

Detección de la pupila mediante la transformada de Hough para la interpretación de movimientos oculares

Jorge Ochoa-Somuano*, Daniel Jafet Vázquez-Bejarano, José Francisco Delgado-Orta, Ángel Salvador López-Vásquez, Omar Antonio Cruz-Maldonado & Ángel Antonio Ayala-Zúñiga

Resumen

La tecnología ha permitido el desarrollo de una infinidad de aplicaciones que permiten al ser humano interactuar con dispositivos como son las computadoras. La mayoría de las soluciones propuestas están dirigidas a personas que tienen capacidades motrices que les permiten interactuar con los dispositivos de entrada, como pueden ser: teclados, ratones, joysticks y mandos de juego, por dar algunos ejemplos. Sin embargo, son pocas las aplicaciones que están dirigidas en proporcionar ayuda a personas con capacidades especiales, por lo que ha surgido la idea de apoyar a personas que tengan problemas motrices y de habla, para que puedan interactuar con computadoras personales o portátiles. Dentro de las actividades de investigación que se realizan en la Licenciatura en Informática de la Universidad de Mar campus Puerto Escondido, se propone la presente investigación como punto de partida para el desarrollo de un software que permita generar documentos de texto con la ayuda de los movimientos oculares realizados por el usuario, de tal forma que se pueda comunicar con otras personas por medio de textos.

Palabras clave: interacción humano-computadora, capacidad motriz limitada, dispositivos de entrada, generación de textos.

Recibido: 29 de agosto de 2020.

Abstract

Technology has allowed the development of an infinity of applications that allow human beings to interact with devices such as computers. Most of the proposed solutions are aimed at people who have motor capacities that allow them to interact with input devices, such as: keyboards, mice, joysticks and game controllers, to give a few examples. However, there are few applications that are aimed at providing help to people with special abilities, so the idea has arisen to support people with speech and motor problems, so that they can interact with personal computers or laptops. Within the research activities carried out in the degree in computer science at the University of Mar campus Puerto Escondido, we propose this research as a starting point for the development of software to generate text documents using eyepieces movements performed by the user, in such a way that the user can communicate with other people through texts.

Key words: human-computer interaction, limited motor ability, input devices, text generation.

Aceptado: 22 de octubre de 2020.

Introducción

La investigación enfocada al procesamiento de imágenes tuvo sus inicios en los años 1920s, y se aplicó primeramente en la industria periodística, cuando se logró por primera vez el envío de imágenes a través de cable submarino entre las ciudades de Londres y Nueva York. Este logro fue posible gracias al desarrollo de la tecnología en las comunicaciones, hecho que permitió reducir el tiempo requerido para transportar una imagen a través del océano Atlántico, de más de una semana a menos de tres horas (González & Woods 2001). Posteriormente, se realizaron las primeras investigaciones para realizar la detección facial en imágenes en los años 1960s, cuando Bledsoe (1964) y su equipo desarrollaron los primeros sistemas de reconocimiento de imágenes; siendo este trabajo precursor al mismo tiempo, de las investigaciones referentes a la interpretación de los movimientos de los ojos iniciado con Yarbus (1967).

Actualmente, el uso de dispositivos e interfaces electrónicas ha pasado a ser parte de la vida diaria en las personas. Las personas que sufren de alguna discapacidad física, no pueden manipular estas interfaces a causa de algún padecimiento que les impide dirigir su cuerpo para realizar determinadas tareas, tales como pulsar una tecla o activar un botón en el dispositivo. En la mayoría de los casos, el único medio para que estos individuos puedan comunicarse con su personal de apoyo es a través del uso de dispositivos electrónicos como parte de sistemas especializados. Esto se ha hecho evidente en las personas cuya única movilidad se limita a la de los ojos. Para estos individuos, el uso de un EGCS (Sistema de Comunicación de Gestos Oculares, por sus siglas en inglés) es de gran importancia, ya que les permite comunicar sus necesidades a otras

personas a través del movimiento de sus ojos. Existen varios tipos de EGCS disponibles en el mercado, sin embargo, normalmente son bastante caros y requieren de un equipo especial para su funcionamiento (Chew & Penver 2019).

En fechas recientes, se han realizado investigaciones enfocadas a la detección de ojos que utilizan diversos enfoques de solución para el proceso de identificación, siendo los más comunes los algoritmos como el de *Haar Cascade*, la transformada de Hough, e inclusive, en algunos trabajos en donde la solución se enfoca en la aplicación de técnicas de umbralización.

Tal es el caso de Yuli *et al.* (2017), en donde se propuso un sistema de seguimiento de la mirada, el cual se utiliza para establecer la comunicación entre personas con discapacidad motriz o que presentan dificultades para hablar y las personas de apoyo que les rodean. El trabajo enfrenta la problemática de una mala interpretación en la comunicación por parte de las personas de apoyo, ante la dificultad de las personas discapacitadas para expresar sus propios deseos de una forma lo suficientemente comprensible para las personas quienes les asisten. Por esto, el sistema propuesto opera con el apoyo de una cámara de una tableta o de un teléfono móvil con el sistema operativo Android. El sistema utiliza el método Haar Cascade para la detección de las características del ojo, donde se detecta primero el ojo derecho que servirá de guía para la detección del ojo izquierdo. Con base en este patrón de identificación, el sistema define una interfaz gráfica de usuario que cuenta con un menú con actividades preestablecidas, como son: comer, beber o encender, dando como salida un audio de la actividad que haya elegido la persona, teniendo como patrón de selección que la persona deberá cerrar el ojo durante más

de 250 milisegundos. Con la solución propuesta, los autores consiguieron un 80 % de éxito en el seguimiento de la mirada a la izquierda, un 80 % para la mirada a la derecha y un 90 % de éxito para la mirada fija.

Alva *et al.* 2017 se propone un sistema para asistir a personas con síndrome de encierro, es decir, pacientes que no pueden mover ninguna parte de su cuerpo excepto los ojos. El trabajo tiene como objetivo establecer una comunicación sencilla entre pacientes paráliticos con sus familiares y amigos, para que el paciente les solicite asistencia para realizar ciertas actividades. Para esto, el sistema propuesto está diseñado para detectar la cara y las pupilas de niños pacientes a través de una *webcam* estándar (o cámara web) utilizando el algoritmo *Haar Cascade* y la Transformada Circular de Hough. El sistema está programado para mostrar las imágenes de diferentes actividades diarias, donde el paciente tiene que mirar una imagen por un tiempo preestablecido para seleccionarlo, y posteriormente, rastrea el punto de vista de la mirada de la persona, haciendo la selección de la imagen después de una confirmación del paciente. Una vez hecha esta confirmación, el asistente notifica el objeto seleccionado vía texto o audio según se elija.

El sistema *GazeSpeak* propuesto por Zhang *et al.* (2017), está enfocado al seguimiento ocular para personas con Esclerosis Lateral Amiotrófica (ELA) o con otros impedimentos motrices. El sistema *GazeSpeak* funciona en un *smartphone*, es de bajo costo, robusto, portátil y fácil de aprender. Está compuesto por tres módulos: reconocimiento de gestos oculares, un motor de texto predictivo y una interfaz de entrada de texto; el módulo principal es para el reconocimiento de gestos oculares.

Por otro lado, Zheng & Usagawa (2018)

propusieron un método rápido de rastreo ocular ligero, de bajo costo, y que responde a una situación en la que se requiere de una alta velocidad de procesamiento. Para ello, el sistema se basa en el uso de una *webcam*. Y aunque el método obtuvo una menor precisión, los autores resolvieron tres problemas que presenta el problema: el alto costo de los dispositivos, la comodidad y la velocidad de carga del cálculo. La *webcam* utilizada cuenta con una resolución de 640 x 480 píxeles, lo que reduce considerablemente el costo de los dispositivos de captura del video. El algoritmo desarrollado adapta el método para lograr el procesamiento de una imagen de baja calidad, en donde la detección de las pupilas se basa en el cambio de la intensidad del color. Los resultados experimentales de la propuesta, muestran que al utilizar la *webcam* se logra una precisión del 78% al 94% para el procesamiento con 25 y 9 secciones respectivamente.

Actualmente DELL (2020) cuenta con un rastreador ocular llamado *Tobii eye*, lo utiliza en equipos *Alienware*, los cuales tienen costos superiores a los \$50,000 (cincuenta mil pesos mexicanos) y permite brindarle diferentes tipos de instrucciones al equipo, una de ellas es que al mirar el logotipo de *alienware* la *laptop* enciende. También se puede adquirir por separado, sus costos fluctúan entre los \$7,500 y los \$30,000 pesos mexicanos.

Por todo lo anterior, observando las aportaciones y las ventajas de la aplicación de las investigaciones previas en sus diferentes entornos, la presente investigación propone una metodología que permite hacer la detección de la pupila para identificar los movimientos oculares. La metodología tiene el objetivo de proveer un sistema que sirva de apoyo a las personas con discapacidad auditiva y vocal, para que se puedan comunicar con la

computadora mediante movimientos oculares. Esto surge como consecuencia de que todas las soluciones actuales encontradas de su tipo son comerciales, y no se cuenta actualmente con herramientas gratuitas disponibles, teniendo además la desventaja de que requieren de equipos muy costosos para su implantación.

Material y métodos

En la figura 1 se muestra la metodología a seguir para la identificación del movimiento ocular. La metodología consta de seis fases funcionales: 1) Adquisición de la imagen, 2) Escala de grises, 3) Negativo de una imagen, 4) Extracción de características, 5) Representación y descripción, y 6) Base de conocimiento.

La metodología se enfoca en el procesamiento de un conjunto de imágenes que deben cumplir con ciertos requerimientos para su captura. Estas alimentan al sistema en la etapa “Adquisición de imágenes”. Posteriormente, se realiza el procesamiento de las imágenes a través de la aplicación de filtros en las etapas “Escala de grises” y “Negativo de una

imagen”, donde estas etapas retroalimentan la entrada del módulo “Base de conocimiento”, a partir del cual se caracterizan las imágenes y se extraen sus características de relevancia en el módulo “Extracción de características”. Por último, las características obtenidas del módulo “Extracción de características” se utilizan para definir la descripción de cada imagen en el proceso de identificación de los ojos, como resultado de la etapa “Representación y descripción”, cuya salida retroalimenta al mismo tiempo a la base de conocimiento con el objetivo de lograr una adaptación del sistema, ante la posibilidad de presentar aspectos multivariantes de las imágenes, originadas por factores físicos como la iluminación y la perspectiva de una imagen existentes en los entornos en donde se realiza su captura, así como de factores referentes a las características propias de los dispositivos de captura que inciden en la calidad de la imagen.

La metodología permite definir un dominio para el problema, a partir del cual se puede obtener como resultado una identificación. Esto se logra a través de la aplicación del procesamiento digital

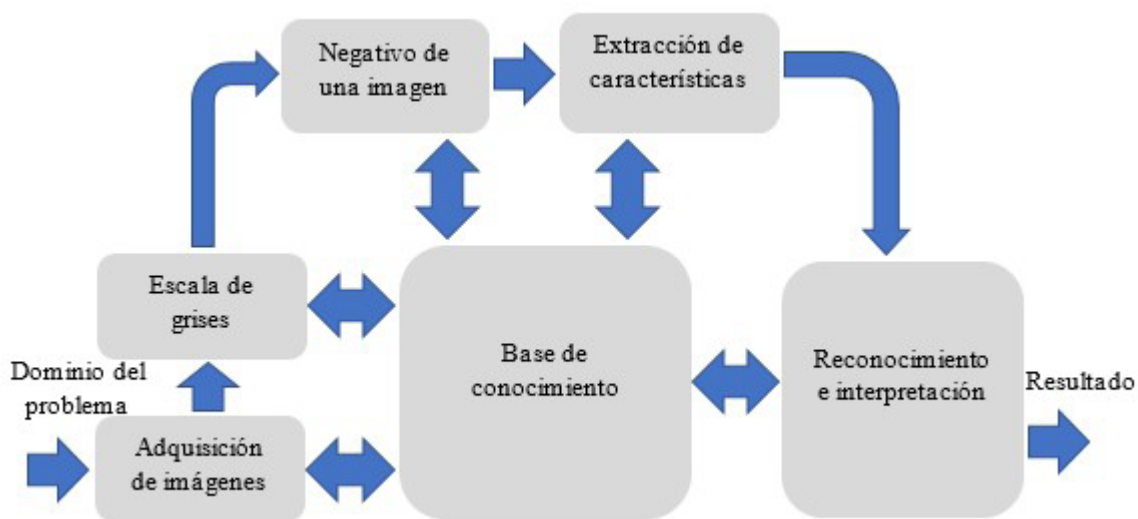


Figura 1. Metodología para la detección de la pupila mediante la transformada de Hough para la detección de movimientos oculares. Tomando como base el diagrama de las etapas para el procesamiento digital de imágenes propuesto por González & Woods (2001).

en las imágenes adquiridas, así como de los enfoques de aprendizaje de máquina empleados en cada una de las etapas que conforman la metodología, cuya descripción se muestra a continuación:

Adquisición de la imagen. Para la obtención de la imagen (frames independientes de una secuencia de imágenes, video) se utilizará la cámara *web* que viene integrada en el equipo de cómputo portable, o bien, una cámara compatible externa, el formato será JPG o PNG y se captura a una distancia aproximada de 50 cm entre la computadora y el usuario, con el objetivo de capturar el rostro completo.

Escala de grises. Partiendo de una imagen original a color (Fig. 2) se obtiene una imagen en escala de grises (Fig. 3) como resultado de obtener la media de las tres matrices de color RGB (rojo, verde y azul) de la imagen, en consecuencia cada pixel en la imagen resultante está en el rango de 0 a 255 (Pereyra 2011).



Figura 2. Imagen original en color RGB, modelo Daniel Jafet.

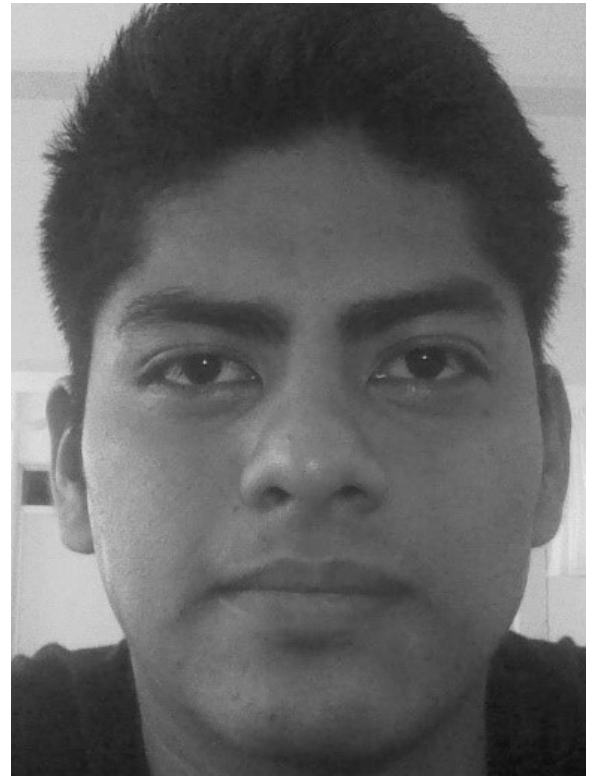


Figura 3. Imagen resultante de aplicar el procedimiento de conversión a escala de grises a la figura 2.

Al convertir la imagen se pueden obtener mejores resultados al aplicar los algoritmos posteriores. La fórmula para convertir a escala de grises una imagen se puede expresar de la siguiente manera (ec. 1):

$$\text{Grises} = (WR * R) + (WG * G) + (WB * B)$$

Asumiendo que R, G y B tendrán los valores de los colores monocromos (Rojo, Verde y Azul) de la imagen a color RGB que son lineales en luminancia. Se establece los valores para cada variable:

$$WR = 0.299$$

$$WG = 0.587$$

$$WB = 0.114$$

Así, la sumatoria de los valores es igual a 1.

$$WR + WG + WB = 1$$

donde WR contiene el valor fijo del color rojo, WG el valor fijo del color verde y WB el valor fijo del color azul.

Negativo de una imagen. Al invertir los niveles de intensidad de una imagen, se produce la equivalente a un negativo fotográfico. Este tipo de procesamiento es particularmente adecuado para realzar los detalles blancos o grises incrustados en las regiones oscuras de una imagen, especialmente cuando las áreas negras son dominantes en tamaño.

Se expresa como en la ecuación 2:

$$s = L - 1 - r$$

donde, r representa el nivel de intensidad de entrada y s el nivel de intensidad de salida.

Extracción de características. Se realizar la transformada de Hough, específicamente para hacer la detección de círculos en la imagen para su interpretación posterior.

En general, se sabe que se necesitan tres parámetros para definir un círculo en una imagen (ver la ecuación 3).

$$C: (x_{center}, y_{center}, r)$$

Donde $(x_{center}, y_{center}, r)$ define la posición del centro del círculo y r es el radio del mismo, con base en estos valores se permite definir completamente un círculo, como se puede ver en la imagen de muestra representada por la figura 5 (Yuen *et al.* 1990).

El detector *Canny Edge* fue desarrollado por John F. Canny en 1986. También se conoce como detector óptimo, este algoritmo de *Canny* tiene como objetivo satisfacer tres principales criterios (Canny 1986).

1. Baja tasa de error: Lo que significa una buena detección sólo de los bordes existentes.

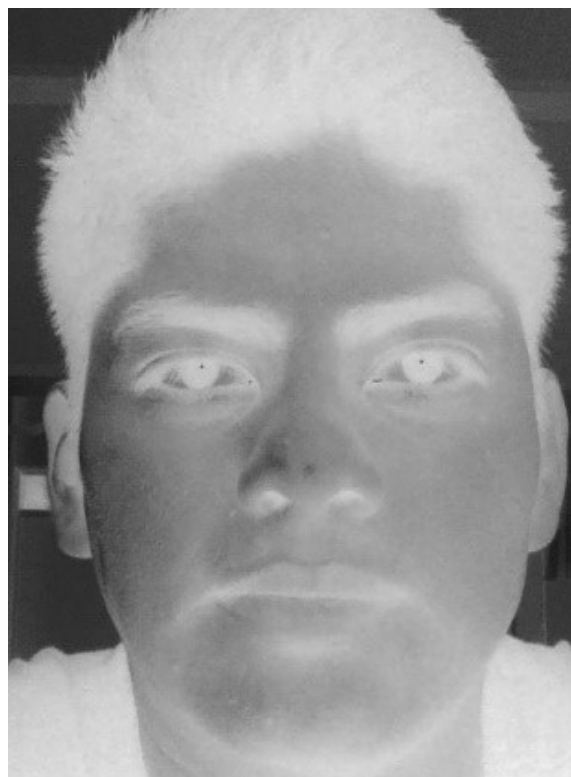


Figura 4. Imagen resultante al aplicar el negativo de una imagen a la figura 3.

2. Buena localización: La distancia entre los píxeles de borde detectados y los píxeles de borde reales debe ser minimizada.

3. Respuesta mínima: Sólo se obtiene una respuesta del detector por borde.

El detector de Canny se aplica a las imágenes en cuatro pasos:

1. Se utiliza el filtro Gaussiano para eliminar el ruido, por ejemplo, se puede usar un filtro de tamaño 5 (ver la ecuación 4).

$$K = \frac{1}{159} \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix}$$



Figura 5. Ejemplo de la aplicación de la transformada de Hough para la detección de círculos en una imagen.

2. Se encuentra el gradiente de intensidad de la imagen, aplicando un par de máscaras de convolución, en las direcciones x e y (ver la ecuación 5).

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} \quad G_y = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}$$

Además, se encuentra la intensidad y dirección del gradiente con las ecuaciones 6 y 7:

$$\text{Ec. 6: } G = \sqrt{G_x^2 + G_y^2}$$

$$\text{Ec. 7: } \theta = \arctang \left(\frac{G_y}{G_x} \right)$$

la dirección redondea a uno de los cuatro ángulos posibles (por ejemplo, 0, 45, 90 o 135).

3. Se aplica una supresión no máxima, es decir, se eliminan los píxeles que no se consideran parte de un borde. Por lo tanto, sólo quedarán líneas finas o llamados bordes candidatos.

4. Histéresis: Es el paso final del algoritmo de *Canny*, se utilizan dos umbrales (superior e inferior):

a) Si el gradiente de un píxel es más alto que el umbral superior, el píxel es aceptado como un borde.

b) Si el valor del gradiente de un píxel está por debajo del umbral inferior, entonces se rechaza.

c) Si el gradiente del píxel se encuentra entre los dos umbrales, entonces será aceptado sólo si está conectado a un píxel que esté por encima del umbral superior.

En la figura 6 se pueden ver los resultados al aplicar el método de detector de bordes *Canny*.

Representación y descripción. En esta última etapa se le da un significado a ciertos movimientos de las pupilas, para que se pueda interpretar y llevar a cabo alguna acción, con base en los movimientos oculares que realice una persona.

Base de conocimiento: interactúa con las demás etapas proporcionándoles información relevante. De este módulo obtienen los datos que necesitan todas las etapas,

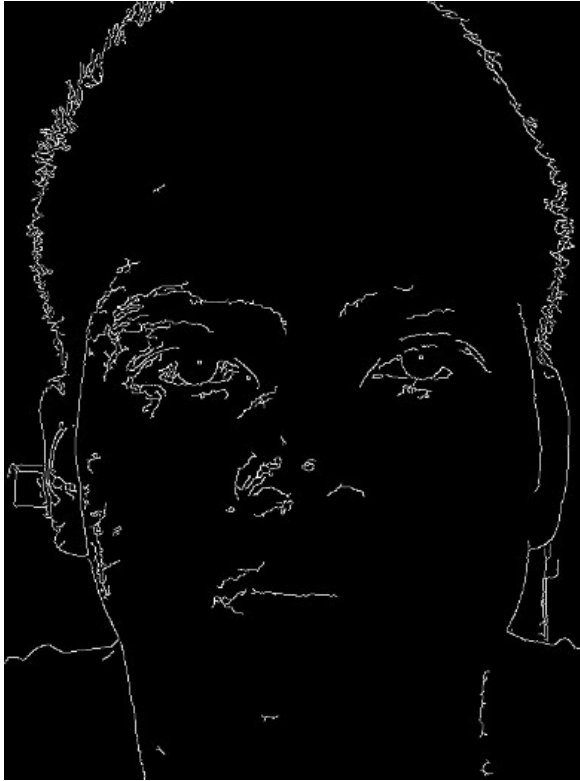


Figura 6. Imagen después de ser procesada con el detector de bordes *Canny*.

para adquirir el conocimiento previo de cómo debe ser el resultado. Este conocimiento además de guiar la operación de cada módulo, ayuda a las operaciones de retroalimentación entre módulos.

Resultados

Como se puede ver en cada uno de los procesos realizados hasta el momento, la detección de la pupila es viable, se puede realizar siguiendo las etapas de la metodología y en los últimos pasos que corresponden al filtro de *Canny* y la transformada de Hough afinar los algoritmos que permitan detectar la pupila sin mayor problema. Finalmente se obtendrá como resultado una imagen como la que se muestra en la figura 7, en este ejemplo se muestra solo una imagen que corresponde a un frame de una secuencia en video, por lo que en la secuencia del video se vería como cambia

la posición de los círculos de color, indicando el seguimiento de la pupila.

Finalmente, lo que se espera realizar como siguiente etapa de la investigación es, interpretar ciertos movimientos de las pupilas realizados por personas para que puedan interactuar con la computadora sin la necesidad de contacto físico.

Discusión

Actualmente no existen, en la Universidad del Mar, trabajos relacionados con la detección de los ojos y específicamente en la detección de las pupilas mediante movimiento ocular. Con la metodología que se propone en el presente documento se espera que las personas puedan interactuar con una computadora.



Figura 7. Resultado esperado en la detección de la pupila de una persona para controlar una computadora por medio de los movimientos oculares.

Los sistemas comerciales para el seguimiento de los ojos actualmente tienen costos muy elevados (<https://www.gazept.com/shop/>) que van de los 1,995 dólares a los 4,550 dólares (\$45,885 - \$104,650 pesos mexicanos), además requieren de dispositivos e instalación especial, lo anterior impide que la mayoría de las personas tengan la capacidad económica de adquirirlos. Sin embargo, el costo e implementación del sistema informático que resulte del presente proyecto de investigación será simbólico, por lo que se podría distribuir de forma gratuita.

Conclusión

La tecnología ha permitido el desarrollo de una infinidad de aplicaciones que permiten al ser humano interactuar con dispositivos como son las computadoras. La mayoría de las soluciones propuestas están dirigidas a personas que tienen capacidades motrices que les permiten interactuar con los dispositivos de entrada, como pueden ser: teclados, ratones, joysticks y mandos de juego, por dar algunos ejemplos.

En los últimos años han aparecido en escena algunos sistemas que permiten a las personas dictar instrucciones a las computadoras por medio de la voz, de tal forma que no se necesita una interacción física con la computadora por medio de los dispositivos mencionados anteriormente, lo único que se necesita es tener un micrófono, que en la mayoría de las veces ya está incluido en la computadora. Sin embargo, son pocas las aplicaciones que están dirigidas en proporcionar ayuda a personas con capacidades especiales, por lo que ha surgido la idea de apoyar a personas que tengan problemas motrices y de habla, para que puedan interactuar con computadoras personales o portátiles,

inicialmente se desarrollará un software que permita generar documentos de texto con la ayuda de los movimientos oculares realizados por el usuarios, de tal forma que se pueda comunicar con otras personas por medio de textos.

Agradecimientos

Agradecemos las observaciones y recomendaciones realizadas por los revisores para mejorar el contenido del presente documento.

Referencias

- Alva, M., N. Castellino, R. Deshpande, K. Sonawane & M. Lopes. 2017. An image based eye controlled assistive system for paralytic patients. Mumbai: 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA). DOI: 10.1109/CSCITA.2017.8066549.
- Bledsoe, W.W. 1964. The model method in facial recognition. Computer Science.
- Canny, A. 1986. Computational Approach to Edge Detection, IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(6), pp. 679-698.
- Chew, M.T. & K. Penver. 2019. Low-Cost Eye Gesture Communication System for People with Motor Disabilities. Auckland, New Zealand: IEEE International Instrumentation and Measurement Technology Conference (I2MTC). DOI: 10.1109/I2MTC.2019.8826976.
- Dell. 2020. Página web oficial de DELL. Consultada el 11 de junio de 2020: <https://www.dell.com/en-us/shop/dell-laptops/alienware-m17-r2-gaming-laptop/spd/alienware-m17-r2-laptop>
- Gonzalez, R., & R. Woods. 2001. Digital Image Processing (2da ed.). Nueva Jersey: Prentice Hall.
- Pereyra, C. G. 2011. Reconocimiento de rostros en un ambiente de iluminación controlado no variante a expresiones faciales. Tesis de licenciatura: Universidad del Mar, campus Puerto Escondido, Oaxaca, México.
- Yarbus, A. L. 1967. Eye Movements and Vision. Moscow: Springer.
- Yuen, H. K., J. Princen, J. Illingworth & J. Kittler. 1990. Comparative study of Hough transform methods for circle finding. Image Vision Comput. 8(1), pp. 71-77.
- Yuli C.R., R. Sigit, T. Harsono, D.C. Adelina, A. Nabilah, & N.P. 2017. Anggraeni. Eye gaze tracking to operate android-based communication helper application.

Surabaya: International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), Surabaya.

Zhang, X., H. Kulkarni, & M.R. Morris. 2017. Smartphone-Based Gaze Gesture Communication for People with Motor Disabilities. Denver, Colorado: CHI Conference on Human Factors in Computing Systems Association. DOI: 3025453.3025790.

Zheng, C., y T. Usagawa. 2018. A Rapid Webcam-Based Eye Tracking Method for Human Computer Interaction. Hangzhou: International Conference on Control, Automation and Information Sciences (ICCAIS). DOI: 10.1109/ICCAIS.2018.8570532.