INTERFAZ HUMANO-COMPUTADOR PARA PERSONAS CON LIMITACIÓN MOTRIZ DE MIEMBROS SUPERIORES BASADA EN GESTOS FACIALES

JOHN ALEX PINO MURCIA
LUIS FELIPE MOCTEZUMA RUIZ
JOHNATAN BURGOS MARTINEZ

FUNDACIÓN UNIVERSITARIA CATÓLICA LUMEN GENTIUM
FACULTAD DE INGENIERÍA
INGENIERÍA DE SISTEMAS

2020**INTERFAZ HUMANO-COMPUTADOR PARA PERSONAS CON LIMITACIÓN MOTRIZ DE MIEMBROS SUPERIORES BASADA EN GESTOS FACIALES**

JOHN ALEX PINO MURCIA
LUIS FELIPE MOCTEZUMA RUIZ
JOHNATAN BURGOS MARTINEZ

Trabajo de grado para optar al título de
Ingeniero de Sistemas

Directores
CARLOS DIEGO FERRÍN BOLAÑOS, M. Sc., Ing.
JOSÉ HERNANDO MOSQUERA DE LA CRUZ, M. Sc., Ing.

FUNDACIÓN UNIVERSITARIA CATÓLICA LUMEN GENTIUM
FACULTAD DE INGENIERÍA
INGENIERÍA DE SISTEMAS
2020

Nota de aceptación:

Aprobado por el Comité de Grado en cumplimiento de los requisitos exigidos por la Fundación Universitaria Católica Lumen Gentium Facultad de Ingeniería, Programa de Ingeniería de Sistemas, para optar al título de Ingeniero en Sistemas.

_____

Jurado

_____

Jurado

Santiago de Cali, 26 de mayo de 2020

DEDICATORIA

John Alex Pino Murcia: A mi esposa Leidy Johana Velásquez e hija María Camila Pino Velásquez por su paciencia, comprensión y solidaridad con este proyecto, por el tiempo que me han concedido, un tiempo robado a la historia familiar. ¡Sin su apoyo este trabajo no hubiera sido posible, por eso, este trabajo es también de ustedes, las amo!

Luis Felipe Moctezuma Ruiz: Le dedico este trabajo de grado a la memoria de mi madre Sonia Ruiz Pisamina, que siempre me impulsó en los momentos más duros de mi vida y me dio fortaleza para continuar con mis estudios.

Jonathan Burgos Martínez: Este trabajo investigativo lo dedico principalmente a Dios y en especial a mi madre Amparo Martínez Mena, por ser la inspiración y darme fuerzas para continuar en este proceso de obtener uno de los anhelos más deseados y mi sueño hecho realidad.

AGRADECIMIENTOS

John Alex Pino Murcia: ¡Gracias especiales a mis padres y a mi esposa, porque fueron mi apoyo moral y motivador día a día, para continuar y finalizar este proceso de formación profesional sin rendirme! En segundo lugar, deseo expresar mi agradecimiento a los directores de esta tesis, Carlos Diego Ferrín Bolaños y José Hernando Mosquera De La Cruz, por todo el apoyo que han brindado para la realización de este trabajo, por el respeto a mis sugerencias e ideas y por la dirección y el rigor que han facilitado a las mismas. Gracias por la confianza ofrecida.

Luis Felipe Moctezuma Ruiz: Agradezco a Dios y mi familia, en especial a mis padres por su constante apoyo, dedicación y motivación durante todo este proceso, poniendo su grano de arena en cada etapa y dando el impulso extra que necesitaba.

Jonathan Burgos Martínez: Agradecemos a Dios por bendecirnos en cada momento, por guiarnos a lo largo de nuestra existencia, ser el apoyo y fortaleza en aquellos momentos de dificultad y de debilidad.

CONTENIDO

LISTA DE FIGURAS

pág.

LISTA DE TABLAS

pág.

GLOSARIO

INTERFAZ HUMANO MÁQUINA: es el medio por el cual, un usuario puede comunicarse con una máquina y que abarca todos los puntos de contacto entre este y el máquina en cuestión.

VISIÓN POR COMPUTADOR: disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica para que puedan ser tratados por un ordenador.

GESTOS DEL ROSTRO: junto con la mirada, es uno de los medios más importantes para expresar emociones y estados de ánimo.

WHATSAPP: aplicación móvil, web y de escritorio tipo red social adquirida por Facebook.

TETRAPLEJICO: persona que ha perdido totalidad o parte de la actividad motora/sensorial de los miembros superiores e inferiores.

LIMITACIÓN MOTRIZ: deficiencia que provoca en el individuo que la padece alguna disfunción en el aparato locomotor.

RELACIÓN DE ASPECTO: la relación de aspecto es un atributo de proyección de imagen que describe la relación proporcional entre la anchura y altura de una imagen.

DISTANCIA EUCLIDIANA: la distancia euclidiana es la distancia “ordinaria” entre dos puntos de un espacio euclídeo, esta se deduce a partir del teorema de Pitágoras. Sirve para definir la distancia entre dos puntos en espacios bidimensionales, tres o más dimensiones, permite hallar la longitud de un segmento definido por dos puntos de una recta, del plano o de espacios de mayor dimensión.

RELACIÓN SEÑAL A RUIDO: es el cociente entre la cantidad de luz y la cantidad de ruido que tiene un píxel.

FILTRO DE MEDIANA: para el suavizado de una imagen, el filtro de mediana reemplaza el valor de gris de un punto por la mediana de los niveles de gris de una cierta vecindad.

KERNEL: es una matriz de coeficientes que asigna una serie de valores a los píxeles vecinos del píxel de referencia.

CORRELACIÓN CRUZADA: es la operación de calcular el producto interno de una plantilla con el contenido de una ventana de imagen (cuando la ventana se desliza sobre todas las posiciones de imagen posibles).

PROFUNDIDAD DE IMAGEN: número de bits que se han puesto a disposición para representar cada píxel en la imagen.

FRAME: Un fotograma o frame es cada una de las imágenes que forman un vídeo. Se expresan con las siglas fps y en hercios (Hz).

RESUMEN

El lenguaje corporal es importante para comunicarse fluidamente con las personas. En el ámbito de la interacción con máquinas, existen sistemas para reconocer automáticamente gestos faciales. En el caso de personas con limitación motriz de miembros superiores los gestos faciales son la única forma de comunicarse con el mundo, sin embargo, las interfaces actuales no tienen en cuenta las reducciones de movilidad que la mayoría de las personas con limitación motriz experimenta durante sus periodos de recuperación. Para contribuir a la solución de este problema, se presenta una interfaz humana máquina que mediante técnicas de visión por computador detecta, sigue y estima la pose de un rostro a partir de imágenes de capturas mediante webcam, para así generar comandos en una aplicación de WhatsApp. Con el fin de evitar fatiga, el sistema detecta de forma automática, utilizando capturas de pantalla y técnicas de reconocimiento de patrones, las regiones de interacción clave dentro de WhatsApp como son la zona de selección de conversación y la de chat. La interfaz es complementada con módulos para configurar su funcionamiento y controlar un cursor. La programación se realiza en Python 3.6.8 y utiliza las librerías OpenCV Versión 4.2.0, DLIB Versión 19.8.1 para procesar imágenes, PyAutogui Versión 0.9.42 para capturar la pantalla del computador, emular eventos de teclado y cursor, y la biblioteca Qt Versión 4.8 para exponer una interfaz de usuario sencilla. El desempeño se evalúa con videos de personas utilizando cuatro comandos de interacción con WhatsApp. En las pruebas se utilizan varios tipos de iluminación, fondos, distancias a la cámara, posturas y velocidades de movimiento. Los resultados muestran que el algoritmo detecta, sigue y estima la pose del rostro en 85 % de los casos. Se producen fallas cuando hay fuentes de luz frente a la cámara, oscuridad o movimiento detrás del rostro. El programa se ejecuta a 1 FPS y utiliza el 35% de un procesador Intel Core i5 y 1.5 GB de RAM. La plataforma es capaz de distinguir rostros de variados tonos, y es muy robusto a cambios de iluminación, presencia de otras personas y oclusiones parciales del rostro en la aplicación WhatsApp. La facilidad de interacción y la rápida curva de aprendizaje, así como su conexión directa con el aplicativo orientado al internet permiten prever que la interfaz desarrollada fortalecerá los procesos de inclusión de más personas con limitación de miembros superiores como los tetrapléjicos y las personas con accidente cerebro vascular en los contextos sociales y de transformación digital.

Palabras clave: Gestos faciales, limitación motriz, interfaz humano-máquina, visión por computador.

ABSTRACT

Body language is important to communicate fluently with people. In the area of machine interaction, there are systems to automatically recognize facial gestures. In the case of people with motor limitation of upper limbs, facial gestures are the only way to communicate with the world, nonetheless current interfaces do not take into account the mobility reductions that most people with motor limitations experience during their periods of recovery. In order to contribute to the solution of this problem, a human machine interface is here described. Our solution uses computer vision techniques to detect, track and estimate head pose from images acquired by a webcam to generate commands into WhatsApp application. For avoiding user fatigue, the system automatically detects key interaction regions within WhatsApp such as the conversation selection area and the chat selection area by using desktop screenshots and pattern recognition techniques. The interface is complemented with modules to configure its operation and commanding computer cursor. Programming is realized through Python 3.6.8, OpenCV Version 4.2.0 and DLIB Version 19.8.1 libraries to process and analyze webcam images, PyAutogui Version 0.9.42 to acquire computer screen, emulate keyboard and cursor events, and Qt library Version 4.8 to expose a simple user interface. Performance is assessed by videos of people using four interaction commands with WhatsApp. Various types of lighting, backgrounds, camera distances, postures, and movement speeds are used during testing. The results show that the algorithm detects, tracks and estimates head pose in 85% of cases. Failures occur when there are light sources in front of the camera, darkness, or movement behind the face. Our application executes at 10 FPS of speed and uses 16% of an Intel Core i5 processor and 2 GB out of 8GB RAM. The platform is able to distinguish faces of similar skin tone, however is not very tolerant to background movements. The ease of interaction and the fast learning curve allow to conceive the interface developed for applications with upper-limb disabled people, such as tetraplegics and people with stroke.

Keywords: Facial gestures, motor limitation, human-machine interface, computer vision.

INTRODUCCIÓN

Una interfaz hombre máquina [1] permite que tanto personas comunes como personas que padecen alguna limitación motriz interactúen con aparatos. Se compone de mecanismos para controlar y consultar el estado del sistema. Los dispositivos más comunes y antiguos son de tipo electro-mecánico: botones, perillas, teclados, ratones, entre otros. Con el tiempo, estos sistemas se han ido haciendo intuitivos, de modo que las personas sin entrenamiento puedan utilizarlos con facilidad, como ocurre con los teléfonos celulares y ordenadores. La inteligencia artificial, reconocimiento de habla y visión computacional son objeto de investigación y en el futuro podrían generar aplicaciones en las que personas y máquinas colaboren para realizar tareas eficientemente en hogares, oficinas, etc. [2].

Existen herramientas para reconocer automáticamente indicaciones visuales simples. Algunas de estas iluminan la escena con luz infrarroja y examinan los reflejos para obtener gestos del rostro [3]. Otras aproximaciones emplean una cámara bidimensional y utilizan métodos de detección de piel [4]. Ninguno de estos procedimientos equipara las capacidades humanas y ambos pueden fallar en presencia de objetos similares o en ciertos ambientes, debido a limitaciones de los sensores y algoritmos. Por otra parte, las webcams están presentes en prácticamente todos los ordenadores modernos, por lo que un desarrollo que haga uso de ellas y resuelva sus problemas podría tener utilidad inmediata para muchos usuarios.

El propósito de este documento es describir una interfaz humano-máquina basada en visión artificial para identificar gestos faciales y zonas de interacción en la aplicación WhatsApp a partir de una webcam y captura de pantallas. El programa detecta, sigue y estima el rostro utilizando detección facial y puntos característicos. El sistema permite desplazar el cursor hacia la zona de contactos y la zona de chat mediante ligeros movimiento laterales del rostro, y también permite hacer scroll-up y scroll-down sobre las mismas zonas con ligeros movimientos hacia arriba y hacia abajo. Abriendo la boca ligeramente se lograr activar y desactivar la IHC.

Para evitar la fatiga en las personas con limitación motriz, en esta propuesta a diferencia de las propuestas encontradas en el estado del arte que controlan la posición del cursor con los movimientos del rostro, se identifican las zonas de interacción independiente de la posición y tamaño de la ventana de WhatsApp en el escritorio del computador, esto permite que los usuarios puedan ubicar el cursor en las zonas de interacción con movimientos menos sostenidos que los que convencionalmente se debe realizar en otro tipo de interfaces.

El documento está estructurado así: en los capítulos 1, 2, 3, 4 y 5 se describen el planteamiento del problema, la justificación, los objetivos tanto generales como específicos, el marco referencial y la metodología de la investigación, respectivamente. Estos capítulos retoman lo concebido durante la fase de anteproyecto y además han sido complementados (específicamente el marco de referencia) para evidenciar lo alcanzado durante la ejecución del primer objetivo de este proyecto. En el capítulo 6 se describen las herramientas y técnicas tanto matemáticas como computacionales utilizadas para concebir la solución propuesta de interfaz gestual la cual es descrita en el capítulo 7. Estos capítulos evidencian lo alcanzado tras la consecución de los objetivos 2 y 3 del proyecto. Finalmente, los capítulos 8, 9 y 10 relacionados con las pruebas experimentales/resultados, conclusiones y recomendaciones futuras respectivamente, permiten evidenciar los alcances y límites de la solución propuesta, lo cual es a su vez el último objetivo de este proyecto.

1. DESCRIPCIÓN DEL PROYECTO

1.1. Definición del problema

En la actualidad el dominio de computadoras se ha vuelto una de las necesidades apremiantes de nuestra sociedad moderna, permitiendo no solo acceder a información relevante, sino también como una herramienta comunicativa y de socialización. Es decir, poder manejar dispositivos computacionales permite interactuar con el resto de la humanidad. Sin embargo, toda esta revolución informática, presenta barreras limitantes que interfieren en el acceso directo a este mundo moderno, un ejemplo de ello lo constituyen la población con alteraciones motoras, especialmente las que presentan limitación motriz de miembros superiores. En Colombia aproximadamente un 2,6% de la población padece algún tipo de limitación según[5]. El diseño de interfaces Humano-Computador para facilitar la interacción de este tipo de personas con aplicaciones orientadas a internet constituye todavía un problema abierto dentro de la comunidad científica. Dentro de las tipologías de IHC más investigadas se encuentran las basadas en gestos, exactamente las basadas en gestos faciales, ya que los miembros superiores no son un mecanismo idóneo en este tipo de personas para la generación de gestos. Por esto en este trabajo se define el problema de investigación mediante la siguiente pregunta abierta: ¿Cuál es el diseño de una interfaz humano-computador que permita a personas con limitación de miembros superiores realizar tareas básicas en una aplicación orientada a internet mediante gestos faciales?

1.2. Antecedentes del problema

Nacionales:

En 2005 en la Universidad Javeriana se reporta el desarrollo de un sistema electrónico capaz de reconocer, en tiempo real, doce gestos realizados por un interlocutor en una escena con iluminación y con fondo controlados. Aun cuando el sistema constituye un gran aporte a línea aquí investigada, el sistema se limita únicamente a gestos de mano. La interfaz es robusta a rotaciones, translaciones y cambios de escala de la mano del interlocutor en el plano de la cámara, estas características son importantes para tener en cuenta en el caso de utilizar gestos faciales [6].

En 2009 en la Universidad Nacional Sede Manizales, se trabajó una interfaz hombre máquina que permitía ofrecer una “mano adicional” para controlar la laparoscopia a un cirujano, cuando se encuentra desarrollando una intervención quirúrgica de este tipo, lo que a menudo se hace muy necesario debido a que, en la mayoría de las veces, éste tiene ambas manos e incluso ambos pies ocupados manipulando instrumentos quirúrgicos. La interfaz utilizaba gestos del rostro, específicamente de la postura de los labios para generar el comando. Como resultado se demostró que un cirujano podía de manera fácil y precisa, controlar el brazo de un robot, haciendo simplemente los gestos faciales adecuados, sin tener que usar interruptores o comandos de voz para iniciar la secuencia de control. Este tipo de ideas se evidencia en desarrollos actuales y constituye un gran aporte si pensamos en pacientes con capacidad de producir gestos faciales, pero con limitaciones de miembro superior [7].

En 2017, en La Universidad del Valle se desarrolló una interfaz audiovisual (audio y voz) para comandar varias aplicaciones orientadas a la Web (Google, Facebook y Gmail). En la parte gestual utilizaron algoritmos convencionales de detección de rostros para la manipulación del cursor y utilizaron la detección de guiño para la emulación de clic derecho. La propuesta se caracteriza por estar centrada en el usuario y ser independiente la aplicación, obteniéndose resultados sobre salientes principalmente con personas normales [8].

Internacionales:

Camera Mouse es un sistema de interfaz que rastrea los movimientos del usuario con una cámara de video y los traduce a los movimientos del puntero del mouse en la pantalla. Se puede simular un clic izquierdo del mouse al pasar el puntero sobre el icono que se va a seleccionar [9].

Facial Mouse es un sistema emulador de ratón basado en el movimiento facial del usuario. Se coloca una cámara web frente al usuario, enfocándose en la cara del mismo. Luego, se utiliza un algoritmo de extracción de movimiento, que es independiente del usuario, para extraer el movimiento facial del video. Este movimiento se usa para mover el puntero del mouse que se controla de una manera relativamente similar a los dispositivos de mouse estándar [10].

FaceMouse es otro sistema que utiliza una cámara web estándar y técnicas de visión por computadora para rastrear la nariz de la persona y usar esto para mover el puntero del mouse (de acuerdo con la dirección del movimiento de la nariz) [11].

Estudiantes del Departamento de Matemáticas e Informática de la Universidad de Las Islas Baleares, en el año 2008, desarrollaron un sistema que permite a las personas con discapacidad motriz acceder a la computadora a través de los movimientos de la cabeza del usuario. El sistema no requiere calibración y detecta automáticamente la cara usando el algoritmo Viola y Jones. A continuación, el método divide la cara en regiones: ojos, cejas, boca y nariz. Se utiliza también un método Gaussiano 3D para detectar la región del color de la piel. Para determinar las regiones de los ojos y las cejas se realiza el umbral de imagen. El único gesto facial a tener en cuenta es el parpadeo. El movimiento del mouse se realiza mediante la posición de la nariz y el parpadeo del ojo puede tener diferentes funciones. De esta manera, las personas sin movimiento en los miembros superiores pueden controlar la computadora [12].

Por lo general, los sistemas de interacción basados en visión por computador detectan algunas partes del cuerpo humano los cuales se utilizan para interactuar con una computadora. La mayoría de los trabajos encontrados en la revisión de la literatura hasta ahora asumen que el usuario puede hacer movimientos, aunque sean pequeños, con la cabeza, la mano o los ojos y así controlar la computadora. Desafortunadamente, en algunos casos de limitaciones severas, no hay movilidad de los miembros superiores y en el mejor de los casos se puede contar con movimientos leves del rostro e inclusive solo ciertos músculos faciales pueden moverse, lo que lleva a que la mayoría de los sistemas de interfaz existentes no sean lo suficientemente robustos para que un usuario con limitación motriz de miembros superiores pueda usar para tareas básicas de interacción, como la elección la selección de una zona especial dentro de una interfaz gráfica de usuario. En casos de restricciones extremas de movimiento, las expresiones y movimientos leves faciales son la única alternativa para interactuar con la computadora [13].

1.3. Formulación del problema

El diseño de interfaces Humano-Computador para facilitar la interacción de personas con limitación motriz de miembros superiores [15] con aplicaciones orientadas a internet constituye todavía un problema abierto dentro de la comunidad científica. Dentro de las tipologías de IHC más investigadas se encuentran las basadas en gestos, exactamente las basadas en gestos faciales, ya que los miembros superiores no son un mecanismo idóneo en este tipo de personas para la generación de gestos. Por esto en este trabajo se define el problema de investigación mediante la siguiente pregunta abierta: ¿Cuál es el diseño de una interfaz humano-computador que permita a personas con limitación de miembros superiores realizar tareas básicas en una aplicación orientada a internet mediante gestos faciales?

2. JUSTIFICACIÓN

Escribir en un ordenador, navegar por Internet, leer un texto en pantalla o mover el ratón son actividades sencillas y rutinarias para la mayoría. En Colombia aunque no se cuenta con una cifra exacta de personas con discapacidad, según informe presentado por el Ministerio de Salud, a partir del año 2002 a través del RLCPD (Registro de Localización y Caracterización de Personas con Discapacidad), se ha determinado que un millón cuatrocientas cuatro mil ciento ocho personas presentan algún tipo de discapacidad [5], es decir, que un dos punto seis por ciento de la población total, padece alguna limitación física que restringe su acceso a estas herramientas, aumentando los índices de exclusión social.

Con este proyecto de grado se busca facilitar la interacción humano computador de las personas con limitación de miembros superiores quienes pueden realizar algún tipo de movimiento o gesto en el rostro. El desarrollo de este tipo de proyecto busca además fortalecer una línea de investigación dentro del grupo Khimera de la Fundación Universitaria Católica Lumen Gentium (Cali-Colombia) en interfaces humano-computador. Independiente de cualquier otro interés la meta final es la de impactar la calidad de vida de las personas (como la de sus familiares) mediante tecnologías que aumenten su nivel de inclusión social.

3. OBJETIVOS

3.1 GENERAL

Desarrollar una interfaz humano computador basada en gestos faciales y detección de zonas de interés en una aplicación orientada al internet para personas con limitaciones motrices de miembros superiores.

3.2 ESPECIFICOS

- Determinar las principales características de una interfaz humano-computador para personas con limitaciones motrices de miembros superiores.

- Implementar una técnica de visión por computador para la identificación de gestos faciales y la detección de zonas de interés en una aplicación de escritorio orientada a internet.

- Desarrollar una interfaz software de generación de comandos para una aplicación de escritorio a partir de gestos y zonas de interés detectadas.

- Ejecutar un plan de pruebas para definir los alcances y limitaciones del sistema.

4. MARCO REFERENCIAL

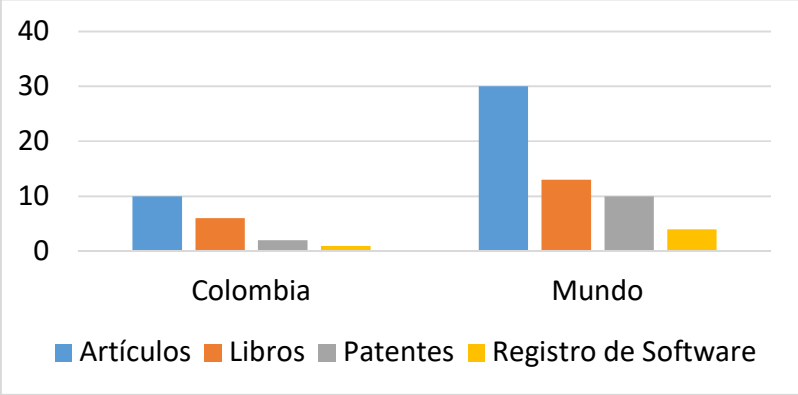
4.1 MARCO TEÓRICO

Las principales teorías alrededor del problema planteado se han alineado en proponer IHC gestuales enfocadas solo en el usuario [8], [14], [15]. En esta propuesta consideramos que debe extraerse información contextual del aplicativo para facilitar la interacción y hacerla más robusta. Por tal razón se ha planteado la siguiente hipótesis: Para facilitar la interacción de los pacientes con limitación de miembro superiores con el computador no solo debe generarse comandos a partir del movimiento y gestos del rostro sino también de información semántica de la aplicación que se desea comandar.

Para identificar las principales teorías en torno a la problemática aquí planteada se realiza una búsqueda en bases bibliográficas como Scopus, Science Direct, IEEE Xplorer utilizando el siguiente criterio de búsqueda tanto en inglés como español: Interfaz Humano Maquina **OR** Computador **AND** Gestos Faciales **AND** Visión Por Computador. Como resultado y después de excluir la bibliografía que no está directamente relacionada, se identifican 87 publicaciones las cuales se distribuyen por categoría como lo muestra la Figura 1.

Dentro de esta búsqueda se resaltan los siguientes aspectos: La construcción de una interfaz hombre máquina requiere variadas consideraciones. Por un lado, las características físicas y psicológicas de las personas deben ser tomadas en cuenta para que el dispositivo sea agradable y seguro. Por otra parte, el sistema se debe poder realizar con los recursos existentes, como presupuesto, tecnologías y conocimientos. Finalmente, el resultado debe garantizar un alto grado de confiabilidad para ser utilizable. Todos estos aspectos son críticos para que un HMI esté al alcance del público objetivo y sea confortable [16]. Los resultados facilitan la utilización de dispositivos, aumentan la seguridad y mejoran la productividad. Cuando la ergonomía no es considerada, los usuarios se pueden cansar, necesitar largos períodos de aprendizaje y sufrir accidentes [17]. Por lo tanto, a la hora de diseñar equipamiento y sistemas interactivos, es importante organizar los recursos para que el producto sea amigable con las personas.

Figura 1. Número de publicaciones nacionales e internacionales por categoría.



Fuente: Los autores con datos de Scopus, Science Direct, IEEE Xplorer

Una interfaz hombre máquina se compone de recursos y conocimientos provenientes de diferentes áreas [2]. Por una parte, el hardware da forma física al dispositivo y proporciona los medios para su operación. Por otro lado, los procedimientos regulan el comportamiento de la aplicación ante las acciones del usuario o condiciones ambientales. A esto se suman consideraciones económicas, que determinan la factibilidad de realizar el diseño e inciden en la conveniencia de utilizar el dispositivo. El resultado, para que sea útil, debe ser accesible y cumplir con su cometido.

Una interfaz tiene que ser confiable para su utilización. Dependiendo del ámbito de aplicación, una falla puede generar desde molestia hasta accidentes. Por ejemplo, un vehículo autónomo puede manejarse por sí mismo y obedecer comandos verbales. En este caso, un error puede costar la vida de los ocupantes y peatones, por lo que este sistema tiene que funcionar a la perfección para que alguien se atreva a utilizarlo. Si bien es relativamente sencillo crear un prototipo para demostrar un concepto, llevarlo a la práctica requiere un nivel de desempeño superior.

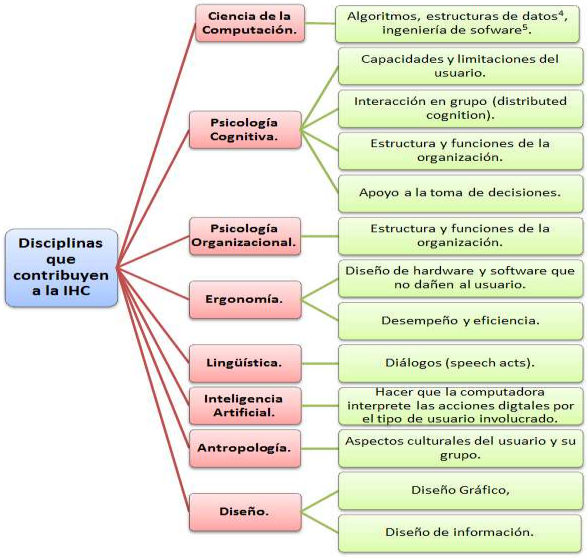
Por lo tanto, una interfaz tiene que estar diseñada con consideraciones ergonómicas para facilitar y optimizar su utilización. También debe poseer una combinación de tecnología, conocimiento y recursos que produzcan los resultados deseados a un costo razonable. Por otro lado, el sistema requiere un alto grado de confiabilidad para ser utilizado en aplicaciones reales. Si se satisfacen estas necesidades, el dispositivo interactivo será amigable, accesible y certero.

4.2 MARCO CONCEPTUAL

4.2.1 Interacción Humano Computador

La Interacción Humano Computador (IHC) [16] es un área de investigación multidisciplinaria (ver Figura 2) enfocada en las modalidades de interacción entre humanos y computadoras. La IHC investiga y trata todos los aspectos relacionados con el diseño y la implementación de las interfaces entre los humanos y las computadoras.

Figura 2. Disciplinas que contribuyen a la concepción, diseño e implementación de sistemas de IHC.



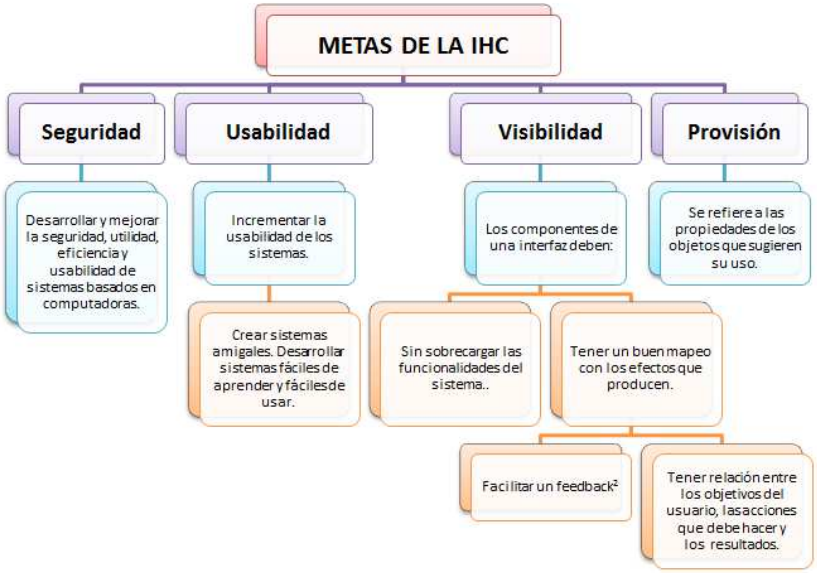
Fuente: Los autores con datos de Scopus, Science Direct, IEEE Xplorer

Esta disciplina que estudia el intercambio de información mediante software entre las personas y las computadoras se encarga del diseño, evaluación e implementación de los aparatos tecnológicos interactivos, estudiando el mayor número de casos que les pueda llegar a afectar. El objetivo es que el intercambio sea más eficiente: minimizar errores, incrementar la satisfacción, disminuir la frustración y, en definitiva, hacer más productivas las tareas que rodean a las personas y los computadores. En la Figura 3 se muestras las principales metas de la IHC.

Es muy importante diseñar sistemas que sean efectivos, eficientes y sencillos a la hora de utilizarlos, dado que la sociedad y en particular las personas con limitación de miembros superiores disfrutarán de estos avances. La dificultad viene dada por

una serie de restricciones que empiezan por el usuario mismo y terminan por la aplicación objetivo que desea manipularse.

Figura 3. Principales metas de los sistemas de IHC.



Fuente: Los autores con datos de J. A. Jacko and D. Wigdor, Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third. Amazon, 2012

Los componentes fundamentales de cualquier sistema de IHC [16], [18], [19] son:

- Usuario: Hay que tener en cuenta que el ser humano tiene una capacidad limitada de procesar información; lo cual es muy importante considerar al hacer el diseño. El ser humano se puede comunicar a través de cuatro canales de entrada/salida: visión, audición, tacto y movimiento. La información recibida se almacena en la memoria sensorial, la memoria a corto plazo y la memoria a largo plazo. Una vez se recibe la información, esta es procesada a través del razonamiento y de habilidades adquiridas, como por ejemplo el hecho de poder resolver problemas o el detectar errores. A todo este proceso afectará al estado emocional del usuario, dado que influye directamente sobre las capacidades de una persona. Además, un hecho que no se puede pasar por alto es que todos los usuarios tendrán habilidades comunes, pero habrá otras que variarán según la persona. En el caso de personas con limitación motriz de miembros superiores la mayor parte de información estará en los movimientos y gestos del rostro, así como se voz.

- Computador: El sistema final de interacción utilizado puede afectar de diferentes formas al usuario. Los dispositivos de entrada actuales permiten introducir texto, como sería el caso del teclado del computador, el teclado de un teléfono, el habla o bien un escrito a mano; dibujos; selecciones por pantalla, con el ratón, por ejemplo. Como dispositivos de salida se contraría con diversos tipos de pantallas, mayoritariamente aquellas que son de mapas de bits, pantallas de gran tamaño de uso en lugares públicos. Los sistemas de realidad virtual y de visualización con 3D juegan un rol muy importante en el mundo de la interactividad persona-computador. También serán importantes los dispositivos en contacto con el mundo físico, por ejemplo, controles físicos, como sensores de temperatura, movimiento, etc. El diseño del sistema IHC puede simplificarse mucho si se limita la aplicación objetivo. En el caso de personas de con limitación motriz las aplicaciones objetivo más importantes son aquellas que contribuyen con la inclusión social del mismos tales como redes sociales, chats, entre otras.

- Origen del proceso interactivo: Es importante que haya una buena comunicación entre usuario y computador, por este motivo la interfaz tiene que estar diseñada pensando en las necesidades del usuario y la aplicación objetivo. Es de vital importancia este buen entendimiento entre ambas partes dado que sino la interacción no será posible.

La compresión de los modelos mentales humanos es otro aspecto importante en la IHC. Los usuarios aprenden y guardan conocimiento y habilidades en formas diferentes, a menudo influenciados por su edad, así como por factores culturales y sociales del contexto. Por esto es importante conocer los componentes principales de los sistemas de IHC (ver Figura 4).

Los sistemas de IHC para personas con limitación motriz se han divido en dos grandes grupos: por voz y por gesto. Aun cuando los sistemas basados en voz son muy robustos para la inserción de texto y la generación de comandos directos para interactuar con elementos gráficos. No son muy efectivos o deseable cuando las personas tienen limitación del habla como es el caso de personas tetrapléjicas con traqueotomía. En esta propuesta se investiga los sistemas basados en gestos ya que para el caso planteado anteriormente pueden valerse de información de movimiento, de gestos y de información semántica de la aplicación objetivo para su concepción, diseño e implementación.

4.2.2 Interacción Humano Computador Gestual

La IHC gestual [18] es una rama de la IHC que junto con la lingüística y las ciencias de la computación tienen como objetivo interpretar gestos humanos a través de

algoritmos matemáticos. Los gestos pueden ser cualquier movimiento corporal o estado, pero comúnmente se originan a partir de la cara o la mano. Enfoques actuales en el campo incluyen reconocimiento de la emoción facial y el reconocimiento de gestos de la mano. Muchos enfoques que se han hecho hacen uso de cámaras y algoritmos para interpretar el lenguaje de señas. El reconocimiento de gestos puede ser visto como una manera para que las computadoras empiecen a entender el lenguaje corporal humano, construyendo así una relación más sólida entre máquinas y seres humanos. Dejando atrás sistemas primitivos como las interfaces de usuario de texto o incluso GUIs (interfaces gráficas de usuario), que aún limitan la mayoría de las entradas informáticas al teclado y el ratón. Dado que esta propuesta se enfoca en personas con limitación de miembros superiores, a continuación, se consideran las IHC basada en gestos faciales.

Figura 4. Componentes principales de los sistemas de IHC.

Fuente: Los autores

4.2.3 Reconocimiento de Gestos utilizando visión artificial

El reconocimiento de gestos [14] permite a seres humanos comunicarse con la máquina (HMI) e interactuar naturalmente sin dispositivos mecánicos. Utilizando el concepto de reconocimiento de gestos, es posible usar los dedos en un espacio libre para relacionar movimientos del cursor con el movimiento del usuario. Esto podría hacer que los dispositivos convencionales de entrada, tales como ratón, teclados e incluso pantallas táctiles sean redundantes. El Reconocimiento de gestos puede llevarse a cabo con técnicas de visión por computador utilizando cámaras o

webcams o con técnicas de procesado de señales utilizando dispositivos vestibles (wearable devices).

4.3 MARCO CONTEXTUAL

Este Proyecto se va llevar a cabo en la ciudad de Cali, será realizado por el semillero *Pionerosec* de la Fundación Universitaria Católica Lumen Gentium (Cali-Colombia), se tiene como fin realizar una interfaz de interacción humano-máquina la cual será manejada por medio de gestos faciales en un computador portátil, se tiene previsto realizar pruebas con sujetos experimentales quienes de forma voluntaria mediante Consentimiento Informado utilizarán la aplicación desarrollada para comandar WhatsApp.

5. METODOLOGÍA

5.1 TIPO DE ESTUDIO

Dado que el proyecto culminará con el desarrollo de un software que se ensayará con personas reales, el tipo de estudio propuesto es de tipo experimental. Básicamente este proyecto se apoyará en trabajos sistemáticos fundamentados en los conocimientos existentes obtenidos por la investigación en interfaces humano-computador y la visión artificial y se dirigirá a la fabricación de un prototipo software de interfaz humano-computador gestual que buscará mejorar lo actualmente reportado en la literatura científica.

5.2 METODO DE INVESTIGACIÓN

La investigación propuesta usará un enfoque científico-cuantitativo [20], específicamente del tipo analítico experimental. Básicamente consiste en apropiar y adaptar conocimiento sobre IHC y Visión por Computador para su aplicación en prototipos software de IHC gestual para personas con limitación motriz de miembros superiores. En la etapa inicial del proyecto de investigación se realizará una revisión bibliográfica en la que se identificarán estudios similares enmarcados en el desarrollo de IHC gestuales. Por un lado, se profundizará técnicas como la detección de rostros, estimación de pose 3D e identificación de elementos gráficos en aplicaciones de escritorio. El estudio general de estas técnicas permitirá no solo identificar la evolución de los algoritmos a través de la historia, sino también determinar la base para las técnicas que se propongan durante el desarrollo del proyecto. Principalmente se considerarán aquellas técnicas simples, y altamente robustas que puedan operar en tiempos adecuados para escenarios reales de interacción. Al final, los resultados del desempeño obtenido con el sistema propuesto en este proyecto se compararán con otros desarrollos alrededor del mundo bajo métricas convencionales. Si durante la ejecución del proyecto se identifican otras métricas, se incluirán en la evaluación del algoritmo propuesto.

5.3 FUENTES Y TÉCNICAS DE RECOLECCIÓN

Para el desarrollo del proyecto se empleará los siguientes dispositivos:

- Computador portátil: Procesador de 2.0 GHz o superior. Cuatro núcleos o superior recomendado. 8 GB de RAM recomendado, Espacio en disco duro: hasta 130 GB de espacio disponible y conexión a internet.
- Cámara web: En el proceso se utilizará la cámara integrada del computador.

A nivel de herramientas de software se utilizará:

- Sistema operativo Windows 10.
- Visual Studio Code 1.40.2
- Python versión 3.6.8
- Bibliotecas con licencia MIT para visión por computador y emulación de instrucciones de sistema operativo.
- WhatsApp versión de Escritorio.

5.4 TRATAMIENTO DE LA INFORMACIÓN

5.4.1 Técnicas estadísticas:

Para poner a prueba nuestra hipótesis se realizarán pruebas con usuarios utilizando la IHC desarrollada y se evaluará tanto el desempeño computacional como la efectividad en la generación de comando, así como la usabilidad del mismo. Para todas las evaluaciones se utilizarán estadísticas de primer orden para describir los resultados. Mediante test de hipótesis determinaremos la significancia estadística [21] en la obtención de algunos resultados.

5.4.2 Presentación de la información.

Se utilizarán tablas, diagramas de barras y tortas estadísticas [21] para representar y comunicar los resultados principales hallados en durante la evaluación de la prueba.

6. TÉCNICAS Y HERRAMIENTAS

La visión artificial o visión por computador [22] es una disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica para que puedan ser tratados por un ordenador. Tal y como los humanos usamos nuestros ojos y cerebros para comprender el mundo que nos rodea, la visión artificial trata de producir el mismo efecto para que los ordenadores puedan percibir y comprender una imagen o secuencia de imágenes y actuar según convenga en una determinada situación. Esta comprensión se consigue gracias a distintos campos como la geometría, la estadística, la física y otras disciplinas. La adquisición de los datos se consigue por varios medios como secuencias de imágenes, vistas desde varias cámaras de video o datos multidimensionales desde un escáner médico.

El objeto principal de estudio de la visión por computador son las imágenes. Por lo tanto, se debe partir de una definición matemática de la misma.

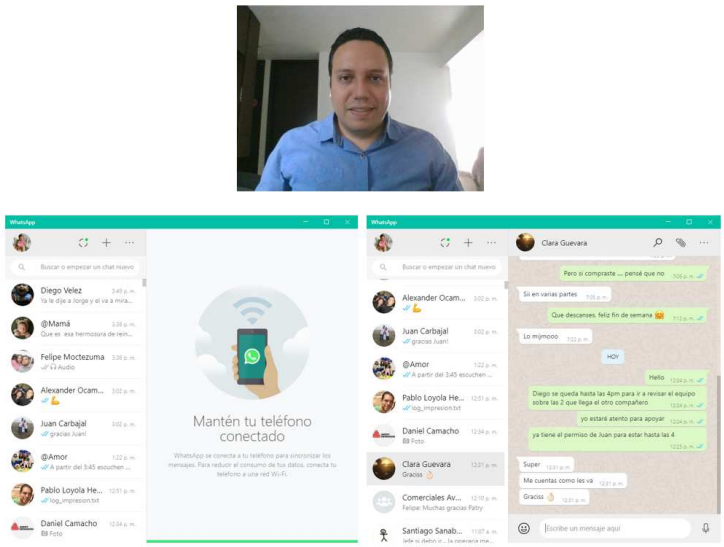
Definición de Imagen Digital [23]: Una imagen digital $I(x, y)$ se define como una matriz de dimensiones W x H (dónde W es el ancho de la imagen "número de columnas" y H es el alto de la imagen "número de filas", conteniendo en cada elemento de la matriz un valor discreto que cuantifica el nivel de información del correspondiente elemento, representado por un número finito de bits.

$$I(x,y).\ x\ e\ y\ \in Z \mid 0\ \leq x\ \leq W-1\ \leq y\ \leq H\ -1 \qquad (1)$$

Para el caso de imágenes con profundidad de 8 bit, los valores de los pixeles van de 0 a 255. Las imágenes a color se forman agrupando 3 matrices como las descritas anteriormente donde cada matriz codifica información de un canal dentro de un espacio de color, típicamente el espacio RGB por sus siglas en inglés de Rojo, Verde y Azul.

En la Figura 5 se observa dos ejemplos de imágenes provenientes de diferentes modalidades de captura: webcam y capturas de pantalla del computador.

Figura 5. Imagen RGB de rostro de 320x240 de resolución y 8 bit de profundidad, obtenida con cámara integrada al PC (arriba). Pantallazos (imágenes RGB de 1346x713 y 8 bit) de la aplicación WhatsApp versión de escritorio sin selección de chat (izquierdo-abajo), con chat seleccionado (derecho-abajo).



Muchas técnicas de análisis de imágenes no operan sobre la información de color sino sobre los niveles de brillo (tonos de gris) en la imagen, por esa razón suele convertirse la imagen de RGB a escala de grises mediante la siguiente expresión matemática:

$$Gray(x,y) = \frac{R(x,y) + G(x,y) + B(x,y)}{3} \quad (2)$$

Donde $Gray(x,y)$ es la versión de la imagen en escala de grises y $R(x,y)$, $G(x,y)$ y $B(x,y)$ son las componentes de color rojo, verde y azul de la imagen.
En este proyecto durante la solución propuesta se encontró un mejor desempeño operando directamente con la versión en escala de grises de las imágenes tanto adquiridas por webcam como las de captura de pantalla del escritorio del computador. Con el fin de mejorar la relación señal a ruido de las imágenes y buscando no perder información de bordes, se utilizó para el caso de la imagen adquirida por webcam un filtro no lineal tipo estadístico (filtro de mediana), el cual opera de la siguiente forma: el valor en la imagen de salida filtrada, $\mathrm{I_M(x,\,y)}$, se obtiene de la mediana de los valores que se encuentra en los ocho vecinos más cercanos (Kernel de 3x3) alrededor del pixel en la posición $\mathrm{(x,\,y)}$ de la imagen de entrada I.

6.1 CORRESPONDENCIA POR PLANTILLA DE IMAGEN

En el procesamiento digital de imágenes la correspondencia por plantilla (*template matching*) [24] se utiliza para encontrar pequeñas partes de una imagen que coincidan con una imagen de plantilla. Se puede utilizar en la fabricación como parte del control de calidad, una forma de navegar por un robot móvil, o como una forma de detectar objetos en las imágenes. La biblioteca PyAutogui [25] utiliza esta técnica para obtener la región de interés, ROI, que contiene la ventana de una aplicación en la captura de pantalla del escritorio del computador. En la Figura 6 se muestra la detección por plantilla de la ubicación y el tamaño de la ventana de la aplicación de Calculador de Windows 10.

Un método básico de correspondencia de plantillas utiliza un pedazo de imagen (plantilla), adaptado a una característica específica de la imagen de búsqueda, que se quiere detectar. Esta técnica se puede realizar fácilmente en imágenes grises o imágenes de borde. La salida de correlación cruzada será más alta en los lugares donde la estructura de la imagen coincide con la estructura de la máscara, donde los valores de imagen grandes se multiplican por los valores de máscara grandes.
Este método normalmente se implementa seleccionando primero una parte de la imagen de búsqueda para usar como modelo: denominaremos la imagen de búsqueda como $S(x, y)$, dónde (x, y) representa las ubicaciones de cada pixel en la imagen de búsqueda. Denominaremos a la imagen modelo cómo $\mathrm{T(xt,\ yt)}$, dónde $\mathrm{(xt,\ yt)}$ representa las ubicaciones de cada pixel en la imagen modelo. Luego, se desplaza el origen del modelo $\mathrm{T(xt,\ yt)}$ sobre cada punto $\mathrm{(x,\ y)}$ en la imagen de búsqueda y se calcula la suma de productos entre los coeficientes en $\mathrm{S(x,\ y)}$ y $\mathrm{T(xt,\ yt)}$ en toda el área contenida por imagen de búsqueda. Como se consideran todas las posiciones posibles de la imagen modelo con respecto a la imagen de búsqueda, la posición con la puntuación más alta es la mejor posición. Este método a veces se denomina filtrado espacial lineal y la imagen modelo se denomina máscara o filtro.
Una forma de lograr mejorar los resultados debido a los problemas de traslación en la imagen es comparar las intensidades de los píxeles usando la medida SAD (Suma de diferencias absolutas). Un píxel en la imagen de exploración con coordenadas $\mathrm{(xs,\ ys)}$ tiene magnitud $\mathrm{Is(xs,\ ys)}$ y un píxel en la imagen modelo con ubicaciones $\mathrm{(xt,\ yt)}$ tiene magnitud $\mathrm{It(xt,\ yt)}$. Por lo tanto, la diferencia absoluta en las intensidades de píxeles se define como:

$$D(x_s, y_s, x_t, y_t) = |I_s\ (x_s, y_s) - \ I_t\ (x_t, y_t)| \tag{3}$$

Ecuación de suma de diferencias absolutas:

$$SAD(x,y) = \sum_{i=0}^{H_T-1} \sum_{j=0}^{W_T-1} D(x+i, y+i, i, j) \tag{4}$$

Donde W_T y H_T son el ancho y altura de la imagen plantilla T. La representación matemática de la idea de recorrer los píxeles en la imagen de búsqueda a medida que se traslada el origen de la plantilla en cada píxel y tomamos la medida SAD es la siguiente:

$$\sum_{x=0}^{H_S-1} \sum_{y=0}^{W_S-1} SAD(x,y) \tag{5}$$

Donde W_S y H_S son el ancho y altura de la imagen de búsqueda S. En este método, la puntuación SAD más baja proporciona la estimación de la mejor posición de la plantilla dentro de la imagen de búsqueda. El método es simple de implementar y comprender, pero es uno de los más lentos.

En el pasado, este tipo de técnica normalmente solo se usaba en soluciones de hardware dedicadas debido a la complejidad computacional de la operación, sin embargo, se puede disminuir esta complejidad ejecutando esta operación en el dominio de frecuencia de la imagen, mediante el uso del teorema de convolución [22]. Otra forma de acelerar el proceso de correspondencia es mediante el uso de una pirámide de imágenes. Esta es una serie de imágenes, a diferentes escalas, que se forman filtrando y sub-muestreando repetidamente la imagen original para generar una secuencia de imágenes de resolución reducida. Estas imágenes de resolución más baja pueden buscarse para la plantilla (con una resolución similarmente reducida), con el fin de obtener posibles posiciones de inicio para buscar en escalas más grandes. Las imágenes más grandes se pueden buscar en una pequeña ventana alrededor de la posición de inicio para encontrar la mejor ubicación de la plantilla. Otros métodos pueden manejar problemas como la traslación, la escala, la rotación de imágenes e incluso todas las transformaciones afines.

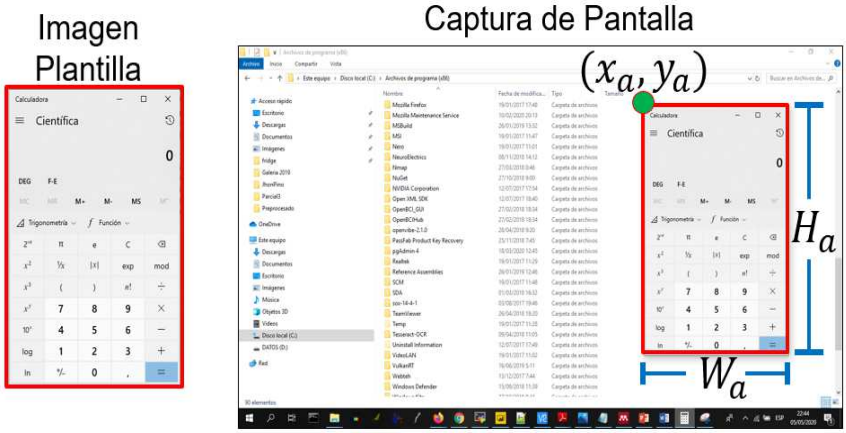
6.2 DETECCIÓN DE ROSTRO HUMANO

La detección de rostros [26], [27] (del inglés *face detection*) es un caso específico de la detección de objetos. La detección de caras por ordenador es un proceso por el cual el ordenador ubica los rostros presentes en una imagen o un vídeo. Generalmente la imagen se procesa en escala de grises y el resultado del algoritmo es similar al mostrado en la Figura 6. Correspondencia por plantilla de imagen de la aplicación Calculadora de Windows 10.

Fuente: Los autores

Figura 7.

Figura 6. Correspondencia por plantilla de imagen de la aplicación Calculadora de Windows 10.



Fuente: Los autores

Figura 7. Proceso de Detección de Rostro a partir de Imagen RGB.

Fuente: Los autores

Este proceso no es tan sencillo como lo haría el sistema visual humano. Según las condiciones en la que se encuentre la imagen durante el proceso de detección

puede suponer algunos problemas. En muchos casos la luminosidad no es la adecuada, aparecen elementos extraños, las caras están de perfil, se encuentran tapadas por algún elemento o por alguna otra cara o en un ángulo complicado. Actualmente existen varios métodos para detección de rostros.

- **Métodos basados en el conocimiento**: Estos métodos representan las técnicas de detección de caras que se basan en una serie de reglas previas definidas por la persona que quiere hacer la detección. Se definen una serie de características sobre las caras a detectar (forma de la cabeza, dos ojos, una nariz). Esto puede suponer un problema y es que, si estas reglas son muy generales, el resultado de una búsqueda en imágenes donde no hay caras, seguramente el resultado dirá que si hay caras y además una cantidad elevada. En el caso en que las reglas establecidas sean muy específicas posiblemente también aparezcan problemas ya que el resultado de la detección será muy bajo.

- **Métodos basados en características invariantes**: Estos métodos utilizan como punto de referencia el color de la piel y la textura, el problema que supone aplicar estos métodos es que si en la imagen aparece ruido o diferentes condiciones de iluminación el algoritmo aplicado no funcionará correctamente. Si se utiliza el color de la piel, los algoritmos que utilizan toda la gama de colores tienen mejor resultado que los que utilizan una escala de grises.

- **Métodos basados en plantillas**: Estos métodos modelan geométricamente la forma del objeto. Las plantillas son las componentes básicas como por ejemplo círculos, elipses. Una vez están definidas las plantillas se evalúa la correspondencia entre la cara y la plantilla. Las principales técnicas son las plantillas deformables y los contornos activos.

- **Métodos basados en apariencia**: Esta técnica en un principio no necesita el conocimiento de las características de la cara de la imagen que se quiere detectar. En los algoritmos utilizados en estos métodos aparecen los conceptos de entrenamiento y de aprendizaje. diferentes métodos para poder realizar la detección de caras por ordenador.

Fue hacia los años 70 cuando aparecieron los primeros algoritmos, basados en técnicas heurísticas y antropométricas [11], [27], pero no eran muy eficientes ya que fallaban bastante y eran muy sensibles a cambios. La investigación se dejó de lado porque todavía no tenía utilidad. Fue en los años 90 cuando, gracias al desarrollo de la tecnología y al descubrimiento de aplicaciones útiles, se reanudó la investigación.1

De los modelos anteriores, el que se utiliza actualmente son los métodos basados en el aspecto ya que son los que dan mejor resultados. Esto es debido a que en función de la variabilidad de la colección de imágenes o muestras con las que se realiza el entrenamiento obtendrán detectores con tasas altas de detección y bajas tasas de falso positivos. Además de una gran robustez, presentan una eficiencia en el sistema de detección y reducción del coste computacional.

En la actualidad uno de los algoritmos más representativos para la detección de objetos o más específicamente de rostros, es el algoritmo Viola & Jones, debido a sus múltiples factores como su bajo coste computacional, ahorro de tiempo considerable y la facilidad en la detección de objetos, se convierte para los programadores de hoy en día, como una de las herramientas más utilizadas. Por otro lado, a pesar de ser un algoritmo supremamente robusto, consta de dos partes fundamentales que facilitan su comprensión, uno de ellos el clasificador cascada que representa una probabilidad de verdaderos positivos del 99,9% y una probabilidad de falso positivos del 3,33%, esta probabilidad de verdaderos positivas es alta a comparación de otros algoritmos debido a que utiliza una representación de la imagen llamada imagen integral, donde se utilizan unos clasificadores en cascada que facilita la detección de cada subregión y así determina si son caras o no. Por otro lado, posee un entrenador de clasificadores basado en Adaboost [30], el cual se enfoca en los datos que fueron erróneamente clasificados y así obtener mejores tiempos en la entrega de información. Este detector se ha hecho muy popular debido a su rapidez a la hora de detectar las caras en imágenes y para su uso en la biblioteca OpenCV [28]. Como resultado de usar está técnica implementada en esta biblioteca de programación, en cada instante de tiempo se tiene un conjunto de ROI's asociadas a todos los rostros detectados en la imagen, esto es:

$$(\{(x, y, h, w)_i | x, y, w, h \in \mathbb{Z}; i = 1,2, \dots, NF; NF \text{ es el número de rostros detectados}\}).$$

6.3 SEGUIMIENTO

El problema de seguimiento de objetos en una escena es un problema ampliamente investigado en la comunidad de visión por computador. El propósito es lograr que en un video se pueda asociar la misma etiqueta al ROI que delimita uno o más objetos de interés. Existen algoritmos basados en correspondencia de características, filtrado de partículas y filtro Kalman [18] que han sido ampliamente utilizados en la literatura pero que tienen una considerable carga computacional. Para el caso de seguimiento de un solo objeto el problema puede simplificarse y lograr una mayor velocidad de ejecución entre frames (imágenes en un video)

consecutivos. Si se fija un ROI con etiqueta cero, su etiqueta puede asociarse al ROI que más cercano esté utilizando una distancia euclidiana para ello:

$$d_i(p_t, p_{t+1}^i) = \sqrt{p_t \cdot p_{t+1}^i} \tag{6}$$

Donde p_t es el centroide de coordenadas (x_t, y_t) del ROI de etiqueta cero en el instante de tiempo t y p_{t+1}^i es el centroide del i-ésimo ROI detectado en el instante de tiempo $t+1$. La etiqueta cero se pasará al i-ésimo ROI cuya distancia d_i sea la mínima entre las posibles.

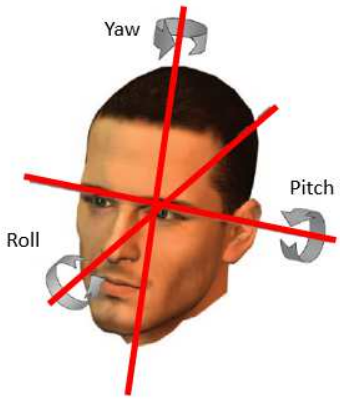
6.4 ESTIMACIÓN DE POSE DE LA CABEZA

Desde una temprana edad las personas muestran la capacidad de interpretar sin esfuerzo la orientación y el movimiento de una cabeza humana, lo que permite inferir las intenciones de otros que están cerca y comprender una importante forma no verbal de comunicación. La facilidad con la que el ser humano realiza esta tarea oculta la dificultad de un problema que ha desafiado los sistemas computacionales por décadas. En un contexto de visión por computador, la estimación de la postura o pose de la cabeza es el proceso de inferir la orientación de una cabeza humana a partir de una imagen digital. (Ver Figura 8). Esto requiere una serie de pasos de procesamiento para transformar una representación basada en píxeles de una cabeza en un alto nivel concepto de dirección. Al igual que otros pasos de procesamiento de imagen del rostro, un estimador ideal de la postura de la cabeza debe demostrar invariancia a una variedad de factores que cambian la imagen. Estos factores incluyen fenómenos físicos como distorsión de la cámara, proyectiva geometría, iluminación de fuentes múltiples no lambertianas [23], así como apariencia biológica, expresión facial y la presencia de accesorios como gafas y sombreros.

Es mucha la cantidad de información que se puede obtener a partir de los gestos y pose de la cabeza de una persona. Por ejemplo, los movimientos rápidos de la cabeza pueden ser un signo de sorpresa o alarma. En las personas, estos comúnmente desencadenan respuestas reflexivas de un observador, lo cual es muy difícil de ignorar incluso en presencia de estímulos auditivos contradictorios. Se pueden hacer otras observaciones importantes estableciendo el foco visual de

atención a partir de una estimación de pose de cabeza. Si dos personas enfocan su atención visual entre sí, a veces denominada mirada mutua, esto es a menudo una señal de que dos personas están involucradas en una discusión. La mirada mutua también se puede usar como un signo de conciencia, por ejemplo, un peatón esperará a que un conductor de automóvil detenido lo mire antes de entrar en un cruce de peatones. Observar la dirección de la cabeza de una persona también puede proporcionar información sobre el medio ambiente. Si una persona mueve su cabeza hacia una dirección específica, existe una alta probabilidad de que esté en la dirección de un objeto de interés. Los niños de tan solo seis meses explotan esta propiedad, conocida como seguimiento de la mirada, mirando hacia la línea de visión de un cuidador como un filtro para el medio ambiente.

Figura 8. Ángulos Roll, Pitch y Yaw para definir la orientación del rostro [29].



Fuente: Werner, F. Saxen, and A. Al-Hamadi, "Landmark Based Head Pose Estimation Benchmark And Method," Proc. - Int. Conf. Image Process. ICIP, vol. 2017-Septe, pp. 3909–3913, 2018, doi: 10.1109/ICIP.2017.8297015

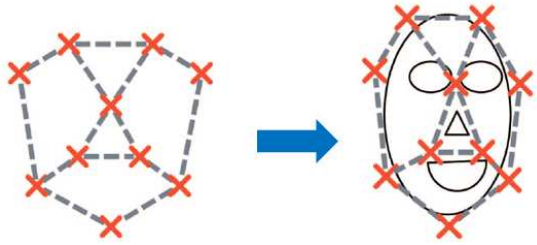
Al igual que el reconocimiento de voz, que ya se ha entrelazado en muchas tecnologías ampliamente disponibles tales como las interfaces hombre máquina, la estimación de la postura de la cabeza es también una herramienta comercial para cerrar la brecha entre los humanos y las computadoras. En la actualidad se logra diferencia en la literatura varias metodologías para abordar esta problemática:

- Los métodos de plantilla de similitud comparan una nueva imagen de una cabeza con un conjunto de ejemplos (cada uno etiquetado con una pose discreta) para encontrar la vista más similar.
- Los métodos de matriz de detectores entrenan una serie de detectores de cabeza cada uno sintonizado con una pose específica y asignan una pose discreta al detector con el mayor apoyo.
- Los métodos de regresión no lineal utilizan herramientas de regresión no lineal para desarrollar un mapeo funcional desde la imagen o los datos de características hasta una medición de pose de cabeza.
- Los métodos de inclusión múltiple buscan variedades de baja dimensión que modelen la variación continua en la postura de la cabeza. Se pueden incrustar nuevas imágenes en estas variedades y luego usarlas para la coincidencia o regresión de plantillas incrustadas.
- Los métodos geométricos utilizan la ubicación de características como los ojos, la boca y la punta de la nariz para determinar la postura a partir de su configuración relativa.

- Los métodos de seguimiento recuperan el cambio de pose global de la cabeza del movimiento observado entre fotogramas de video.
- Los métodos híbridos combinan uno o más de estos métodos antes mencionados para superar las limitaciones inherentes a cualquier enfoque individual.

Los métodos anteriores han considerado la estimación de la postura de la cabeza como un problema de detección de señal, asignando una región rectangular de píxeles de imagen a una orientación de postura específica. Los modelos flexibles [29] adoptan un enfoque diferente. Con estas técnicas, un modelo no rígido se ajusta a la imagen de manera que se ajuste a la estructura facial de cada individuo. Además de las etiquetas de pose, estos métodos requieren datos de entrenamiento con rasgos faciales anotados que les permite hacer comparaciones a nivel de rasgo en lugar de a nivel de apariencia global. En la Figura 9 se presenta una ilustración conceptual.

Figura 9. Los modelos flexibles se ajustan a la estructura facial del individuo en el plano de la imagen. La postura de la cabeza se estima a partir de comparaciones a nivel de características o de la instanciación de los parámetros del modelo.



Fuente: P. Werner, F. Saxen, and A. Al-Hamadi, “Landmark Based Head Pose Estimation Benchmark And Method,” Proc. - Int. Conf. Image Process. ICIP, vol. 2017-Septe, pp. 3909–3913, 2018, doi: 10.1109/ICIP.2017.8297015

Para estimar la pose con los métodos de la plantilla de similitud la vista de un nuevo cabezal se superpone en cada plantilla y se utiliza una métrica basada en píxeles para comparar las imágenes. Sin embargo, incluso con un registro perfecto, las imágenes de dos personas diferentes no se alinearán exactamente, ya que la ubicación de los rasgos faciales varía entre las personas. Ahora, considere una plantilla basada en un gráfico deformable de puntos característicos locales (esquinas de los ojos, nariz, esquinas de la boca, etc.), para entrenar este sistema, las ubicaciones de los rasgos faciales se etiquetan manualmente en cada imagen de entrenamiento, y los descriptores de rasgos locales como los de Gabor [22] se pueden extraer en cada lugar. Estas características se pueden extraer de las vistas de varias personas, y se puede lograr una invariancia adicional almacenando un grupo de descriptores en cada nodo. Esta representación se ha denominado Grafo

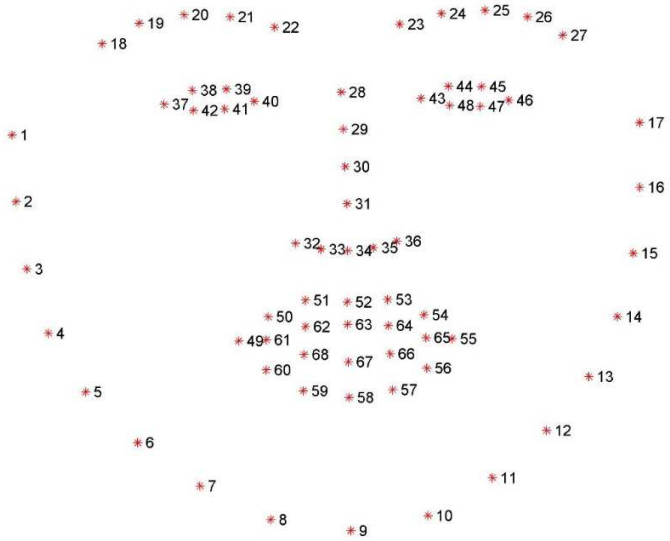
de Grupo Elástico [29] y posee la capacidad de representar objetos no rígidos o deformables.

Para comparar un grafo de grupo con una nueva imagen de la cara, el grafo se coloca sobre la imagen y se deforma de forma exhaustiva o iterativa para encontrar la distancia mínima entre las entidades en cada ubicación de los nodos del grafo. Este proceso se llama Correspondencia de Grafo Elástica (por sus siglas en inglés EGM). Para la estimación de la postura de la cabeza, se crea un grafo de grupo diferente para cada postura discreta, y cada una de estas se compara con una nueva vista de la cabeza. El grafo de grupo con la máxima similitud asigna una pose discreta a la cabeza. Debido a que EGM utiliza características ubicadas en puntos faciales específicos, hay significativamente menos variabilidad entre sujetos que con puntos no alineados. Esto hace que sea mucho más probable que la similitud entre los modelos equivalga a la similitud en la pose. Una desventaja de este método es que la estimación de la postura es discreta, lo que requiere muchos grafos de grupo para obtener estimaciones finas de la postura de la cabeza. Desafortunadamente, comparar muchos grafos de grupo, cada uno con muchas deformaciones, es computacionalmente costoso en comparación con la mayoría de las otras técnicas de estimación de la postura de la cabeza.

En este proyecto se utiliza la biblioteca DLIB [30] el cual trabaja con un modelo flexible de 68 puntos característicos, (ver Figura 10).

A partir de este modelo es posible definir algunas relaciones de aspectos que pueden contribuir a determinar por ejemplo si se ha producido un guiño del ojo o una apertura de boca. Para el caso de la boca, el coeficiente de relación de aspecto se denomina MAR (*Mouth Aspect Ratio*) y está definido a partir de los puntos 61, 62, 63, 64, 65, 66, 67 y 68, ver Figura 10. Sesenta y ocho puntos característicos

faciales del modelo flexible utilizado en la biblioteca DLIB.

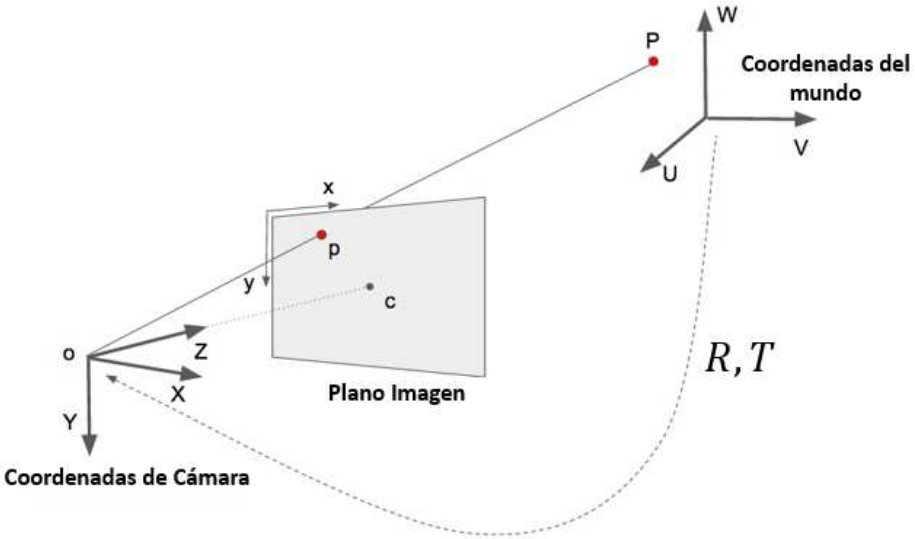


Fuente: X. Ren, J. Ding, J. Sun, and Q. Sui, “Face Modeling Process Based On Dlib,” in 2017 Chinese Automation Congress (CAC), 2017, pp. 1969–1972, doi: 10.1109/CAC.2017.8243093

Figura 11, mediante la siguiente ecuación:

$$MAR = \frac{\|p_{61} - p_{68}\| + \|p_{63} - p_{67}\| + \|p_{64} - p_{66}\|}{2\|p_{65} - p_{61}\|} \quad (7)$$

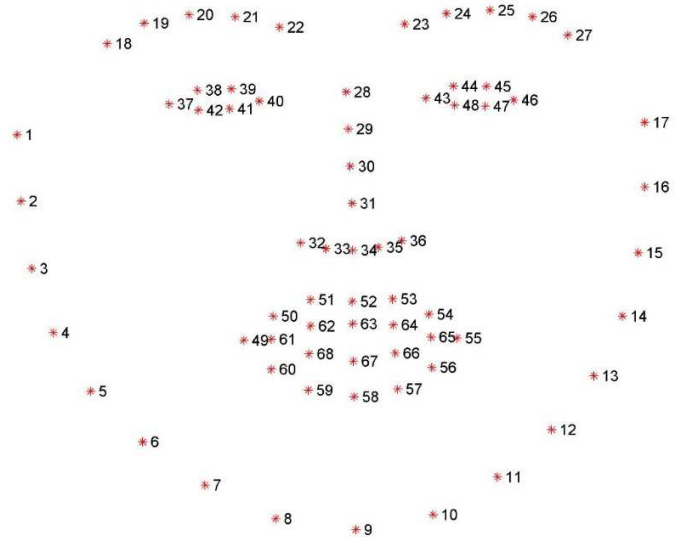
Ahora bien, estimar la pose de un objeto 3D significa encontrar 6 números: tres para la traslación y tres para la rotación. La rotación también puede expresar mediante una matriz de nueve parámetros. Existen varios algoritmos para la estimación de pose. El primer algoritmo conocido se remonta a 1841. La idea general se concibe teniendo en cuenta el modelo pin-hole [22], (ver Figura 12). Existen tres sistemas de coordenadas involucrados. Las coordenadas 3D de las diversas características faciales que se muestran en la Figura 12. Modelo pin-hole para proyección de puntos 3D en la imagen de una cámara.



Fuente: R. Szeliski, Computer Vision, vol. 42. London: Springer London, 2011

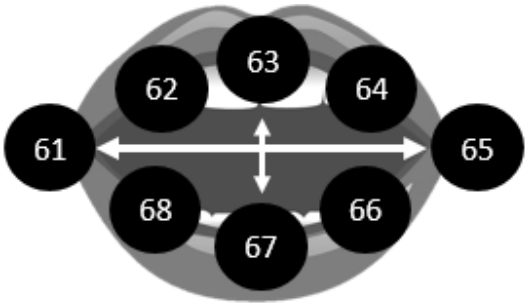
Figura 13 están en coordenadas del mundo. Si se supiera la rotación y la traslación (es decir, la pose), (R, T), se puede transformar los puntos 3D en coordenadas del mundo (U, V, W) en puntos 3D en coordenadas de cámara (X, Y, Z). Los puntos 3D en las coordenadas de la cámara se pueden proyectar en el plano de la imagen (es decir, el sistema de coordenadas de la imagen (x, y)) utilizando los parámetros intrínsecos de la cámara (distancia focal, centro óptico, etc.).

Figura 10. Sesenta y ocho puntos característicos faciales del modelo flexible utilizado en la biblioteca DLIB.



Fuente: X. Ren, J. Ding, J. Sun, and Q. Sui, “Face Modeling Process Based On Dlib,” in 2017 Chinese Automation Congress (CAC), 2017, pp. 1969–1972, doi: 10.1109/CAC.2017.8243093

Figura 11. Puntos de interés para definir la relación de aspecto de la boca con base en el modelo flexible de la biblioteca DLIB.



Fuente: X. Ren, J. Ding, J. Sun, and Q. Sui, “Face Modeling Process Based On Dlib,” in 2017 Chinese Automation Congress (CAC), 2017, pp. 1969–1972, doi: 10.1109/CAC.2017.8243093

$$R = \begin{pmatrix} \cos(\alpha)\cos(\beta) & \cos(\alpha)\sin(\beta)\sin(\gamma) - \sin(\alpha)\cos(\gamma) & \cos(\alpha)\sin(\beta)\cos(\gamma) - \sin(\alpha)\sin(\gamma) \\ \sin(\alpha)\cos(\beta) & \sin(\alpha)\sin(\beta)\sin(\gamma) + \cos(\alpha)\cos(\gamma) & \sin(\alpha)\sin(\beta)\cos(\gamma) - \cos(\alpha)\sin(\gamma) \\ -\sin(\beta) & \cos(\beta)\sin(\gamma) & \cos(\beta)\cos(\gamma) \end{pmatrix} \quad (8)$$

$$T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (9)$$

R y T en las ecuaciones 8 y 9 son matriz de rotación y vector de traslación respectivamente, asociados a la pose del rostro. Los coeficientes r_{ij} de R vienen relacionados con los ángulos yaw (α), pitch (β) y roll (γ) mediante la ecuación 8. (t_x), (t_y) y (t_z) son las componentes del vector T a lo largo de los ejes X, Y y Z. En ausencia de distorsión radial, las coordenadas (x, y) del punto p en las coordenadas de la imagen están dadas por:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (10)$$

Donde f_x, f_y, c_x, y c_y son las distancias focal y centro óptico en los ejes X e Y. La relación entre los puntos del mundo real a las coordenadas de la cámara viene dado por:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{pmatrix} \begin{bmatrix} U \\ V \\ W \\ 1 \end{bmatrix} \quad (11)$$

Utilizando un conjunto conocido de puntos 2D en la imagen es posible proponer una función de costo asociada al error de reproyección asociada al modelo pin-hole y estimar los valores de asociados a la pose minimizando dicha función de costo. Una forma de obtener estos valores minimo es mediante el algoritmo de optimización de Levenberg-marquardt [31], [32]. En openCV [33] es posible obtener los parámetros de pose utilizando la funcion *solvePnP*.

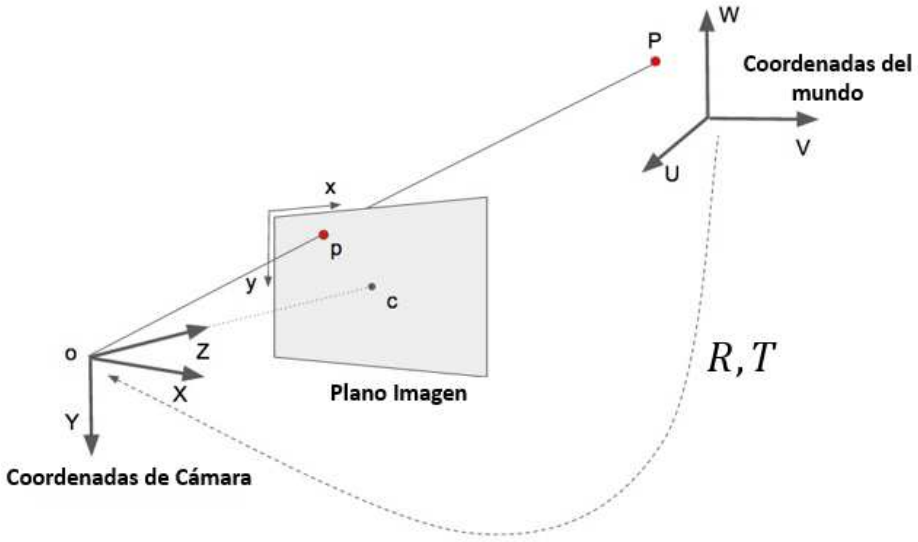
Una vez hallado los parámetros de pose, para determinar los parámetros roll, pitch y yaw directamente desde la matriz de rotación se utilizan las siguientes expresiones:

$$roll = tan^{-1}\left(\frac{r_{21}}{r_{11}}\right) \tag{12}$$

$$pitch = tan^{-1}\left(\frac{-r_{31}}{\sqrt{(r_{32})^2 + (r_{33})^2}}\right) \tag{13}$$

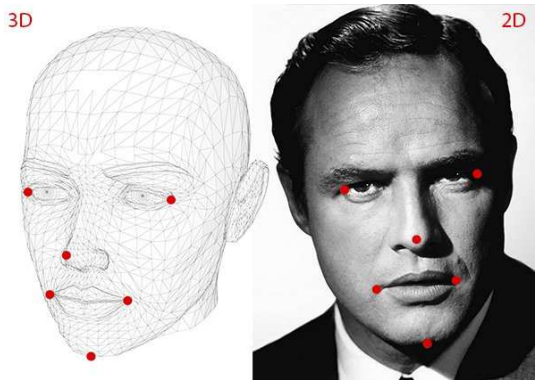
$$yaw = tan^{-1}\left(\frac{r_{32}}{r_{33}}\right) \tag{14}$$

Figura 12. Modelo pin-hole para proyección de puntos 3D en la imagen de una cámara.



Fuente: R. Szeliski, Computer Vision, vol. 42. London: Springer London, 2011

Figura 13. Puntos 3D en el espacio y sus correspondientes puntos 2D en el plano imagen de un rostro.



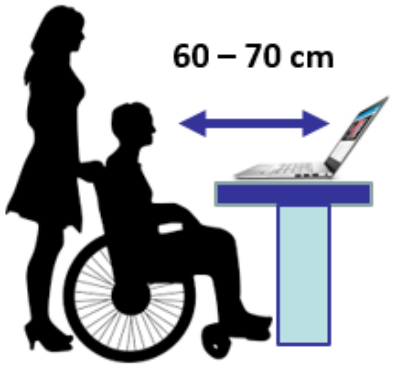
Fuente: R. Szeliski, Computer Vision, vol. 42. London: Springer London, 2011

7. INTERFAZ GESTUAL PARA WHATSAPP

Para concebir la solución de interfaz gestual primero se definió el software WhatsApp de Escritorio como aplicación objetivo para comandar, esto con base en una encuesta preliminar sobre la preferencia de aplicación de las personas con limitaciones motrices para comunicarse con el mundo exterior. Las condiciones de interacción se establecen en la Figura 14. Para tener una buena composición en la imagen capturada mediante la cámara integrada en un laptop sobre una mesa, se define una distancia promedio entre 60 a 70 cm medidos desde la pantalla del computador, esta debe estar inclinada no más de 20° respecto a la normal de la superficie de apoyo del computador.

La primera opción para interactuar con WhatsApp es la de controlar la posición del cursor con ayuda de los movimientos del rostro dado que estos pueden ser estimados utilizando la técnica de detección rostro descrita en la sección 6.3; sin embargo, esto puede resultar tedioso para una persona con limitación motriz ya que debe estar moviendo constantemente la cabeza para seleccionar con el cursor las zonas de interés.

Figura 14. Condiciones de disposición geométrica para interacción con la interfaz gestual.



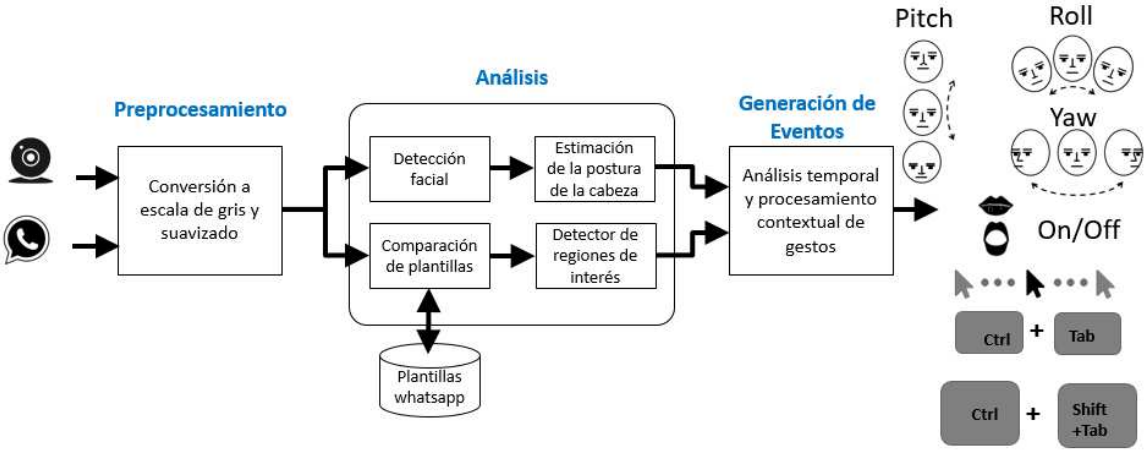
Fuente: Los autores con imagen vectorial en http://centromujer.republica.com/salud/quien-cuida-al-cuidador.html

Para hacer más fácil la interacción se propone realizar desplazamientos específicos del cursor aprovechando que en la aplicación WhatsApp existen dos zonas de interacción claramente diferenciadas: zona de contactos, y la zona de chat (o conversación). Por esta razón se propone no solamente procesar la imagen del rostro capturada mediante una webcam sino también la captura de pantalla del computador donde la persona está visualizando WhatsApp.

Figura 15. Diagrama de bloques principales de la solución propuesta.

Fuente:Los autores

Figura 16. Técnicas implementadas en cada una de las etapas de la solución propuesta.



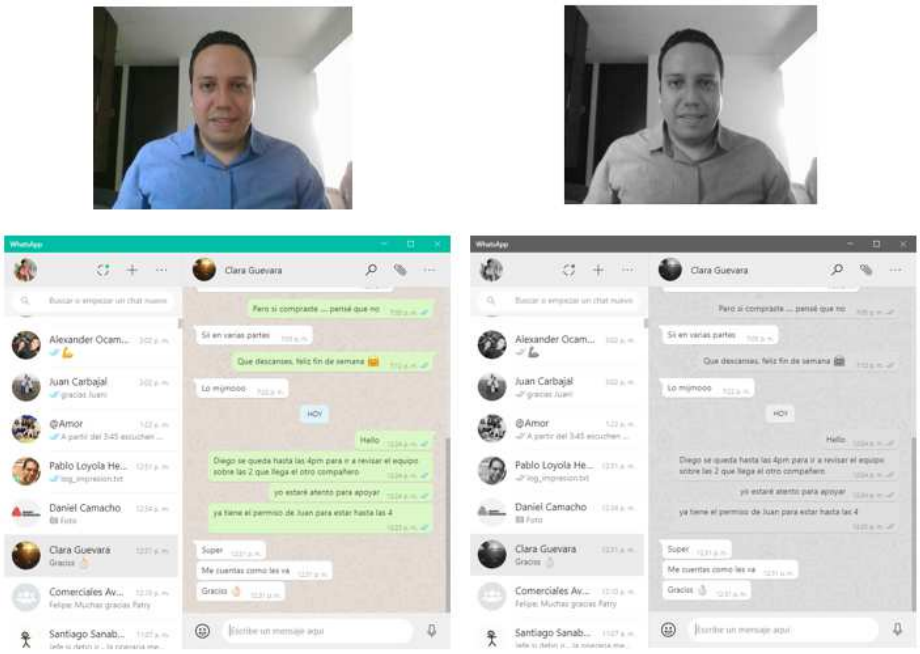
Fuente: Los autores con https://creately.com/es/home/

Adicionalmente se propone como comandos de interacción: i) activar o desactivar la interfaz gestual con la apertura de la boca, ii) ubicar el cursor en la zona de conversación o en la zona de contactos mediante movimientos laterales y sostenidos del rostro, iii) hacer scroll up/down en cualquiera de estas zonas mediante movimientos verticales y sostenidos del rostro. El efecto de hacer scroll up/down en la zona de conversación es la de permitir subir y bajar en la conversación que se ha desarrollado entre dos personas o en un grupo personas. En caso de que el scroll up/down se realice en la zona de contactos permitirá seleccionar un contacto a la vez, esto implica emular 'Ctrl+TAB' o 'Ctrl+Shift+TAB'

cada que vez que el cursor se desplaza de forma discreta entre los diferentes contactos.

La solución propuesta contempla tres etapas: preprocesamiento, análisis y generación de eventos, (ver Figura 15). Las técnicas utilizadas en cada una de estas etapas se pueden ver en detalle en la Figura 16.

Figura 17. Conversión a escala de grises (derecha) de las imágenes de entrada (izquierda).



Fuente: Los autores

7.1 PRE-PROCESAMIENTO

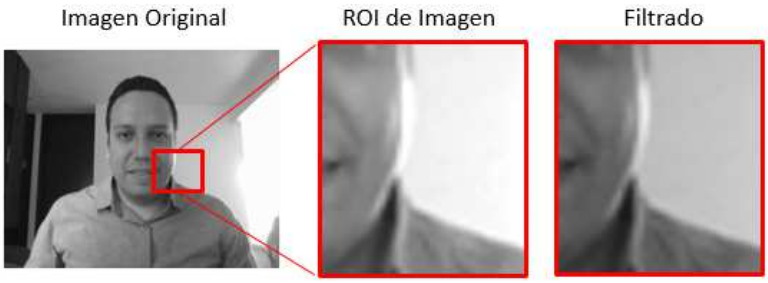
La etapa de pre-procesamiento tiene dos propósitos fundamentales: reducir la dimensionalidad de los datos y mejorar la relación señal a ruido de las imágenes. La reducción de dimensionalidad consiste en pasar de imágenes de tres canales, RGB, a imágenes de un solo canal en escala de grises. Esto se realiza tanto en la imagen adquirida por webcam y la captura de pantalla de la aplicación WhatsApp. (Ver Figura 17). La mejora de la relación señal a ruido se lleva a cabo con la aplicación un filtrado no lineal tipo mediana estadística el cual conserva los bordes únicamente en la imagen adquirida por la webcam. (Ver Figura 18). El propósito es mejorar la tasa de detección de rostros principalmente cuando las condiciones de iluminación se reducen. Se probaron diferentes tamaños de kernel (3x3, 5x5, 7x7 y 9x9) para el filtro de mediana. Para cada uno se estimó que el tiempo promedio de

ejecución por imagen son los siguientes: 3x3 (151 ms), 5x5 (185 ms), 7x7 (243 ms) y 9x9 (272 ms). Se seleccionó finalmente el de tamaño 3x3 porque se obtuvo la misma efectividad en la detección de rostros comparándola con los demás tamaños y a que su velocidad de ejecución es la más rápida comparada con el resto.

7.2 ANÁLISIS

En esta etapa están las técnicas que permiten detectar el rostro y su pose, así como el detector de regiones de interés para la aplicación WhatsApp. En la Figura 19 se muestra el resultado de detección y estimación de pose obtenida con las técnicas descritas en las secciones 6.2 y 6.4.

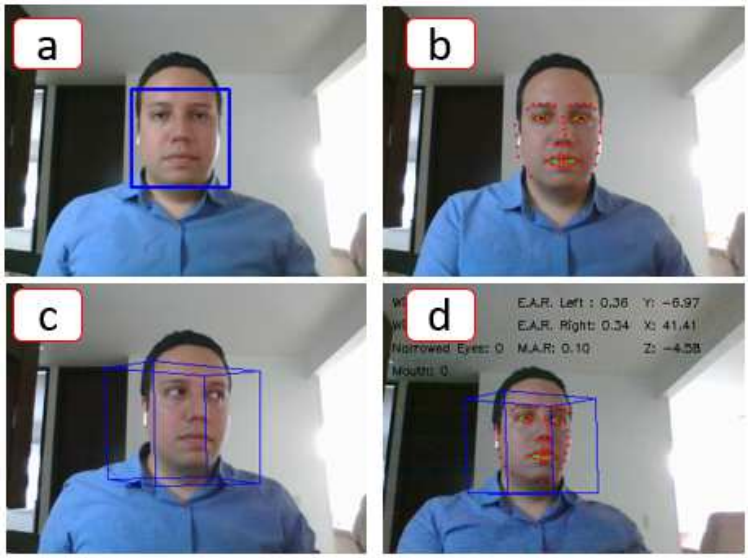
Figura 18. Efecto de aplicación de filtrado de mediana.



Fuente: Los autores

Figura 19. Etapas del proceso de análisis de imagen adquirida por webcam: a) detección de rostro, b) detección y correspondencia de puntos de interés del rostro, c) estimación de pose y d) fusión de resultados de detección de rostro, puntos característicos y estimación de pose.

Figura 19 Etapas del proceso de análisis de la imagen



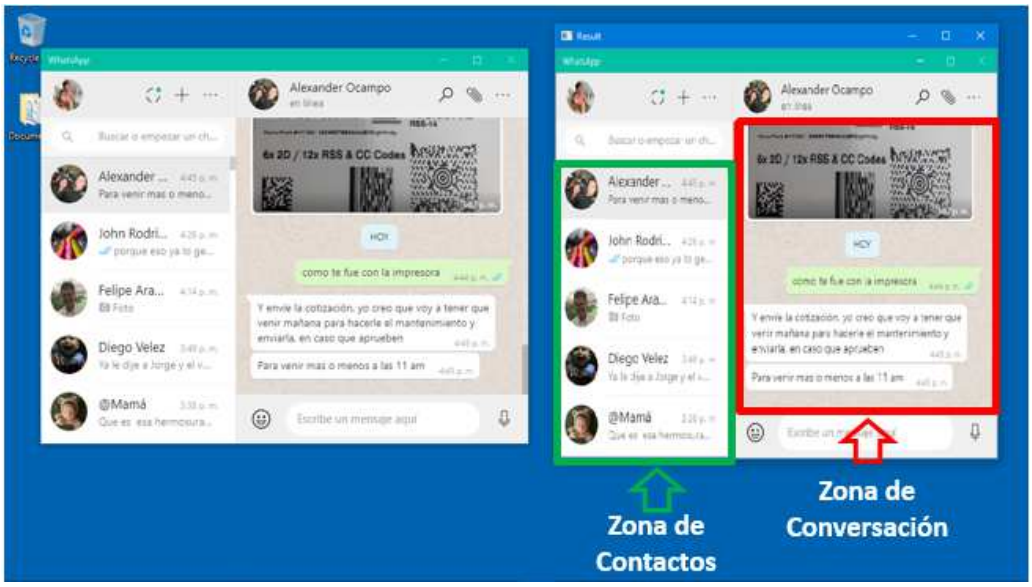
Fuente: Los autores

En este módulo se incluye también la técnica de seguimiento descrita en la sección 6.3 seguir el rostro y las zonas de interacción de un frame al siguiente durante la captura constante de imágenes.

La detección de zonas de interés se realiza primero obteniendo la región de interés de toda la aplicación de WhatsApp en la captura de pantalla del escritorio del computador, para esto se utiliza la técnica de correspondencia de plantillas descrita en la sección 6.1. En la Figura 20 se observa el resultado de este proceso.

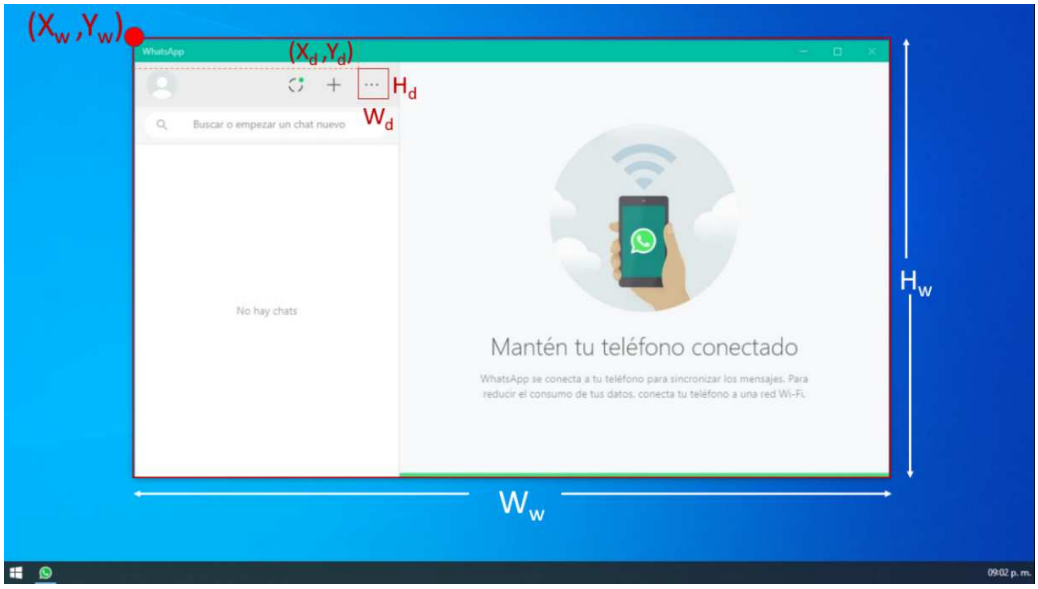
Para localizar las zonas de interacción, primero se obtienen las coordenadas de la ventana de WhatsApp, la esquina superior de la ventana (X_w, Y_w) y su ancho y alto (W_w, H_w). Estos parámetros son el punto de partida para tener acotado el espacio de detección. Después de analizar los elementos dentro de la ventana, se encontró que el botón de menú podía ser el punto de referencia para separar la ventana en dos zonas y así delimitar las regiones donde se encuentran las zonas de contacto y de conversación. Para encontrar los datos de la posición del botón de menú toma una muestra del mismo y utilizando la correspondencia por plantillas se obtiene su posición con respecto a la pantalla completa, esto se referenciará al origen de coordenadas de la ventana de WhatsApp para ubicar su posición dentro de la misma, a saber (X_d, Y_d, W_d, H_d) (ver Figura 21).

Figura 20. Detección de zonas de interés en aplicación de WhatsApp.



Fuente: Los autores

Figura 21. Detección de ROI de la ventana de WhatsApp y del ROI asociado al botón menú.



Fuente: Los autores

Con base en los datos del botón menú, se logra separar la ventana de WhatsApp en dos zonas, esto se logra a partir del cálculo de la posición de la esquina superior derecha del icono del menú, la cual se obtiene mediante la siguiente expresión:

$$E_{tr} = Y_d + W_d \tag{15}$$

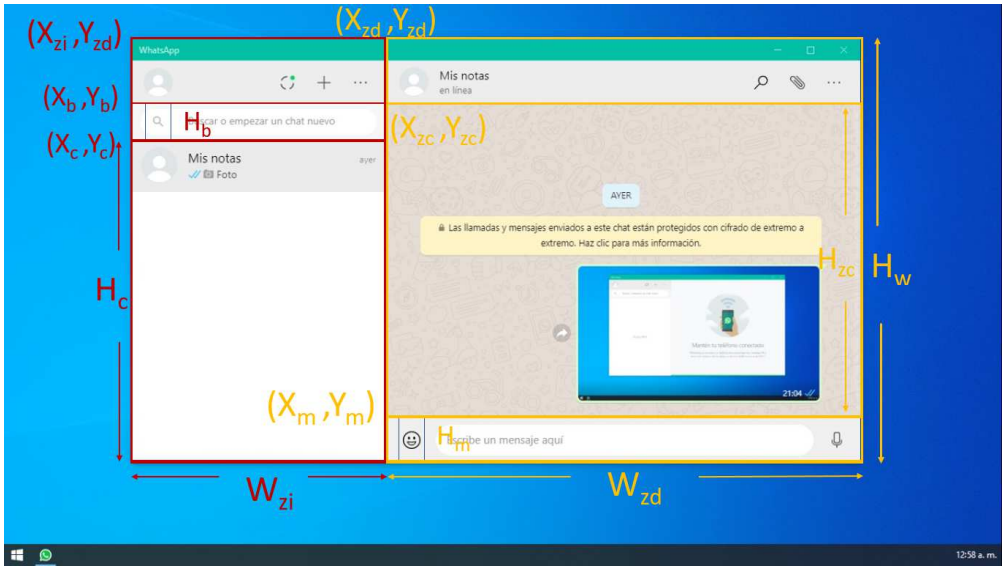
A este valor se le añaden 16 pixeles que es el valor que le falta a la coordenada del ícono para para encontrar la coordenada en X de la zona Derecha (X_{zd}), y el borde derecho de la zona izquierda, de esta forma se tiene que ($X_{zd} = E_{tr} + 16$). Lo anterior permite calcular el ancho de ambas zonas (W_{zi} y W_{zd}) de la siguiente forma: ($W_{zi} = X_{zd} - X_w$) y ($W_{zd} = W_w - X_{zd}$). (Ver Figura 22).

Una vez separadas las dos zonas se procede a identificar dos elementos más en la aplicación de WhatsApp para obtener los ROI's finales asociados a la zona de contactos y la zona de interacción. Utilizando nuevamente correspondencia de plantillas se localiza el icono de búsqueda de contactos y con esto se logra delimitar mejora la zona de contactos. Para la zona derecha se localiza el icono de emoticones y seguidamente la zona de conversaciones. Si el ROI del icono de búsqueda es (X_b, Y_b, W_b, H_b) y el del icono de emoticón es (X_m, Y_m, W_m, H_m), se puede estimar la posición de las otras dos zonas de interacción de la siguiente forma. Primero, para la zona de contactos se calcula la coordenada Y_c sumando del área de búsqueda la coordenada Y_b más la altura del área H_b y para la altura de la zona de contactos H_c se usa la altura de la ventana menos la recién calculada coordenada Y_c. Con esto se completa los parámetros de la zona de contactos ya que la posición en X (X_c) es igual X_{zi} y el ancho es igual a W_{zi}. Por otro lado, para la zona de conversación se empieza tomando las coordenadas ya conocidas e igualándolas con las que se necesita ya que las coordenadas en X e Y se igualan de la siguiente forma, X_{zc} será igual a X_{zd}, Y_{zc} será igual a Y_b y W_{zc} será igual a W_{zd}. Para calcular la altura habrá que tomar Y_m y restarle Y_{zc}, de esta forma ($H_{cz} = Y_m - Y_{zc}$), y de esa forma se obtienen las coordenadas de la zona de conversación.

Figura 22. Primera división de la ventana de WhatsApp utilizando criterios geométricos.

Fuente: Los autores

Figura 23. Identificando nuevas plantillas para mejorar la estimación de las zonas de contacto y de conversación.



Fuente: Los autores

7.3 GENERACIÓN DE EVENTOS

En esta etapa se utilizan las señales de roll, pitch y yaw, así como los ROI's de interacción en WhatsApp detectados y la relación de aspecto de la boca (MAR) por sus siglas en inglés *Mouth Aspect Ratio*, seguidos en el bloque de análisis.

Se determinó de forma heurística que la mejor forma de posicionar el cursor en el centro del ROI de la zona de conversación debe ocurrir cuando el usuario realice un movimiento en yaw mayor de 15 °. Para posicionar el cursor en el centro del ROI de la zona de contactos, el usuario deberá hacer un movimiento en yaw menor de -6 °. En caso que el cursor se encuentre en la zona de conversación, con un movimiento en pitch mayor de -5 ° se interpretará como scroll up y un movimiento en pitch menor de -12 ° será interpretado como scroll down. En caso que el cursor se encuentre en la zona de contactos, se emulará en Ctrl+TAB (siguiente abajo) o Ctrl+Shift+TAB (siguiente arriba) para activar la conversación con el contacto inmediatamente superior o inferior respectivamente.

Para detectar la boca abierta. Se estableció de forma heurística un umbral del MAR para la boca cerrada de 0.6, al abrir la boca este valor aumentará. Se determinó que si el usuario permanece con la boca abierta con el umbral superior a 0.6 y durante 3 frames consecutivos, esto se interpretará como una orden para desactivar (cerrar) la interfaz gestual.

7.4 IMPLEMENTACIÓN

La implementación de los diferentes módulos se realizó utilizando el lenguaje Python 3.6.8, el código fuente quedó disponible en: https://github.com/cdfbdex/hciVisualGesture bajo licencia MIT y se creó tanto un instalador del aplicativo como un manual de usuario para su operación.

8. PRUEBAS EXPERIMENTALES Y RESULTADOS

Para evaluar el desempeño, robustez y usabilidad de la interfaz de usuario primero se realizaron pruebas de eficiencia computacional con la interfaz gestual en tres equipos de cómputo de baja (Intel Pentium, 4 GB RAM), media (Intel Core i5, 8 GB RAM) y alta gama (Core i7, 16 GB RAM). Luego se estudió el comportamiento de la interfaz frente a cambios de iluminación, presencia de otros usuarios y casos de oclusión tanto del rostro como en la aplicación de WhatsApp. Finalmente, en un escenario controlado de pruebas se solicitó a diez personas que usaran la interfaz durante cinco minutos ejecutando los comandos que fueron previamente enseñados. Durante los dos primeros minutos se instruyó a los sujetos experimentales respecto a los comandos básicos que la interfaz soporta y seguidamente se dejó que los usuarios activaran/desactivaran la interfaz gestual, seleccionaran contactos y navegaran en sus conversaciones. En todos los casos mediante consentimiento informado (ver ANEXO) se les pidió a los sujetos que abrieran su propia cuenta de WhatsApp para que la navegación fuera más familiar para los mismos.

8.1 DESEMPEÑO COMPUTACIONAL

Para determinar si la velocidad de ejecución de la interfaz gestual es independiente de la gama del computador que puede utilizar eventualmente una persona con limitación motriz, se seleccionaron tres equipos con características de baja, media y alta gama. Además, se estudió si la velocidad de ejecución de la interfaz es igual ejecutando solo WhatsApp (además del aplicativo de interfaz gestual) y WhatsApp junto con tres aplicaciones más (Microsoft Word, Mozilla Firefox, Reproductor de Windows Media Player).

Se guarda el tiempo ejecución de cada par de imágenes (webcam y captura de pantalla) que pasa por el flujo de ejecución (pre-procesado, análisis, generación de eventos) de la interfaz gestual. Este proceso se realiza 525 veces en cada uno de los escenarios descritos arriba, a saber, WhatsApp ejecutándose solo y con tres aplicaciones más. Los valores promedios para cada uno de los tipos de computadores se muestran en la Tabla 1. Se observa que la rapidez de ejecución de la interfaz gestual es aproximadamente 2 segundos más rápida en el computador de media y alta gama comparado con el de baja gama. Se encontró también que en cualquier tipo de computador la interfaz gestual requiere 1.5 GB en RAM para su ejecución y aproximadamente un 35 % de porcentaje de ocupación de la CPU en los computadores de media y alta gama y de 65 % en el computador de baja gama.

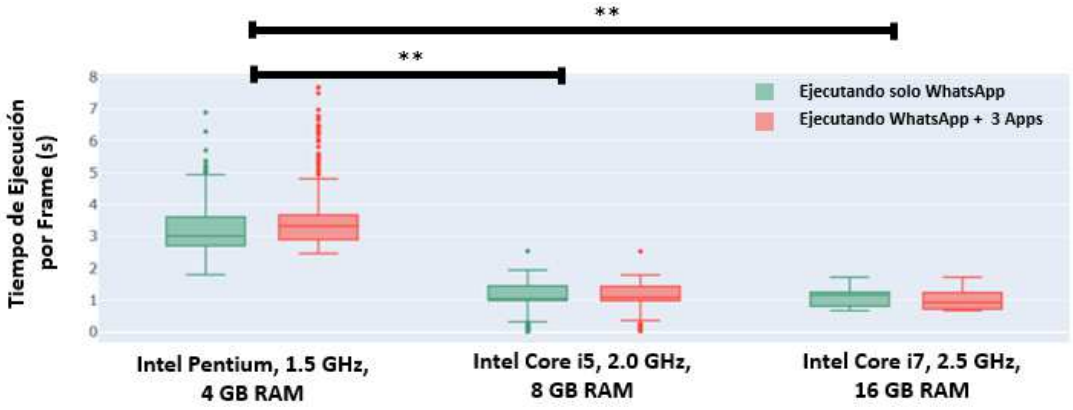
Tabla 1. Tiempo promedio y desviación estándar de ejecución de interfaz gestual.

Características del PC	**Tiempo promedio y desviación estándar de Ejecución por frame cuando solo se ejecuta WhatsApp, (s).**	**Tiempo promedio y desviación estándar de Ejecución por frame cuando se ejecuta WhatsApp y 3 aplicaciones más, (s).**
Intel Pentium, 1.5 GHz, 4 GB RAM (baja gama)	3.23 (±0.73)	3.52 (±0.98)
Intel Core i5, 2.0 GHz, 8 GB RAM (media gama)	1.14 (±0.41)	1.17 (±0.38)
Intel Core i7, 2.5 GHz, 1 6 GB RAM (alta gama)	1.05 (±0.25)	1.00 (±0.26)

Fuente: Los autores

Para determinar si existe una diferencia estadísticamente significativa entre los valores promedios de ejecución por frame, se comprobó primero la normalidad de los datos mediante una prueba de t-Student [21] y se procedió a realizar una prueba de hipótesis sobre los valores promedios de ejecución. En la Figura 24 se observa la distribución de tiempos medidos para cada uno de los tipos de computadores en cada uno de los escenarios mediante diagrama de cajas. Se encontró que no es posible establecer diferencia de tiempos promedios entre los escenarios donde se ejecuta solo WhatsApp y WhatsApp más tres aplicaciones. Esto se debe principalmente a que en ningún caso el porcentaje de ocupación de la CPU está al 100 % lo que significa que la aplicación de interfaz gestual puede ejecutarse junto con otras aplicaciones simultáneamente y no afectar el desempeño de la CPU. Por otro lado, se observa que la diferencia de promedios es estadísticamente significativa (p-valor < 0.001) entre el computador de baja gama y los de media gama y alta gama. No se encontró diferencia de tiempos de ejecución promedio entre el computador de media gama y el de alta gama, lo que indica que es posible operar la interfaz con un tiempo de ejecución promedio por frame de aproximadamente 1.2 segundos desde un computador cuyo precio en la actualidad no supera los 1.6 millones de pesos colombianos (este valor se obtuvo consultando los principales proveedores de la ciudad de Santiago de Cali).

Figura 24. Diagrama de cajas de tiempo de ejecución por frame de la aplicación de interfaz gestual. (** p-valor < 0.001).



Fuente: Los autores

8.2 ROBUSTEZ

La interfaz gestual depende de dos técnicas claves para su funcionamiento: la detección de rostros y la detección de zonas de interacción. Por esta razón en este proyecto se estudia la robustez de estos bloques de forma cualitativa.

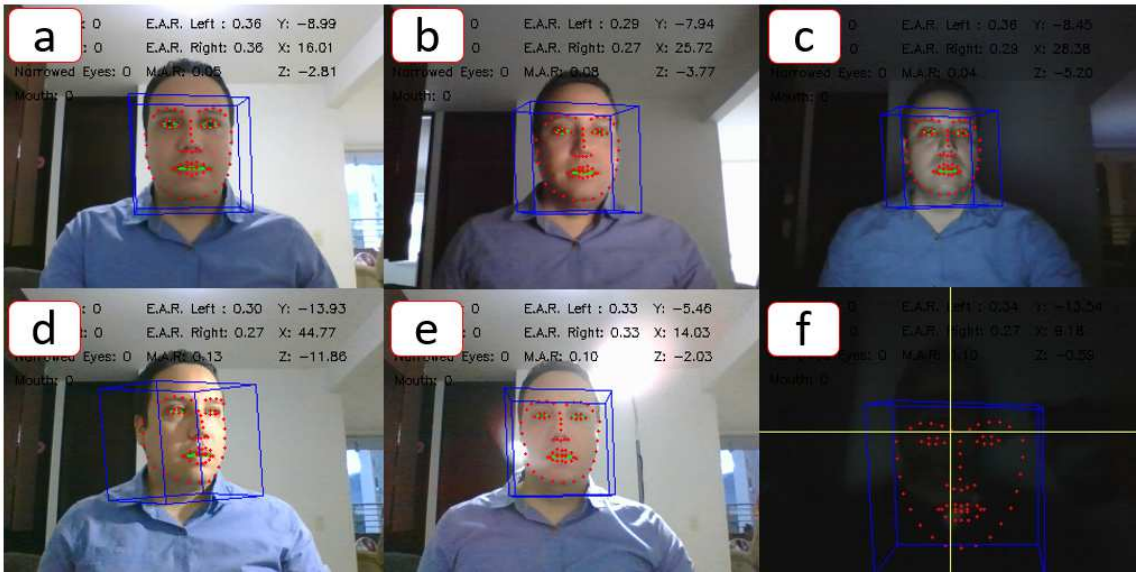
En la Figura 25 se observa que la detección de rostros funciona muy bien en diferentes condiciones de iluminación, incluso en completa oscuridad utilizando solo la luz emitida por el computador con el nivel de brillo al máximo. Para este caso el sujeto se ubicó a 80 cm de la pantalla del computador. La técnica falla cuando las condiciones de iluminación son muy bajas como es el caso de f) en la misma figura.

El criterio de seguimiento de una sola etiqueta permite que la interfaz no sea comandada por un segundo sujeto en presencia de otras personas en la misma escena del sujeto principal (quien debe operar la interfaz). La Figura 26 muestra este hecho.

La detección de rostro funciona bien hasta distancia menores de los 90 cm, después de este valor aun cuando se detecta el rostro el algoritmo de detección de pose empieza a fallar debido a que la detección de puntos característicos en el rostro no es tan robusta a esta distancia. Finalmente, cuando la distancia es mayor de los 120 cm el algoritmo de detección de rostros falla completamente. (Ver Figura 27). Es importante aclarar que en este caso y para lograr que la interfaz gestual ejecutara rápidamente la imagen de entrada de webcam es redimensionado siempre a 320x340 antes de ser aplicada dentro del flujo de pre-procesamiento análisis y

generación de eventos. Trabajar a una resolución mayor mejorará la efectividad de la detección en distancias superiores a los 120 cm, sin embargo, hará que el aplicativo de interfaz gestual ejecute más lentamente.

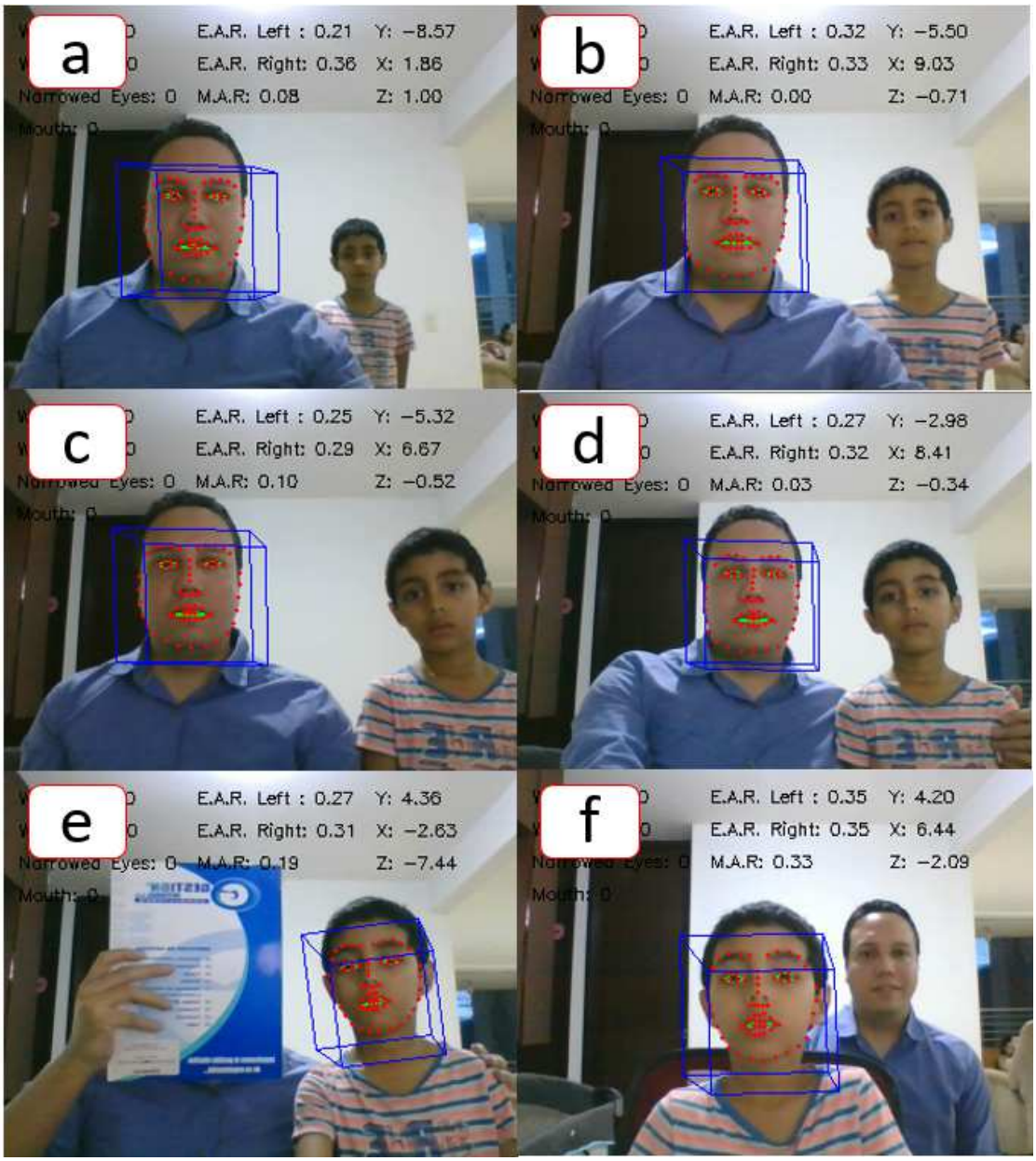
Figura 25. Detección de rostro en condiciones variadas de iluminación. a) Iluminación cuasi-uniforme dentro de habitación durante el día con lámpara superior, b) sin iluminación de lámpara, c) durante la noche con el brillo del computador al máximo, d) lámpara iluminando lateralmente, e) lámpara detrás del sujeto y de frente a la cámara y d) durante la noche con el brillo del computador al mínimo.



Fuente: Los autores

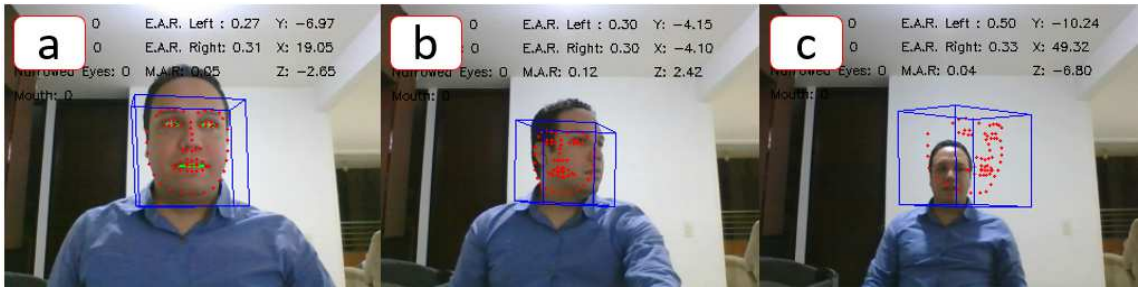
Las oclusiones son críticas en cualquier aplicación de detección de objetos utilizando imágenes, y la detección de rostro no es la excepción, sin embargo debe resaltarse que durante las pruebas se pudo observar que casos extremos como el uso de mascarillas, ponerse la mano en el rostro o incluso aparecer a medio cara en la imagen son casos en los que la detección del rostro no falla, sin embargo la detección de puntos característicos sobre el mismo sí, afectando directamente la etapa de estimación de pose del rostro. (Ver Figura 28). Poseer cabello largo, el uso de gafas con lentes transparentes e incluso desaparecer parte del rostro en la imagen, siempre que se alcance a ver la nariz y boca, no afectan la detección y estimación de pose del rostro.

Figura 26. Algoritmo de seguimiento de una sola etiqueta. En a), b), c) y d) se observa que aun cuando hay dos rostros la etiqueta se cede de frame a frame utilizando el criterio de mínima distancia al primer sujeto de izquierda a derecha. e) la etiqueta es cedida al segundo sujeto ya que el primero desapareció de la escena. f) la etiqueta la conservará el segundo sujeto mientras se cumpla el criterio de mínima distancia.



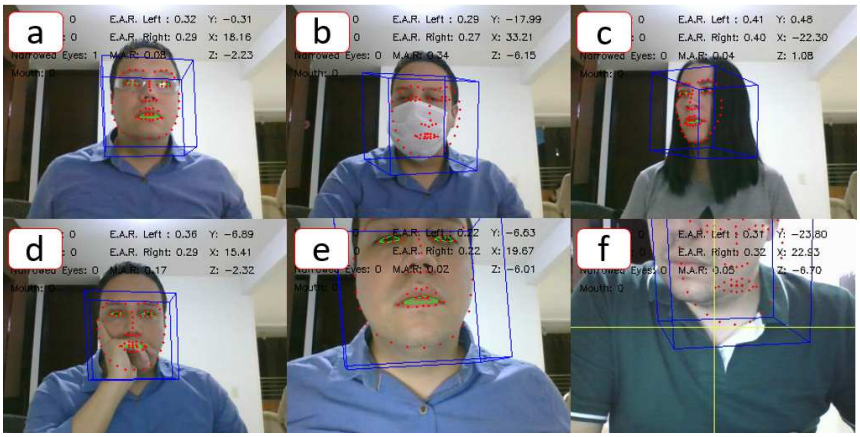
Fuente: Los autores

Figura 27. Efecto de la distancia en la efectividad del algoritmo de detección de rostros. a) < 90 cm, b) >90-120 cm y c) > 120 cm.



Fuente: Los autores

Figura 28. Casos de oclusión sobre el rostro. a) uso de gafas con lentes transparentes, b) uso de mascarilla, c) abundante cabello, d) mano sobre el rostro, e) parte del rostro (se alcanza a detectar ojos) por fuera de la imagen y f) parte del rostro (no se alcanza a detectar ojos) por fuera de la imagen.



Fuente: Los autores

En el caso de la detección de zonas de interés en la aplicación de WhatsApp a partir de la captura de pantalla se observó que la detección de dichas zonas es invariante a escala y traslación, (ver

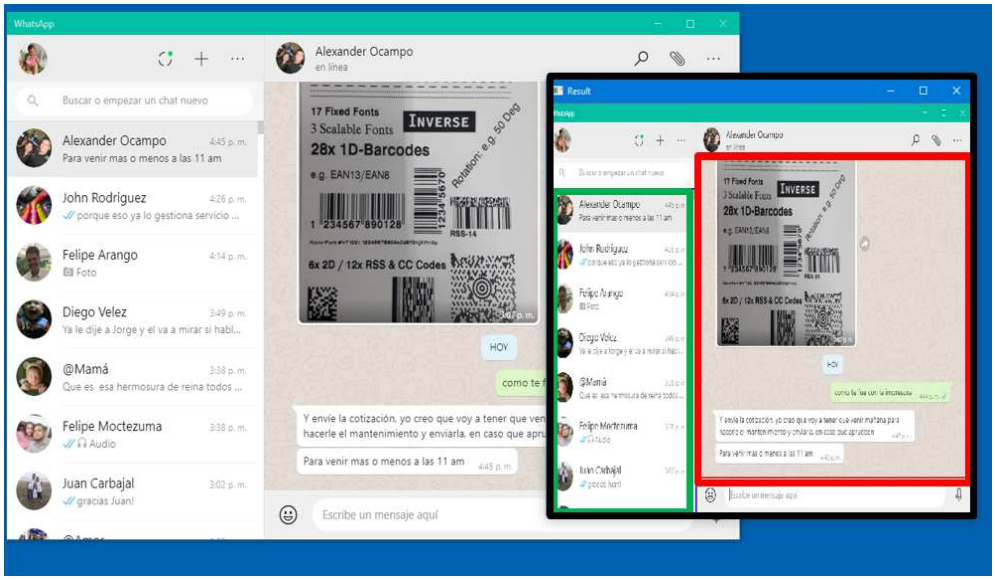
Figura 29 y

Fuente: Los autores

Figura 30). Esto se debe principalmente a que la técnica depende de poder identificar por correspondencia de plantillas los iconos de menú, búsqueda y

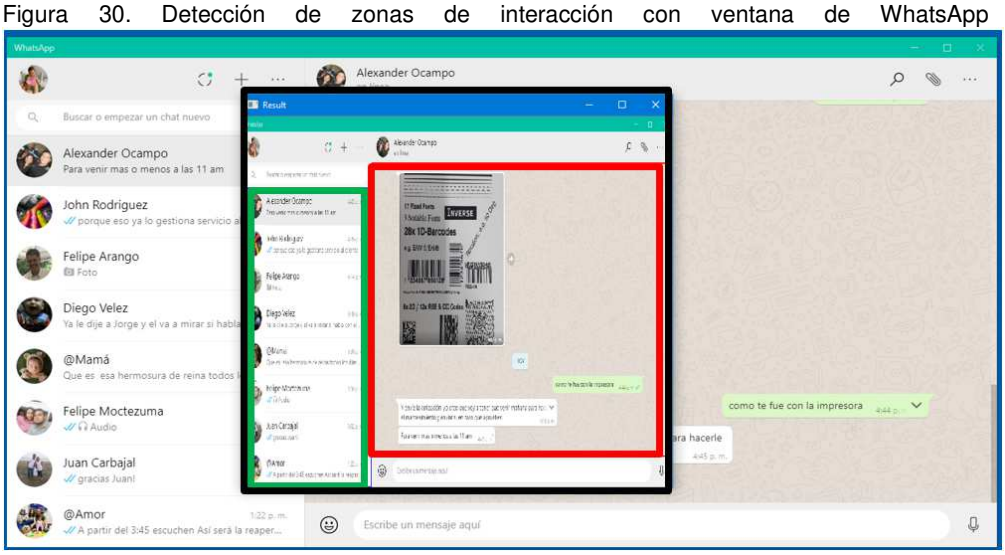
emoticón. En cualquier caso que no se logre observar con claridad estos íconos, en la captura de pantalla, se tendrá un resultado no deseado tal y como se evidencia en la Figura 31 y Figura 32. La Figura 33 es un caso que soporta nuestra tesis que el algoritmo solo depende de la correcta identificación de los íconos ya mencionados. En este caso la ventana de la aplicación no obstruye ninguno de los íconos y por tal razón se obtiene una correcta identificación de las zonas de interacción. Sin embargo, en caso que la aplicación de WhatsApp se encuentre en segundo plano, los comandos de interacción no tendrán ningún efecto, por el contrario, es posible tener algún efecto no deseado sobre la aplicación que esté en primer plano en ese momento.

Figura 29. Detección de zonas de interacción con ventana de WhatsApp reducida de tamaño y trasladado de origen de coordenadas.



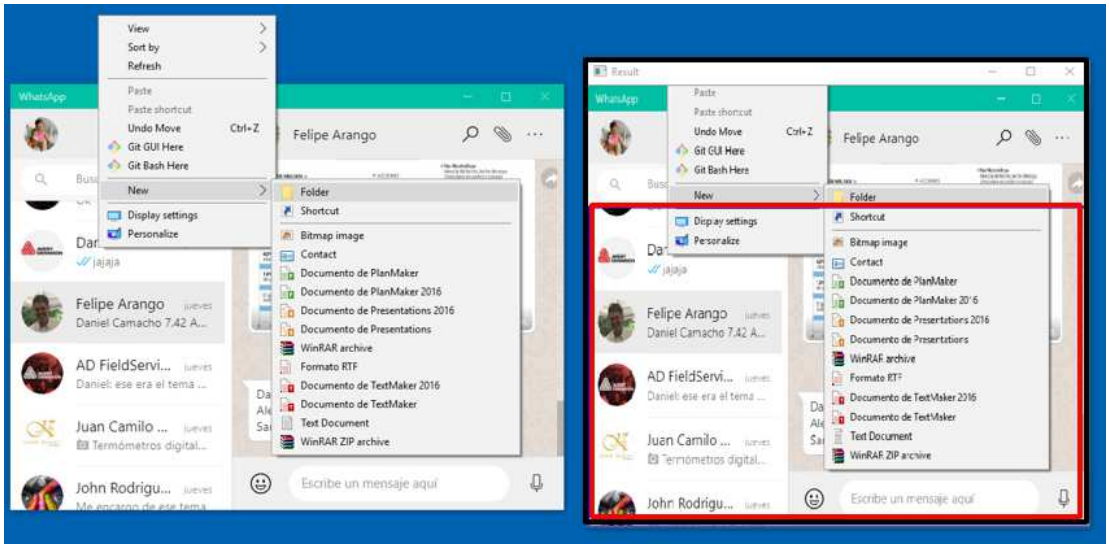
Fuente: Los autores

Figura 30. Detección de zonas de interacción con ventana de WhatsApp maximizada.



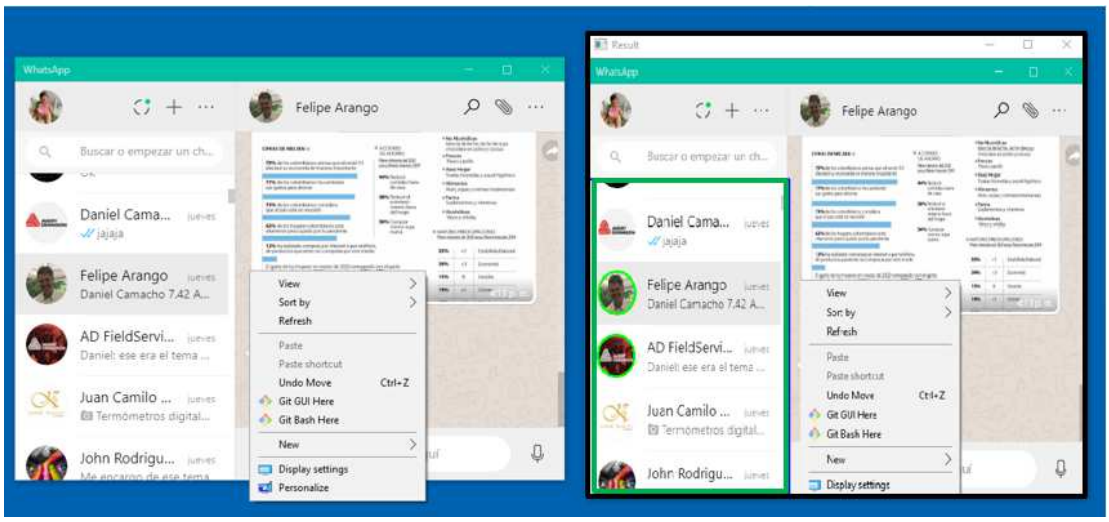
Fuente: Los autores

Figura 31. Detección errada de zonas de interacción debido a la presencia de elementos contextuales del sistema operativo cubriendo el icono de menú.



Fuente: Los autores

Figura 32. Detección de una sola zona de interacción debido a la presencia de elementos contextuales del sistema operativo cubriendo el icono de emoticón.



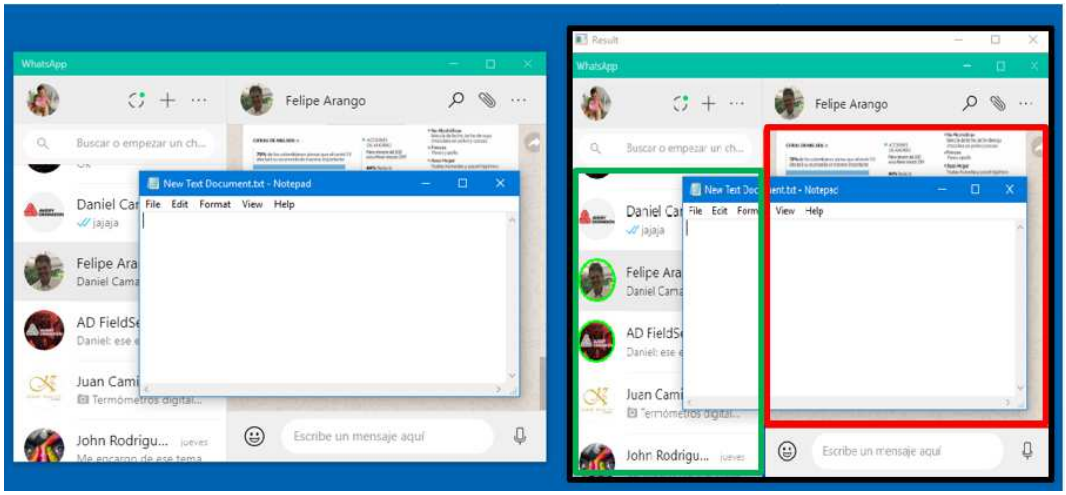
Fuente: Los autores

8.3 PRUEBAS DE USABILIDAD

Con el fin de explorar la facilidad de uso de la interfaz gestual propuesta se realizó un experimento con onces personas (seis mujeres, cinco hombres) cuyas edades

varían de los 12 años hasta los 67 años. Con ayuda de un consentimiento informado (ver ANEXO) se daba cuenta del experimento en que los sujetos iban a participar y una vez firmado el mismo se procedía con la explicación del procedimiento experimental, el cual consistía durante los primeros 3 minutos en explicar el uso del mismo. Durante las pruebas las personas abrían su cuenta de WhatsApp para estar más familiarizados con los contactos que iban a seleccionar durante las pruebas.

Figura 33. Detección de zonas de interacción con presencia de ventana de aplicación Bloc de Notas de Windows que no cubre ninguno de los iconos claves para la detección de plantillas.



Fuente: Los autores

La prueba consistió en utilizar los comandos de interacción desarrollados en la propuesta de solución, a cada usuario se le solicitó seleccionar 10 contactos personales o grupales de la zona de contactos y que navegara de arriba hacia abajo y de abajo hacia arriba las conversaciones. La supervisión del experimento siempre se llevó a cabo desde la parte trasera del computador para evitar violar la intimidad de las conversaciones de los sujetos experimentales.

El tiempo promedio de interacción con WhatsApp mediante la interfaz gestual fue en promedio de 10 minutos. Una vez finalizado el ejercicio de interacción con ayuda de un formulario, Tabla 2, se recogía información sobre la percepción de los usuarios sobre la solución de interfaz gestual propuesta. De las respuestas obtenidas en forma porcentual para cada una de las encuestas de la Figura 34, Figura 35, Figura 36, Figura 37 y Figura 38, se observa que en general la percepción de los sujetos experimentales es que la aplicación de interfaz gestual es amigable, con una curva rápida de aprendizaje y permite interactuar con WhatsApp de forma efectiva bajo los comandos que actualmente permite la misma. Sin embargo, se percibe que la interfaz no es tan rápida en su ejecución, de hecho, algunos sujetos

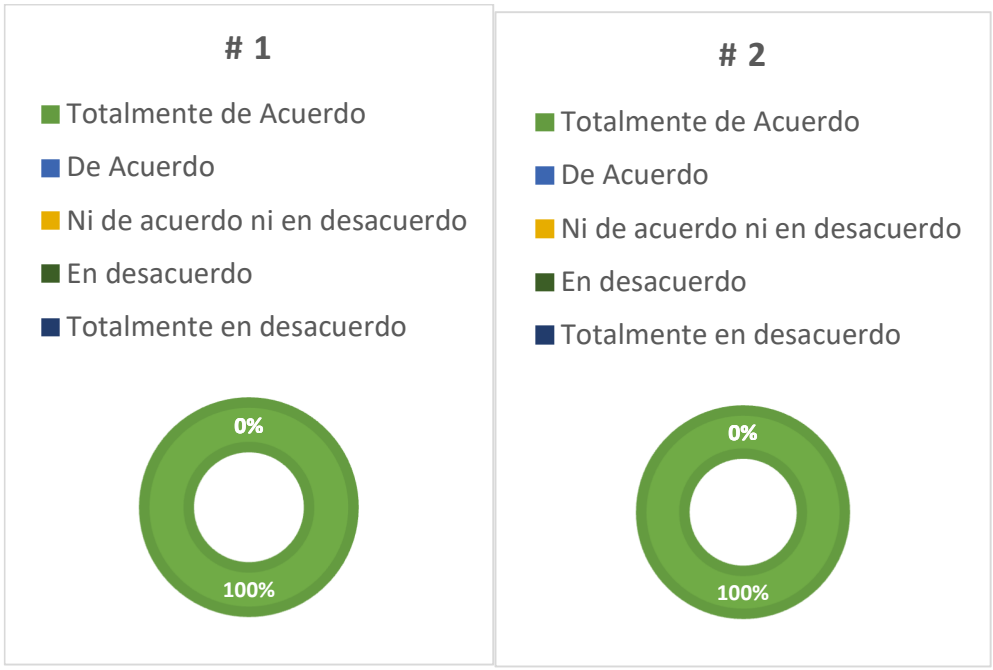
manifestaron que consideran debe mejorarse esta parte dentro de la interfaz para hacerla más intuitiva.

Tabla 2. Formulario de encuesta para las pruebas de usabilidad.

No.	Encuesta	Totalmente en desacuerdo	En desacuerdo	Ni de acuerdo ni en desacuerdo	De acuerdo	Totalmente de acuerdo
1	La Interfaz detecta el rostro.					
2	La interfaz detecta y posiciona correctamente el cursor en las zonas de contactos y mensajes en WhatsApp de escritorio.					
3	La interfaz reconoce todos los comandos de navegación (Mov.Izquierda; Mov.Derecha; Mov.Arriba; Mov.Abajo).					
4	La interfaz reconoce el comando gestual para salir de la interfaz (boca abierta).					
5	La ejecución de los comandos en la interfaz es rápida.					
6	Teniendo en cuenta que la interfaz está pensada para personas con tetraplejia, Considera que la herramienta ayuda establecer comunicación con WhatsApp de escritorio.					
7	La interfaz es de fácil uso para el usuario.					
8	La interfaz se puede utilizar en cualquier ambiente de iluminación sin interferencia en la detección del rostro.					
9	La interfaz cumple el objetivo de realizar tareas básicas de navegación en la aplicación WhatsApp desktop.					
10	Pensando en la necesidad de las personas con limitaciones motrices de miembros superiores, considera que la interfaz es una herramienta que se puede recomendar.					

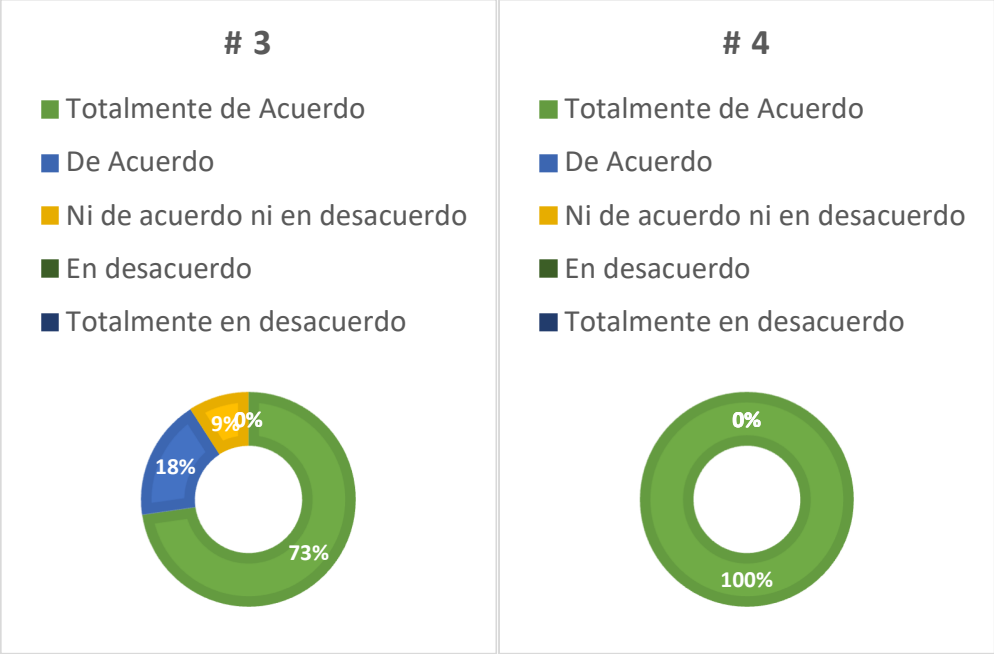
Fuente: Los autores

Figura 34. Porcentajes de respuesta para las preguntas 1 y 2 de la encuesta.



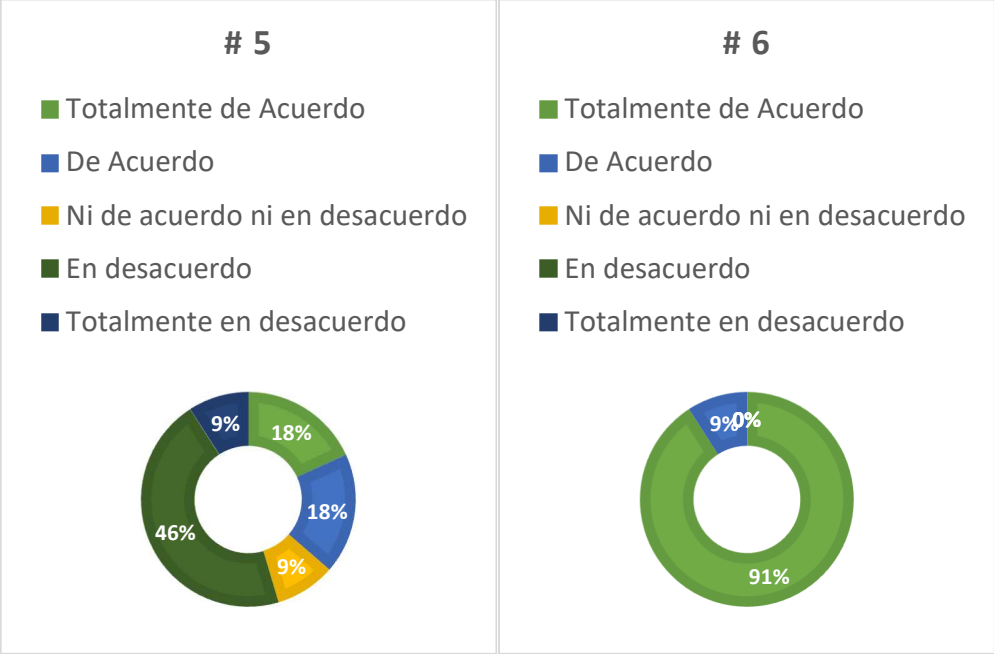
Fuente: Los autores

Figura 35. Porcentajes de respuesta para las preguntas 3 y 4 de la encuesta.



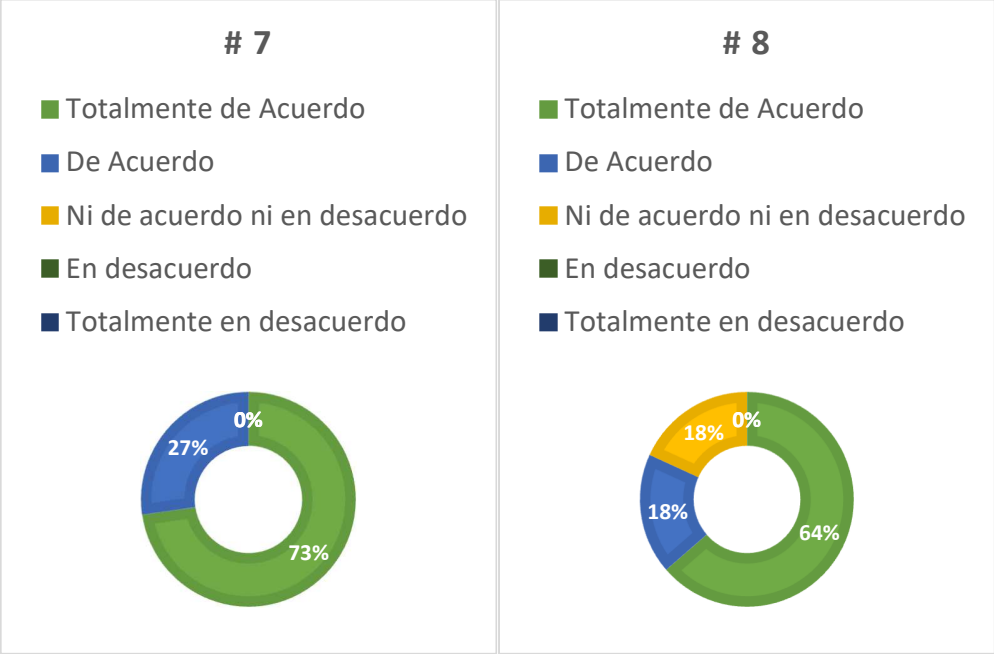
Fuente: Los autores

Figura 36. Porcentajes de respuesta para las preguntas 5 y 6 de la encuesta.



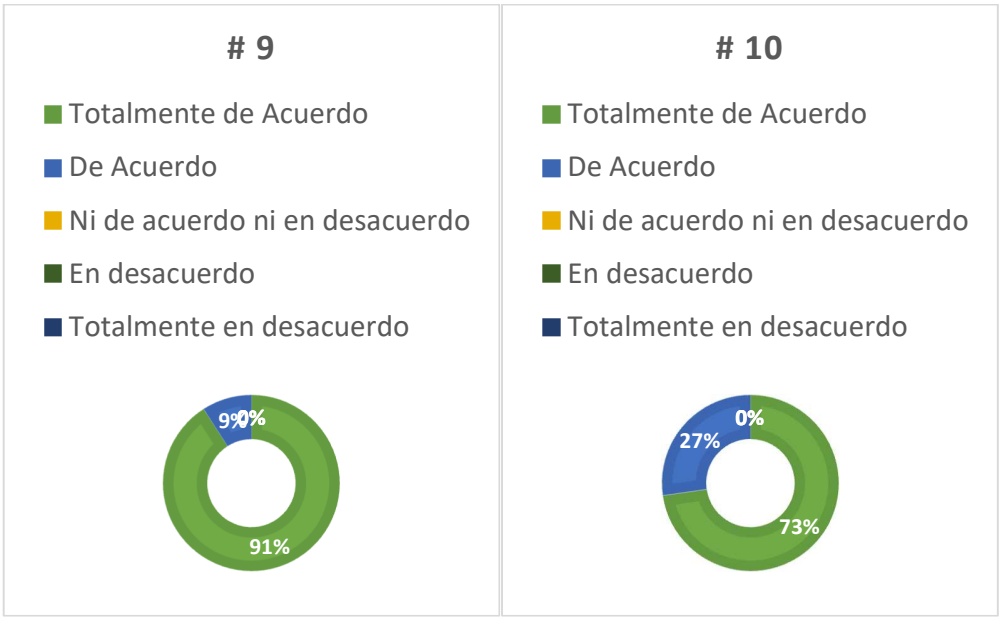
Fuente: Los autores

Figura 37. Porcentajes de respuesta para las preguntas 7 y 8 de la encuesta.



Fuente: Los autores

Figura 38. Porcentajes de respuesta para las preguntas 7 y 8 de la encuesta.



Fuente: Los autores

En todos los casos los sujetos estuvieron de acuerdo en que ven en la aplicación aquí propuesta un gran potencial para ser usado en personas con limitaciones motrices. Sin embargo, consideran que deben incluirse otros gestos como el guiño del ojo, el parpadeo intermitente, etc., dentro de la interfaz gestual para ampliar sus capacidades.

9. CONCLUSIONES

Se desarrolló una interfaz humano computador basada en gestos faciales y detección de zonas de interés en una aplicación orientada al internet para personas con limitaciones motrices de miembros superiores. La interfaz permite mediante cinco comandos seleccionar contactos, navegar en las conversaciones de la aplicación WhatsApp de escritor, así como y activar/desactivar la interfaz utilizando únicamente gestos del rostro.

Se determinaron las principales características de una interfaz humano-computador para personas con limitaciones motrices de miembros superiores. La interfaz cuenta con principios ergonómicos para facilitar y optimizar su utilización. También posee una combinación de tecnología, conocimiento y recursos que produjeron resultados deseados en equipos de cómputo con precios inferiores a los 1.5 millones pesos colombianos.

Se implementó una técnica de visión por computador para la identificación de gestos faciales y la detección de zonas de interés en una aplicación de escritorio orientada a internet. Se logró la detección automática a partir de una captura de pantalla de la aplicación WhatsApp las zonas de contacto y conversaciones mediante una combinación de heurísticas y correspondencia por plantilla de imágenes. La detección de gestos se logró utilizando técnicas robustas de detección de rostros y estimación de pose mediante modelos flexibles.

Se construyó una interfaz software de generación de comandos para una aplicación de escritorio a partir de gestos y zonas de interés detectadas. Esta interfaz es capaz de correr en fondo y no interfiere visualmente con la aplicación WhatsApp. La velocidad de ejecución de la misma es de 1 Hz, y en un equipo de media gama ocupa el 35 % de CPU y 1.5 GB RAM.

Se llevaron a cabo pruebas de desempeño computacional, robustez y de usabilidad. Primero, las pruebas de desempeño permitieron identificar que en un equipo de media gama es posible trabajar la interfaz gestual y que se pueden tener más aplicaciones ejecutándose simultáneamente sin que esto reduzca la velocidad de ejecución de esta. Segundo, las pruebas de robustez evidenciaron que la interfaz gestual puede trabajar en variadas condiciones de iluminación y con algunos casos de oclusión tanto a nivel de detección de gestos faciales como detección de zonas de interés. Por último, las pruebas de usabilidad permitieron entender que los usuarios reconocen la funcionalidad de la interfaz implementada, así como su potencial para personas con limitación motriz, sin embargo, sugieren que debe ampliarse el número de eventos y de capacidades de la interfaz para ser completamente útil además de sugerir una mejora en los tiempos de ejecución para manejo sea más fluido.

En general, los autores de este trabajo consideran que los hallazgos encontrados en la ejecución de este proyecto son claves para su aplicación en personas con limitaciones motrices de miembro superior y que deben integrarse a la parte de reconocimiento de voz (el cual es otro proyecto que se está realizando en la Fundación Universitaria Lumen Gentium) para complementar su espectro de aplicación.

10. RECOMENDACIONES

Con el fin de ampliar el alcance obtenido y de igual forma superar las limitaciones de la solución propuesta en este proyecto de investigación se identifican los siguientes trabajos futuros:

- Incluir más eventos a partir de la identificación de otros gestos del rostro, por ejemplo, los que involucran los ojos.

- Mejorar el tiempo de ejecución de la interfaz gestual mediante la re-implementación de algunas técnicas de análisis de imágenes para que escalen mejor las capacidades de la CPU.

- Integrar reconocimiento por voz para permitir el ingreso de texto utilizando técnicas de procesamiento del lenguaje natural.

BIBLIOGRAFÍA

[1] D. Lalanne and J. Kohlas, Eds., Human Machine Interaction, vol. 5440. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[2] K. Warwick, “The Future of Human-Machine Interaction: Implant Technology,” 2011, pp. 11–19.

[3] C. Manresa-Yee, P. Ponsa, J. Varona, and F. J. Perales, “User experience to improve the usability of a vision-based interface,” *Interact. Comput.*, vol. 22, no. 6, pp. 594–605, 2010, doi: 10.1016/j.intcom.2010.06.004.

[4] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, “Vision-based hand-gesture applications,” *Commun. ACM*, vol. 54, no. 2, pp. 60–71, Feb. 2011, doi: 10.1145/1897816.1897838.

[5] M. de S. y P. S. Colombia, “Sala Situacional Situación de las Personas con Discapacidad,” 2018.

[6] N. Balsero, D. Botero, J. Zuluaga, and C. Parra Rodríguez, “Interacción Hombre-Máquina Usando Gestos Manuales en Texto Real,” *Ing. y Univ.*, vol. 9, no. 2, pp. 101–112, 2005.

[7] William Alfredo Castrillón Herrera, “Implementación de una Interfaz Hombre-Máquina para el Control de un Brazo Robótico Mediante Posturas Labiales,” Universidad Nacional de Colombia Sede Manizales, 2009.

[8] J. H. Mosquera, H. Loaiza, S. E. Nope, and A. D. Restrepo, “Identifying Facial Gestures to Emulate a Mouse: Navigation Application on Facebook.,” *IEEE Lat. Am. Trans.*, vol. 15, no. 1, pp. 121–128, Jan. 2017, doi: 10.1109/TLA.2017.7827915.

[9] M. Betke, J. Gips, and P. Fleming, “The Camera Mouse: Visual tracking of Body Features to Provide Computer Access for People with Severe Disabilities,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 10, no. 1, pp. 1–10, 2002, doi: 10.1109/TNSRE.2002.1021581.

[10] T. Granollers and M. García, “Computer Vision Interaction for People with Cerebral Palsy,” *An Interdiscip. J. Humans ICT Environ.*, vol. 2, no. April, pp. 38–54, 2006.

[11] E. Perini, S. Soria, A. Prati, and R. Cucchiara, “FaceMouse: A Human-Computer Interface for Tetraplegic People,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3979 LNCS, pp. 99–108, 2006, doi: 10.1007/11754336_10.

[12] J. Varona, C. Manresa-Yee, and F. J. Perales, “Hands-Free Vision-Based

Interface for Computer Accessibility," *J. Netw. Comput. Appl.*, vol. 31, no. 4, pp. 357–374, 2008, doi: 10.1016/j.jnca.2008.03.003.

[13] A. Matos, V. Filipe, and P. Couto, "Human-Computer Interaction Based on Facial Expression Recognition: A Case Study in Degenerative Neuromuscular Disease," *ACM Int. Conf. Proceeding Ser.*, pp. 8–12, 2016, doi: 10.1145/3019943.3019945.

[14] H. Mosquera, H. Loaiza, S. Nope, and A. Restrepo, "Identifying Facial Gestures to Emulate a Mouse: Control Application In a Web Browser," in *2016 XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, 2016, pp. 1–6, doi: 10.1109/STSIVA.2016.7743345.

[15] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *Int. J. Eng. Technol.*, vol. 7, no. 2, pp. 29–32, 2018.

[16] J. A. Jacko and D. Wigdor, Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications, Third. Amazon, 2012.

[17] E. Al Flaspöler E., Hauke A, Pappachan P., Reinert D., The Human Machine Interface as an Emerging Risk. European Agency for Safety and Health at Work, 2010.

[18] N. Ismael Fernando Escalona, "Interfaz Humano Máquina Controlada por Gestos," *Repos. Univ. Chile*, p. 70, 2014.

[19] J. Cannan and H. Hu, "Human-Machine Interaction (HMI) A Survey," *Sch. Comput. Sci. Electron. Eng. Univ. Essex*, 2010.

[20] C. Hernandez and P. Baptista, Metodología de la Investigación, 6th ed. 2010.

[21] Z. Ali and Sb. Bhaskar, "Basic Statistical Tools in Research and Data Analysis," *Indian J. Anaesth.*, vol. 60, no. 9, p. 662, 2016, doi: 10.4103/0019-5049.190623.

[22] R. Szeliski, *Computer Vision*, vol. 42. London: Springer London, 2011.

[23] C. Demant, B. Streicher-Abel, and C. Garnica, Industrial Image Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[24] R. Brunelli, Template Matching Techniques in Computer Vision: Theory and Practice. Wiley, 2009.

[25] A. Sweigart, "PyAutoGUI Documentation." Read the Docs, p. 25, 2020.

[26] S. Zafeiriou, C. Zhang, and Z. Zhang, "A Survey on Face Detection In The Wild: Past, Present And Future," *Comput. Vis. Image Underst.*, vol. 138, no.

March, pp. 1–24, 2015, doi: 10.1016/j.cviu.2015.03.015.

[27] A. Kumar, A. Kaur, and M. Kumar, “Face Detection Techniques: a Review,” *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 927–948, 2019, doi: 10.1007/s10462-018-9650-2.

[28] Mario Cifrek, Ivan Culjak, David Abram, Tomislav Pribanic, Hrvoje Dzapo, “A Brief Introduction To Opencv,” 2012.

[29] P. Werner, F. Saxen, and A. Al-Hamadi, “Landmark Based Head Pose Estimation Benchmark And Method,” *Proc. - Int. Conf. Image Process. ICIP*, vol. 2017-Septe, pp. 3909–3913, 2018, doi: 10.1109/ICIP.2017.8297015.

[30] X. Ren, J. Ding, J. Sun, and Q. Sui, “Face Modeling Process Based On Dlib,” in *2017 Chinese Automation Congress (CAC)*, 2017, pp. 1969–1972, doi: 10.1109/CAC.2017.8243093.

[31] Xiao-Shan Gao, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng, “Complete Solution Classification for The Perspective-Three-Point Problem,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 930–943, Aug. 2003, doi: 10.1109/TPAMI.2003.1217599.

[32] V. Lepetit, F. Moreno-Noguer, and P. Fua, “EPnP: An accurate O(n) Solution to the PnP Problem,” *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, 2009, doi: 10.1007/s11263-008-0152-6.

[33] Adrian Kaehler and G. Bradski, Learning OpenCV: Computer Vision in C++ with the OpenCV Library. Sebastopol: O’Reilly Media, Inc., 2013.

ANEXO

CONSENTIMIENTO INFORMADO PARA PARTICIPACIÓN EN INVESTIGACIÓN

INTERFAZ HUMANO-COMPUTADOR PARA PERSONAS CON LIMITACIÓN MOTRIZ DE MIEMBROS SUPERIORES BASADA EN GESTOS FACIALES

I. INFORMACIÓN

Usted ha sido invitado a participar en la investigación **Interfaz Humano-Computador para Personas con Limitación Motriz de Miembros Superiores Basada en Gestos Faciales.** Para decidir participar en esta investigación, es importante que considere la siguiente información. Siéntase libre de preguntar cualquier asunto que no le quede claro. Si no tiene preguntas ahora, usted podrá hacerlas en cualquier momento, a través de los datos de contacto abajo relacionados. Su participación en este estudio es completamente voluntaria.
El objetivo de este estudio es Desarrollar una interfaz humano computador basada en gestos faciales y detección de zonas de interés en una aplicación orientada al internet para personas con limitaciones motrices de miembros superiores.

Este estudio de investigación está dirigido a 4 personas con limitaciones en extremidades superiores. Aproximadamente, participarán en este estudio un total de 4 sujetos con tetraplejía y 10 personas sanas.

Participación: Si usted autoriza participar en este estudio se le aplicarán las siguientes técnicas de recolección de la información:

1. Se harán pruebas para implementar una interfaz humano computador basada en gestos faciales y detección de zonas de interés en una aplicación orientada al internet para personas con limitaciones motrices de miembros superiores. Para esto el sujeto estará ubicado frente a una cámara de video y monitor de computador, para así poner a prueba el programa desarrollado.

Durante el desarrollo de la técnica o procedimiento se tomará anotaciones por parte del investigador o se hará uso de una grabadora de voz o cámara de video con el fin de revisar en detalle la información obtenida, sin embargo, esta grabación podrá ser interrumpida y/o retomada en el momento que usted lo considere, sin que esto conlleve algún perjuicio para usted o la institución que representa.

Riesgos: El estudio que se desarrollará ha sido clasificado como una investigación. Dado que el proyecto culminará con el desarrollo de una interfaz humana computador la cual se ensayará con personas reales, el tipo de estudio propuesto es de tipo experimental. Básicamente este proyecto se apoyará en trabajos sistemáticos fundamentados en los conocimientos alrededor de los sistemas de reconocimiento de gestos de la cara que permitirá la creación de una interfaz humano computador basada en gestos faciales y detección de zonas de interés en una aplicación orientada al internet para personas con limitaciones motrices de miembros superiores Según el Artículo 11 de la Resolución 008430 de 1993, esta investigación no representa riesgo alguno para los participantes; sin embargo, siempre existe el riesgo de violación de la confidencialidad y por ello hemos adoptado todos los mecanismos necesarios para minimizar este riesgo, con la firma de los respectivos acuerdos de confidencialidad.

Beneficios: Usted no recibirá ningún beneficio directo, ni recompensa alguna, por participar en este estudio. No obstante, su participación permitirá generar información para determinar las principales características de una interfaz humano-computador para personas con limitaciones motrices de miembros superiores, implementar una técnica de visión por computador para la identificación de gestos faciales y la detección de zonas de interés en una aplicación de escritorio orientada a internet, desarrollar una interfaz software de generación de comandos para una aplicación de escritorio a partir de gestos y zonas de interés detectadas y ejecutar un plan de pruebas para definir los alcances y limitaciones del sistema.

Voluntariedad: Su participación es absolutamente voluntaria. Usted tendrá la libertad de preguntar lo que considere pertinente para proseguir, como también de detener su participación en cualquier momento que lo desee. Esto no implicará ningún perjuicio para usted.

Confidencialidad: La información recolectada no será usada para ningún otro propósito, además de los señalados anteriormente, sin su autorización previa y por escrito. Todas sus opiniones y aportes serán confidenciales, y mantenidas en estricta reserva. En las presentaciones y publicaciones de esta investigación, su nombre no aparecerá asociado a ninguna opinión particular. Hasta donde nos es posible, sus respuestas serán confidenciales y minimizaremos los riesgos asignando un código a su formulario y limitando el acceso a sus datos personales únicamente al investigador principal.

Conocimiento de los resultados: Usted tiene derecho a conocer los resultados de esta investigación. Para ello, una vez sean sistematizados y publicados los resultados de la investigación, se le hará llegar una copia del enlace que le permitirá acceder a la misma a la dirección de correo electrónico por usted aportada.

Datos de contacto: Si requiere más información o comunicarse por cualquier motivo relacionado con esta investigación, puede contactar al investigador principal responsable de este estudio:

Nombre: _____ Teléfonos: _____ E-mail: _____
También puede comunicarse con el Comité de Ética de la Investigación que aprobó este estudio:

Presidente (a) del Comité de Ética de la Fundación Universitaria Católica Lumen Gentium –Unicatólica-: Fabio Alberto Enríquez Martínez

Carrera 122 No. 12 - 409, Cali, Valle del Cauca, PBX (2) 5552767 Ext. 1110

Correo Electrónico: ceiunicatolica@unicatolica.edu.co

II. FORMULARIO DE CONSENTIMIENTO INFORMADO

Yo, ……., acepto participar en el estudio INTERFAZ HUMANO-COMPUTADOR PARA PERSONAS CON LIMITACIÓN MOTRIZ DE MIEMBROS SUPERIORES BASADA EN GESTOS FACIALES, en los términos aquí señalados.

Declaro que he leído y comprendido, las condiciones de mi participación en este estudio. He tenido la oportunidad de hacer preguntas y estas han sido respondidas en su totalidad por el investigador responsable. También sé que se respetará la buena fe, la confidencialidad e intimidad de la información por mí suministrada, lo mismo que mi seguridad física y psicológica. No tengo dudas al respecto.

_____ _____

Firma Participante Firma Investigador Responsable

c.c. _____ c.c. _____

Correo electrónico: _____

Lugar y Fecha: _____