

## Aplicación de regresión polinomial como método de predicción de datos en señales obtenidas a partir de movimientos de cabeza

Luis Alberto Hernández Montiel, Edmundo Bonilla Huerta,  
Edwyn Martínez Carrillo, Roberto Morales Caporal

Tecnológico Nacional de México,  
Campus Apizaco, Tlaxcala,  
México

{D23370018, edmundo.bh, M22371203,  
roberto.mc}@apizaco.tecnm.mx

**Resumen.** En este artículo, se presenta un modelo basado en regresión polinomial como método de predicción de datos, para homogeneizar las señales obtenidas de movimientos de cabeza que realiza una persona al ser estimulada para generar una emoción. Primero, se estimula a la persona proyectándole vídeos para generar la emoción de disgusto. Después, se rastrean los macro y micro movimientos que realizó el participante durante la estimulación, mediante la transformación de las muestras del dominio espacial a un dominio frecuencial utilizando análisis de vídeo y momentos de Hu. A las señales obtenidas de este proceso, se les aplica regresión polinomial como técnica de predicción de datos para homogeneizar las señales a un tamaño estándar. Los resultados que genera este algoritmo muestran una predicción más ajustada al patrón que sigue la señal generada por los movimientos corporales de una persona.

**Palabras clave:** Movimientos de cabeza, regresión polinomial, análisis de vídeo, momentos Hu.

### Application of Polynomial Regression as a Data Prediction Method in Signals Obtained from Head Movements

**Abstract.** In this article, a model based on polynomial regression as a data prediction method to homogenize the signals obtained from head movements made by a person when stimulated to generate an emotion is presented. First, the person is stimulated by projecting videos to generate the emotion of disgust. Then, the macro and micro movements performed by the participant during the stimulation are tracked by transforming the samples from the spatial domain to the frequency domain using video analysis and Hu moments. To the signals obtained from this process, polynomial regression is applied as a data prediction technique to homogenize the signals to a standard size. The results generated by this algorithm show a more accurate prediction of the pattern that follows the signal generated by the body movements of a person.

**Keywords:** Head movements, polynomial regression, video analysis, Hu moments.

## 1. Introducción

El reconocimiento de emociones es un campo en la computación que ha crecido en los últimos años. Diversas investigaciones se han enfocado en obtener una mejor comprensión de las diferentes emociones que expresa una persona durante el día. Una de las principales áreas donde se muestran diferentes expresiones es el rostro, ya que gracias a las gesticulaciones que la persona hace, se puede obtener una gran variedad de datos de diferentes emociones. Pero, las distintas partes del cuerpo pueden dar información sobre estos cambios de estados emocionales, tales como el torso, las manos, la piel (cambio de temperatura) e incluso los pies. En este trabajo, se propone el estudio de los movimientos de la cabeza de una persona como método de reconocimiento de emociones. Primero se estimula una emoción en un participante, utilizando vídeos seleccionados específicamente para la emoción de asco. Después se filman las reacciones (movimientos) de la cabeza del participante, convirtiéndolas en una señal que es reestructurada con regresión polinomial para encontrar la emoción que está presentando la persona.

## 2. Estado del arte

La necesidad de reconocer emociones de forma automática da lugar a proponer nuevos estudios utilizando estas las reacciones del cuerpo como datos para analizar una emoción y crear nuevas terapias o dispositivos para una mejor calidad de vida [1]. Pero hacer un estudio de las emociones que expresa una persona no es una tarea fácil. En muchas ocasiones, una persona reacciona con una expresión muy pronunciada y otras no realizan ninguna, además de que todas las personas se expresan de forma diferente en una misma situación. Esto implica la implementación de sistemas que utilicen más de un dispositivo para reconocer dicha emoción.

Con este enfoque, se ha propuesto la utilización de sistemas multimodales [2,3] capaces de analizar diferentes partes del cuerpo al mismo tiempo. Otra técnica propuesta es el estudio del parpadeo. En los trabajos de Maffei [4] y Demiral [5], se analiza la asociación que existe entre el parpadeo y el estado afectivo de la persona, para encontrar una emoción específica. Otros trabajos proponen la utilización de rasgos gestuales no verbales [6] para detectar emociones, entre ellos los movimientos oculares [7], movimientos de pies [8] y movimientos corporales [9].

A pesar de las diferentes técnicas propuestas, aún no se llega a una solución concreta; por lo tanto, siguen apareciendo nuevas propuestas que ayudan a reconocer emociones. En este trabajo, se propone el análisis de los movimientos de la cabeza de una persona para crear señales que puedan ayudar a reconocer una emoción. Primero, se estimula al participante utilizando vídeos específicos para la emoción de disgusto, después se rastrean esos movimientos para transformarlos en una señal. El siguiente paso es utilizar la regresión polinomial como método de predicción de datos para transformar la señal y así dejarla lista para su análisis. El método se describe a continuación.

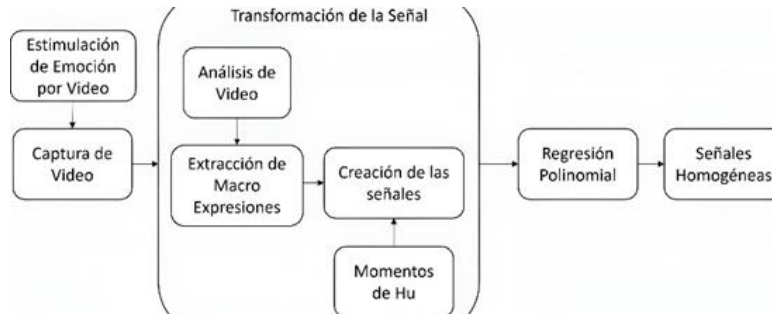


Fig. 1. Diagrama general del método propuesto.

### 3. Materiales y métodos

Encontrar patrones en un vídeo es una tarea compleja, ya que muchos de los vídeos tienen poca iluminación, ruidos ambientales o desfases de tiempo, además de que requiere un costo computacional elevado. Este problema genera la necesidad de crear algoritmos capaces de trabajar bajo estas condiciones y extraer características relevantes de una secuencia de vídeo. Para dar solución a este problema, se propone la transformación del vídeo de su forma espacial a una forma de frecuencia para formar una señal que capture las macro y micro expresiones generadas por un participante. Para ello, se propone un algoritmo combinando técnicas de visión artificial, momentos de Hu y regresión polinomial para registrar y mejorar las señales generadas por los diferentes movimientos de cabeza que realiza una persona que ha sido grabada. La figura 1 muestra el proceso general del algoritmo propuesto.

#### 3.1. Creación de la señal

Para crear una señal que capture los movimientos de cabeza que realizó el participante, se proponen 3 fases. Primero, se estimula a la persona proyectándole vídeos para generar una emoción específica. En la fase dos se hace la transformación del vídeo de un espacio visual a uno de frecuencias. En la fase tres se homogeneizan las señales aplicando regresión polinomial como técnica de predicción de datos. Con estas tres fases se pretende hacer un preprocesado para clasificar las señales. Cada una de las fases se describe a continuación.

##### 3.1.1. Estimulación de emociones y captura de video

En esta fase, se estimula al participante proyectándole vídeos para generarle la emoción de disgusto. La toma de muestras para este estudio se realizó en un área controlada en un espacio de 2x4 m, evitando corrientes de aire, ruidos ambientales y se iluminó con luz natural. La recolección de muestra se realizó entre las 9:00 y 11:00 a. m. en un lapso de dos semanas a un grupo de 24 estudiantes conformado por 13 hombres y 11 mujeres del Instituto Tecnológico de Apizaco, con un rango de edad de 18 a 27 años.

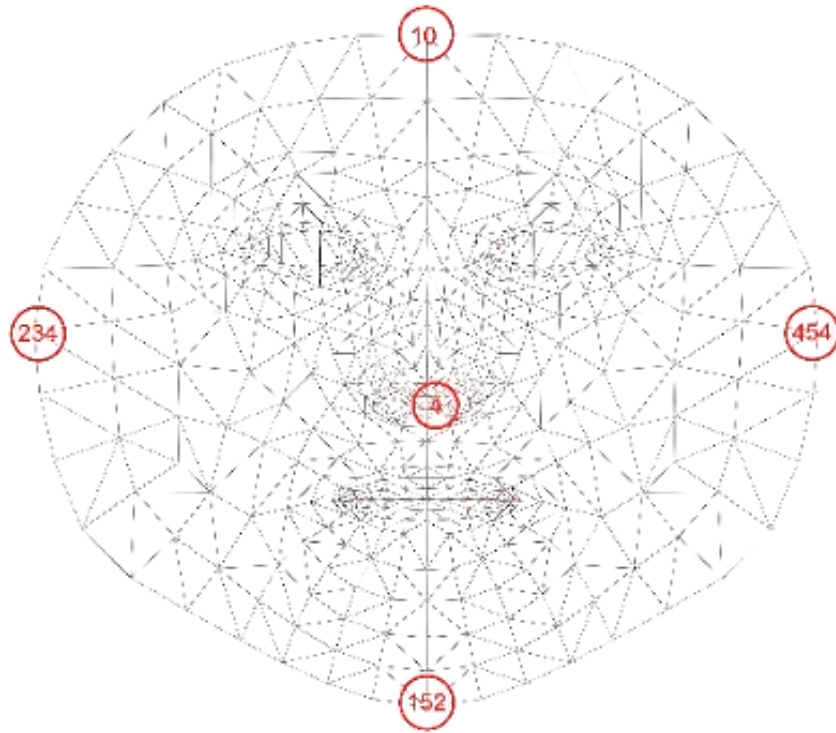


Fig. 2. Malla generada por face\_mesh de mediapipe.

Para la estimulación de la emoción, se proyectó un vídeo de los 70 que contiene la base de datos de LATEMO-E [10]. Este vídeo tiene el porcentaje más alto de vistas para la emoción de “Disgusto”, el vídeo es “Planet Terror”, con 79%. El vídeo clip tiene una duración promedio entre 1 y 2 minutos. Los criterios de selección de este vídeo fueron los siguientes: (1) la película fue sugerida con más frecuencias, (2) la coherencia del argumento de la escena con el tema que se está abordando. El material multimedia fue presentado en una laptop Lenovo IdeaPad Slim 3 con una pantalla de 15.6” y altavoces integrados. La cámara utilizada para grabar fue la cámara web integrada a la laptop con una resolución de 1080p enfocada a la parte superior del cuerpo, colocada a 1 metro de distancia del participante.

### 3.1.2. Captura del movimiento

Para transformar el vídeo a una señal, se hace lo siguiente. Primero se implementa la librería OpenCV para filmar los movimientos que hace el participante al ser estimulado. Esta librería se utiliza en aplicaciones de visión por computadora en tiempo real [11]. Con la implementación de esta librería capturamos los movimientos de cabeza del participante. Se utilizan solo los movimientos de esta parte del cuerpo porque es una de las partes que tienen más reacciones durante la estimulación.



**Fig. 3.** Puntos seleccionados con mediapipe para rastrear el movimiento.

El siguiente paso es analizar el vídeo para rastrear los movimientos de la cabeza, para ello utilizamos la librería Mediapipe [12]. Esta librería es un framework basado en aprendizaje automático para procesar datos de vídeo, dibujando una serie de puntos conectados entre sí para crear distintas soluciones como detectores de pose, manos, caras, u objetos, además de que tiene una función específica para cada parte del cuerpo lo que permite trabajar con ellas por separado. Para rastrear el movimiento de cabeza, se utiliza la función mediapipe fase\_mesh [13]. Esta librería crea una malla de puntos que cubren la cabeza (figura 2).

Para detectar el movimiento de la cabeza, se localizan cinco puntos de referencia de toda la cabeza, los cuales son la frente, la barba, los laterales cercanos a las orejas y la nariz. Con el uso de estos puntos, se logra rastrear los movimientos de la cabeza del participante. Al tener todos los puntos localizados dentro del vídeo, el siguiente paso es convertir el movimiento capturado en forma de frecuencia y crear una señal a partir de esos movimientos.

### 3.1.3. Transformación de la señal

Después de rastrear los movimientos (plano visual del vídeo), el siguiente paso es transformarlos en una señal (plano de frecuencias). Esto se realiza de la siguiente forma. Para transformar los movimientos de cabeza encontrados, lo primero que se hace es encontrar las coordenadas de cada uno de los puntos seleccionados y etiquetarlas para ver los movimientos que hace cada uno de esos puntos de interés. Después de que se obtiene el ancho y alto del fotograma del vídeo, el siguiente paso es utilizar los puntos de referencia (landmarkers) que fueron encontrados en la zona de interés de la persona. Estos landmarks se obtienen al aplicar la siguiente ecuación:

$$\begin{aligned}(x) &= LMs[i].x * w, \\(y) &= LMs[i].y * h,\end{aligned}\tag{1}$$

donde “ $x$ ”, “ $y$ ” son las coordenadas de interés, son las landmarks generadas por la función de mediapipe,  $[i]$  son los puntos con los que se rastreará el movimiento. “ $h$ ” es el alto y “ $w$ ” es el ancho del fotograma.

Con esta ecuación es posible obtener distintos pares de coordenadas en los ejes ( $x$ ,  $y$ ). El siguiente paso es aplicar momentos invariantes. Los momentos invariantes



Fig. 4. Señal obtenida por el movimiento de cabeza del participante.

propuestos por Hu [14] extraen características para reconocer objetos que no se encuentran en la misma posición. En nuestro caso, utilizamos el momento de Hu para calcular el punto exacto donde el landmark genera un movimiento de la siguiente forma [14]:

$$m_{pq} = \sum_x \sum_y (x)^p (y)^q f(x, y), \quad (2)$$

donde  $m_{pq}$  es el momento resultado dentro Del fotograma, “ $p$ ” y “ $q$ ” son el orden del momento, “ $x$ ”, “ $y$ ” son las coordenadas correspondientes al píxel dentro del fotograma. Al aplicar este método, se ha transformado el vídeo de su plano visual a un plano de frecuencias, creando diferentes señales para cada participante. El siguiente paso es homogeneizar esas señales, lo cual se explica a continuación.

### 3.1.4. Homogeneización de la señal

Al aplicar los pasos anteriores, se genera una señal que ha capturado los macro y micro movimientos de los participantes durante la estimulación de una emoción y las ha transformado en una señal con diferentes tamaños de longitud, unas más largas que otras. Esto sucede porque el algoritmo solo rastrea los movimientos que hace el participante al ser estimulado sin importar el tamaño del videoclip, además de que un participante puede tener una reacción en un punto del videoclip diferente al otro participante y, aunque el algoritmo registra las dos reacciones, no crea la misma señal para los dos. Esto genera un problema ya que, al no tener señales con tamaños iguales, se puede crear un sobre entrenamiento en un algoritmo de clasificación.

Para solucionar este problema, se propone utilizar un método de regresión polinomial que sirve para homogeneizar las señales. La regresión polinomial es un modelo de análisis de regresión en el que la relación entre la variable independiente  $X$  y la variable dependiente  $Y$  se modela con un polinomio de  $n$ -ésimo grado en  $X$ . Para

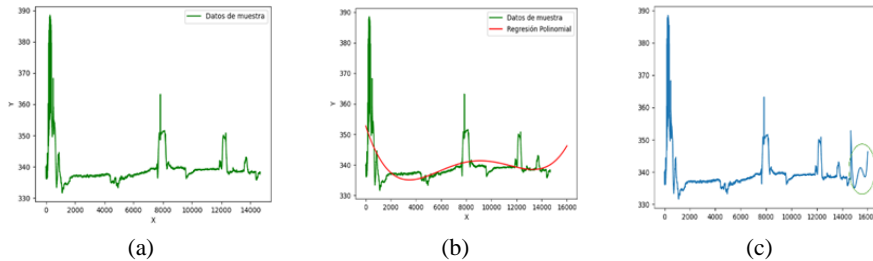


Fig. 5. Resultado al aplicar regresión polinomial.

obtener la regresión polinomial se debe calcular el polinomio con la siguiente ecuación [15]:

$$y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \dots + \beta_p X^p + \varepsilon, \quad (3)$$

donde  $\varepsilon$  representa el error en la estimación o la diferencia entre el valor estimado y el valor observado. El polinomio resultante de la ecuación 3, puede ser abordado mediante una regresión lineal múltiple aplicando las ecuaciones 4 y 5:

$$X_1 = X, X_2 = X^2, \dots, X_p = X^p, \quad (4)$$

$$Y = \beta_0 + \beta_1 X^1 + \beta_2 X^2 + \beta_3 X^3 + \beta_p X^p + \varepsilon. \quad (5)$$

Para la obtención de la regresión lineal múltiple, es posible plantear el modelo a partir de una matriz para  $n$  muestras de datos, tal como se presenta en la ecuación 6:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}. \quad (6)$$

De este modo, mediante la notación de matrices es posible expresar la ecuación 6, tal como se muestra en la ecuación 7:

$$Y = X\beta + \varepsilon. \quad (7)$$

Dado que el objetivo de la ecuación matricial 7, es encontrar el vector de coeficientes  $\beta$ , es posible mediante propiedades matriciales (matriz traspuesta y matriz inversa) expresar la ecuación 8:

$$\beta = (X^T X)^{-1} X^T Y. \quad (8)$$

Así, el vector resultante  $\beta$  contiene los diferentes coeficientes ( $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ ) del polinomio presentado en la ecuación 3. Al aplicar esta técnica, se puede asegurar que todas las señales tendrán el mismo tamaño y no hay grandes variaciones entre ellas, preparándolas para su clasificación.

## 4. Experimentos y resultados

En esta sección se muestran los resultados más relevantes de cada una de las fases del proceso. Los experimentos fueron realizados en el lenguaje de programación Python 3.11.5, utilizando las librerías mediapipe, opencv y scikit-learn. El algoritmo fue desarrollado en una laptop Dell inspiron 3000 con un procesador Ryzen 7 y 24 GB de RAM. Los mejores resultados son descritos a continuación. Se obtuvo un video por cada participante, resultando 24 videos diferentes. Para transformar el video de su plano visual a una señal, se hace un análisis del video, lo cual nos permite obtener los diferentes movimientos que hace el participante durante la estimulación.

Este análisis se realiza utilizando los puntos que generan las funciones de la librería MediaPipe y los resultados de este paso se muestran en la figura 3. En la figura 3 se observan los 3 puntos seleccionados para la cabeza, los cuales se encuentran en la barba, frente, laterales (cercanos a los oídos) y finalmente la nariz. Con estos 3 puntos es posible abarcar toda la zona de interés, en este caso la cabeza. Al tener los puntos específicos seleccionados, el siguiente paso es rastrear los movimientos que hace el participante durante la proyección del videoclip.

Para solucionarlo, se buscan las coordenadas de cada punto (landmarkers) dibujados en la figura 5 utilizando la ecuación 1. Al obtener esas coordenadas, el siguiente paso es aplicar momentos invariantes de Hu (ecuación 2), los cuales convierten el comportamiento de las coordenadas en una frecuencia dentro de un hiperplano. Esto indica que cada vez que el participante hace un movimiento (macro o micro), el momento de Hu lo transforma en una perturbación dentro de la señal que se está capturando.

A mayor movimiento, mayor es la perturbación dentro de la señal; por el contrario, pequeñas o nulas perturbaciones se capturan cuando la persona está sin movimiento. La figura 4 muestra la señal generada por la parte del cuerpo seleccionada. La figura 4 muestra la señal obtenida al aplicar los momentos de Hu a las coordenadas de los landmarks marcados en la figura 3. La imagen muestra que un mayor movimiento se transforma en una perturbación más grande que el resto de la señal. Esto permite crear señales con picos más grandes debido a las diferentes perturbaciones que los momentos de Hu han identificado.

Al analizar los videos de cada uno de los participantes, se puede notar que las 24 señales creadas no son iguales, tienen diferentes tamaños, lo cual puede generar un error al momento de entrenar un algoritmo de aprendizaje. Para solucionar este problema, aplicamos el método de regresión polinomial (ecuación 3), el cual predice datos que nos ayudan a complementar las señales obtenidas por los momentos de Hu y que todas las señales tengan un tamaño estándar. En este experimento, se utilizó un polinomio de grado 4, ya que fue el que mejor ajustó los nuevos datos a los datos originales. La figura 5 muestra el resultado al aplicar la regresión polinomial de grado 4.

### 4.1. Movimientos de cabeza

La figura 5 muestra el resultado de aplicar regresión polinomial de grado 4 a una señal. La imagen está dividida de la siguiente forma: a la izquierda (a) muestra la señal de entrada que fue obtenida por los momentos de Hu. En el centro (b), se muestra la curva obtenida al aplicar la regresión polinomial de grado 4. A la derecha (c), se muestra



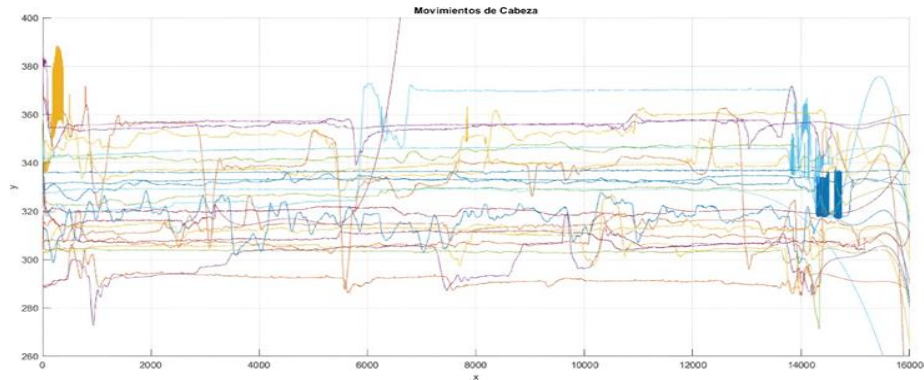


Fig. 6. Resultado al aplicar regresión polinomial.

la señal resultante de este proceso. Los nuevos datos están encerrados en el círculo verde. Como se puede ver en la figura 5, la regresión polinomial crea una curva siguiendo la forma de la señal, tratando de ajustar los picos para generar datos que tengan características similares.

Los datos resultantes crean una nueva señal (figura 8-c círculo verde), la cual es anexada a la señal original y de esta forma se amplifica dicha señal. Al aplicar este método, se ha logrado homogeneizar las señales a un mismo tamaño, lo cual se puede ver en la figura 6. La figura 6 muestra el resultado al aplicar la regresión polinomial a todas las señales. Como se puede ver, todas las señales se han homogeneizado a un mismo tamaño, dejando las señales listas para su análisis.

## 5. Conclusiones y trabajos futuros

Analizar los macro y micro movimientos de alguna parte del cuerpo que realiza una persona representa un problema difícil, ya que las personas pueden expresarse de diferente manera en las mismas situaciones. De esta forma, crear una señal a partir de ese movimiento puede ser complicado. Esto se debe a que el participante puede tener una reacción muy visible o no tener ninguna.

Para abordar este problema, se propone transformar las micro y macro expresiones que realiza esa persona, centrándose en el movimiento de su cabeza cuando se le induce a experimentar una emoción específica. El algoritmo transforma el espacio visual del vídeo en una frecuencia en la que los movimientos se perciben como patrones y se utilizan para discretizar una señal.

Después, se aplica un método de predicción de datos para homogeneizar las señales a un mismo tamaño y así evitar un sobreajuste al momento de clasificar los datos. Los resultados indican que el algoritmo mapea los macro y micro movimientos que realiza la persona al ser estimulada; después, crea nuevos datos a partir del análisis de la señal y los agrega a la misma para tener que todas las señales tengan un mismo tamaño y así estar listas para su análisis.

## 5.1. Trabajos futuros

En trabajos futuros se pretende analizar el movimiento de más partes del cuerpo, además de estimular a los participantes con vídeos de diferentes emociones. Se recolectarán más muestras y se implementarán algoritmos de aprendizaje máquina para clasificar esas señales para encontrar diferentes emociones.

## Referencias

1. Soriano-Méndez, J.J., Riaño-Gómez, D.A., Salazar-Morales, O.: Ratón USB para personas tetraplégicas controlado con el movimiento de la cabeza. *Ingeniería*, vol. 19, no. 2 (2014). DOI: 10.14483/udistrital.jour.reveng.2014.2.a02.
2. Gunes, H., Piccardi, M.: Bi-modal emotion recognition from expressive face and body gestures. *Journal of Network and Computer Applications*, vol. 30, no. 4, pp. 1334–1345 (2007). DOI: 10.1016/j.jnca.2006.09.007.
3. Zatarain-Cabada, R., Barrón-Estrada, M.L., Cárdenas-López, H.M.: Reconocimiento multimodal de emociones orientadas al aprendizaje. *Research in Computing Science*, vol. 148, no. 7, pp.153–165 (2019)
4. Maffei, A., Angrilli, A.: Spontaneous blink rate as an index of attention and emotion during film clips viewing. *Physiology and Behavior*, vol. 204, pp. 256–263 (2019). DOI: 10.1016/j.physbeh.2019.02.037.
5. Demiral, Ş.B., Kure-Liu, C., Benveniste, H., Tomasi, D., Volkow, N.D.: Activation of brain arousal networks coincident with eye blinks during resting state. *Cerebral Cortex*, vol. 33, no. 11, pp. 6792–6802 (2023). DOI: 10.1093/cercor/bhad001.
6. Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., Scherer, K.: Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, vol. 2, no. 2, pp. 106–118 (2011). DOI: 10.1109/t-affc.2011.7.
7. Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardana, G., Chukoskie, L., Snider, J., Jayarathna, S.: Eye movement and pupil measures: A review. *Frontiers in Computer Science*, vol. 3 (2022). DOI: 10.3389/fcomp.2021.733531.
8. Hoffmann, E.R.: A comparison of hand and foot movement times. *Ergonomics*, vol. 34, no. 4, pp. 397–406 (1991). DOI: 10.1080/00140139108967324.
9. Senecal, S., Cuel, L., Aristidou, A., Magnenat-Thalmann, N.: Continuous body emotion recognition system during theater performances. *Computer Animation and Virtual Worlds*, vol. 27, no. 3-4, pp. 311–320 (2016). DOI: 10.1002/cav.1714.
10. Michelini, Y., Acuña, I., Guzmán, J.I., Godoy, J.C.: LATEMO-E: A film database to elicit discrete emotions and evaluate emotional dimensions in Latin-Americans. *Temas em Psicologia*, vol. 27, no. 2, pp. 473–490 (2019). DOI: 10.9788/tp2019.2-13.
11. Bradski, G.: The OpenCV library. *Dr. Dobb's Journal*, vol. 25, no. 11, pp. 120–125, (2000).
12. Google AI for Developers: Guía de soluciones de MediaPipe (2024)
13. Maffei, A., Angrilli, A.: Spontaneous blink rate as an index of attention and emotion during film clips viewing. *Physiology & Behavior*, vol. 204, pp. 256–263 (2019). DOI: 10.1016/j.physbeh.2019.02.037.
14. Sharma, S., Choudhary, S., Kumar-Sharma, V., Goyal, A., Malik-Baliyar, M.: Image watermarking in frequency domain using hu's invariant moments and firefly algorithm. In: *International Journal of Image, Graphics and Signal Processing*, vol. 14, no. 2, pp. 1–15 (2022). DOI: 10.5815/ijigsp.2022.02.01.
15. Chanchí-Golondrino, G.E., Campo-Muñoz, W.Y., Sierra-Martínez, L.M.: Aplicación de la regresión polinomial para la caracterización de la curva del COVID-19, mediante técnicas de machine learning. *Investigación e Innovación en Ingenierías*, vol. 8, no. 2, pp. 87–105 (2020)