	Positivo
INNOVACIÓN	Negativo	Negativo	Positivo	Positivo
RELEVANCIA	Negativo	Negativo	Neutral	Neutral
VALOR	Negativo	Negativo	Negativo	Positivo
ORIGINALIDAD	Positivo	Neutral	Negativo	Positivo
CALIDAD	Negativo	Neutral	Neutral	Neutral
INTENCIÓN DE COMPRA	Negativo	Neutral	Positivo	Positivo

Este tipo de ejercicios pueden ser el parteaguas para una nueva investigación en la que se mejore la base de datos con la que se alimentó la red neuronal y además, construir modelos con los cuales se puedan agrupar o clasificar las emociones.

4. Discusión

Al analizar los datos obtenidos durante los distintos experimentos, se observa que el sistema puede predecir una emoción al igual que un humano la percibe, aunque no es tan rápida la respuesta a la que podría presentar un humano.

Por otro lado, el sistema tiene un error del 5%, lo que puede provocar que no se elija la emoción correspondiente en el lapso simultáneo de tiempo y la predicción no se haga de la forma correcta. Como futuro trabajo, se puede mejorar la precisión del sistema cambiando la configuración de la red neuronal convolucional o intentando el entranamiento con alguna red ya establecida con ResNet, además de buscar un mayor número de datos de entrenamiento, dado que entre mayor sea la base de datos, la precisión aumenta.

5. Conclusiones

El sistema es entrenado con una base de datos ya existente, por lo que sería viable observar el comportamiento del sistema si es entrenado con otra base de datos distintos y en consecuencia quizás deben hacerse cambios al algoritmo para que genere una respuesta más acertada ante ciertos eventos.

Dados los resultados de la segunda prueba, se puede realizar la categorización de emociones negativas y positivas, quedando la categoría de la siguiente forma:

- Emociones negativas: Enojo, desprecio, miedo y tristeza.
- Emociones positivas: Felicidad y sorpresa.

Cabe señalar, que la emoción neutral, no se encuentra en ninguna de las categorías, debido a que probablemente no represente algún valor agregado para la prueba.

6. Bibliografía y Referencias

- [1] Universidades, P.F. (2020). Estrategias IA 2020. ThreePoints.
- [2] C. Heth. (2007). *Psychology: The Science of Behavior*.
- [3] Ekman, P. (1981). Facial expressions of emotions. *Studies in psychology*.
- [4] Vásquez, A. C., Quispe, J. P., & Huayna, A. M. (2009). Procesamiento de lenguaje natural. *Revista de investigación de Sistemas e Informática*.
- [5] Cebrián Chuliá, L. (2016). Reconocimiento de Emociones mediante técnicas de aprendizaje profundo (Doctoral dissertation).
- [6] Vallez, M., & Pedraza, R. (2007). *El Procesamiento del Lenguaje Natural en la Recuperación de Información Textual y áreas afines*. Hipertext. ((Lahoz-Beltrá, 2010).
- [7] Suárez, O. S., Riudavets, F. J. C., & Aguiar, J. R. P. (2004). Relaciones morfológicas sufijales para el procesamiento del lenguaje natural. Mileto Ediciones.
- [8] Quintero, C., Merchán, F., Cornejo, A., & Galán, J. S. (2018). *Uso de Redes Neuronales Convolucionales para el Reconocimiento Automático de Imágenes de Macroinvertebrados para el Biomonitorio Participativo*. KnE Engineering. (Quintero, 2018)
- [9] Rich, E., Knight, K., Calero, P. A. G., & Bodega, F. T. (1994). *Inteligencia artificial* (Vol. 1). McGraw-Hill.
- [10] Nilsson, N. J., Morales, R. M., Méndez, J. T. P., & Aris, E. P. (2001). *Inteligencia artificial: una nueva síntesis* (Vol. 2). España.: McGraw-Hill. (Nilsson, 2001).

- [11] Martínez Guerrero, M. (2018). Reconocimiento facial para la autenticación de usuarios (Doctoral dissertation).
- [12] Benítez, R., Escudero, G., Kanaan, S., & Rodó, D. M. (2014). Inteligencia artificial avanzada. Editorial UOC.
- [13] Díez, R. P., Gómez, A. G., & de Abajo Martínez, N. (2001). Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva. Universidad de Oviedo.
- [14] Lahoz-Beltrá, R. (2010). Bioinformática: Simulación, vida artificial e inteligencia artificial. Ediciones Díaz de Santos.
- [15] Shavar, B. A., & Atwell, E. S. (2005). Using corpora in machine-learning chatbot systems. *International journal of corpus linguistics*.
- [16] Klette, R., Schlüns, K., & Koschan, A. (1998). Computer vision: three-dimensional data from images (Vol. 31). Singapore: Springer.
- [17] Arévalo, V., González, J., & Ambrosio, G. (2004). La Librería de Visión Artificial OpenCV. Aplicación a la Docencia e Investigación. Base Informática.
- [18] Hernando Pericás, F. J. (1993). Técnicas de procesamiento y representación de la señal de voz para el reconocimiento del habla en ambientes ruidosos.
- [19] Díaz, A. B. M. (2004). Reconocimiento facial automático mediante técnicas de visión tridimensional (Doctoral dissertation, Universidad Politécnica de Madrid).
- [20] Cabello Pardos, E. (2004). Técnicas de reconocimiento facial mediante redes neuronales (Doctoral dissertation, Informática).
- [21] Alpaydin, E. (2020). Introduction to machine learning. MIT press.