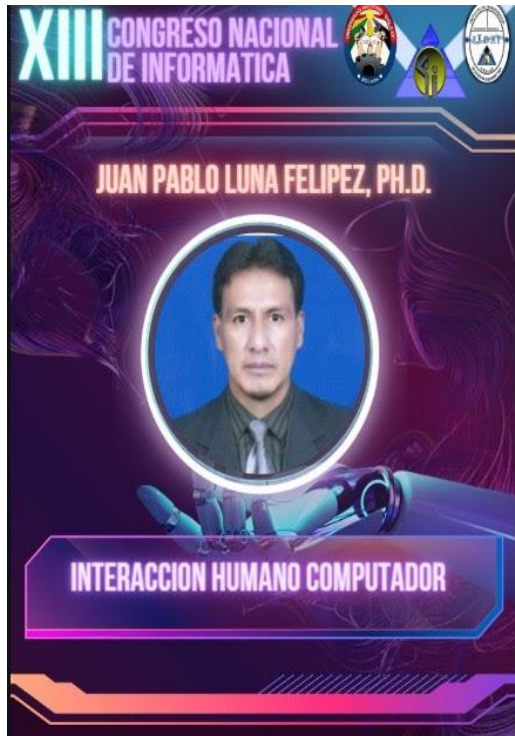


INTERACCIÓN HUMANO COMPUTADOR



Juan Pablo Luna Felipez, Ph.D.

Ingeniería Informática

Universidad Nacional “Siglo XX”

Llallagua, Bolivia

jplunaf@gmail.com

RESUMEN. –

Actualmente dependemos de las computadoras para realizar tareas diariamente y mejorar las interfaces de interacción entre el humano y la computadora es una necesidad muy importante

La detección de gestos de la mano es un proceso complejo, donde los métodos existentes aún no son suficientes ni óptimos, han tenido poco avance, falta fiabilidad en la detección, precisión, presenta desafíos para su implementación y cuya necesidad es innegable como medio de interacción natural actual, por tanto es un tema abierto a la investigación

La investigación propone un método que permite el reconocimiento de gestos de la mano a partir de imágenes bidimensionales, aportando con una solución innovadora empleando la proporción áurea independientemente del tamaño de la mano a partir de una imagen adquirida por una cámara monocular simple, sin calibración previa y de baja resolución.

También aporta de forma novedosa con la detección de puntos límites para determinar la posición de los límites de los dedos de la mano, así como un modelo de almacenamiento y reconocimiento de gestos de la mano y varias estrategias de aceleración en diferentes etapas del método, de forma que acelera el reconocimiento del gesto.

Se presentan también prototipos que implementan el método, contruidos sobre un engine altamente utilizable como es Unity, el cual permite construir y obtener productos sobre diversas plataformas como Linux, Windows, Android, IOS, Web, Xbox, etc. logrando una alternativa altamente útil y atractiva que puede aplicarse en lenguaje de señas, sistemas de realidad virtual, aumentada, medicina, robótica, etc.

El método trabaja en tiempo real ya que se logró que los cálculos no sean costosos computacionalmente y se probó con diferentes personas y varios programas que se desarrollaron con entornos tridimensionales que requieren una alta velocidad de respuesta en tiempo real, logrando resultados altamente satisfactorios.

1. INTRODUCCIÓN

De acuerdo con Andrade (2016) las primeras interfaces de usuario eran por línea de comando, donde se interactuaba con la computadora mediante el teclado, en los 90 como afirma Peralta (2012) las interfaces gráficas de usuario, junto al ratón y teclado se constituyeron en la interfaz predominante de interacción

Otro hito de interacción según Vanderdonckt, Coutaz, Calvary & Stanculescu (2007) son las pantallas táctiles y multitáctiles

Actualmente existe la tendencia denominada interfaz natural de usuario NUI, según Wigdor & Dennis (2011) busca que el usuario interactúe con la computadora como interactúa con otras personas y el entorno a partir de gestos humanos. Desde el 2006 existe una comunidad para expandir la investigación y desarrollo de las NUI (Community, s.f.).

Un campo del reconocimiento de gestos humanos, es el reconocimiento de gestos de manos, que según Andrade (2016) en los últimos años ha atraído un creciente interés por sus aplicaciones en diversos campos como: interacción humano computador, robótica, lenguaje de señas, realidad virtual, medicina, etc.

Algunos trabajos relacionados son:

Hand Gesture Recognition Using sEMG and Deep Learning, Nadia(2021)

Reconocimiento estático de gestos manuales basado en redes neuronales, Pinto et al (2019)

Real Time Static Gesture Detection Using Machine Learning, Sandipgiri (2019)

Modelo computacional para reconocimiento de lenguaje de señas en un contexto colombiano, Ortiz(2020)

Hand Gesture Recognition using fusion of SIFT and HoG with SVM as a Classifier, Ansari(2017)

Un enfoque inteligente para el reconocimiento de gestos manuales, (Andrade, 2016)

El objetivo de la investigación fue: Desarrollar un método para el reconocimiento de gestos de la mano a partir de imágenes bidimensionales que permita mejorar la interacción humano computador

Reconocimiento de Gestos

La mano realiza gestos según Andrade(2016) se dividen en estáticos y dinámicos

Existen Métodos de reconocimiento

Basados en sistemas 3D: Con Leap Motion Huitzil, Pajaro & Ramirez (2017) y Raziq y Latif (2016)

Con Kinect Yang (2016), Realpe (2013) , Barros (2014), Rodríguez (2014) y Ploufee (2015)

Con cámaras RGB-D Nuñez (2019)

Con dispositivos wearables Betancur (2013) y Motoche (2018)

Basados en Imágenes

Con Redes Neuronales Sandipgiri (2019) y (Pinto et Al, 2019)

Con Histogramas de Frecuencias Villa, Valencia y Berrio (2018) y Ronchetti (2016)

Con clasificadores Haar Cascada Styles: Gurav et. al. (2015)

Con SVM Ansari (2017) y Berru (2019)

Con SIFT Villa, Valencia & Berrio (2018) y Otros

Aplicaciones: Según Ronchetti (2016) y Huitzil, Pajaro & Ramirez (2017) se aplica en: monitoreo de pacientes, control de videojuegos, navegación y manipulación de entornos virtuales, lenguaje de señas, navegar por la web, manipulación de robots, drones, aplicaciones médicas, etc.

Proporción Áurea

Serie de Fibonacci: Según Chacón (2014) serie que comienza por dos números uno, se suman los dos últimos números consecutivos para formar el siguiente, tiene la forma 1, 1, 2, 3, 5, etc.

Si se halla la proporcionalidad entre dos números consecutivos que tienden al infinito se acerca a 1. 6180

Serie de Lucas: Según (Hutchison& Hutchison, 2010) las series donde se suman los dos últimos términos para formar otro se denomina serie de Lucas, tiene la forma: 2, 1, 3, 4, etc. también tiende a 1. 6180.

Serie de Littler: Según Hutchison& Hutchison (2010) es una serie especial, donde se suman los dos últimos términos de la serie para formar uno nuevo, empieza en 1 y 1.3. y tiene la forma: 1, 1.3, 2.6, 5.9, etc. también tiende a 1. 6180.

Phi: Existe una proporcionalidad que se encuentra en varios aspectos de la vida y el universo y se denomina Proporción áurea o Phi Φ y se aproxima a 1.618. Euclides define Phi como: “Se dice que una recta está dividida en media razón y extrema razón cuando la longitud de la recta total es al segmento mayor, como el segmento mayor es al segmento menor” (Proporción Áurea, 2020)

Garijo & de Maria (2017) presentan algunas propiedades matemáticas de la proporción áurea

Proporción áurea de la Mano

Autores como Wang et al (2017), Díaz (2017) Park et. Al (2003) y otros concluyen que, en el cuerpo humano y en la mano se cumple la proporcionalidad áurea, también indica que, “Tomando la medida de menor a mayor la falange última de los dedos está en proporción áurea respecto a la siguiente, y está a la siguiente”. Según (Phiproporcion, 2020) no todas las personas tienen las dimensiones de su cuerpo exactamente en la proporción phi, pero en promedio tienden a phi

La mano humana se compone de metacarpo y falanges y cada dedo por: falangeta, falange y falangina, excepto el pulgar, según Priyanka Katyal et al.(2019), los dedos índice medio y anular, se hallan en una serie de Littler y el meñique en serie de Fibonacci

2. MATERIALES Y MÉTODOS

El método propuesto comprende 5 etapas:

a) Obtención de Datos

Captura de Imágenes y Región de Interés ROI: Se detecta y activa la cámara recuperando las imágenes en forma de matriz de píxeles, se define una ROI que es una región rectangular de la imagen donde se ubicará la mano y permitirá un procesamiento más veloz, se ubicará en la parte superior derecha y será modificable.

Obtención de imágenes del fondo y de los gestos: Se captura la imagen del fondo sin la mano, ya que permanecerá fija, luego se captura las imágenes con gestos, obteniendo la matriz de píxeles de la ROI, también se crea una copia para actualizar la imagen.

b) Segmentación

Desenfoque Gaussiano: Se aplica un desenfoque gaussiano para reducir el ruido de la imagen empleando un operador de vecindad a partir de una matriz de convolución

Conversión a escala de Grises: Luego se convierte a escala grises mediante la ecuación de la luminancia, dada por: $Y=0.299*R+0.587*G+0.114*B$

Resta de la imagen del gesto y fondo: Se realiza la resta de la imagen del gesto con la imagen del fondo, obteniendo el gesto en una matriz diferencia.

Conversión a Imagen Binaria: Se convierte la matriz diferencia a blanco y negro, empleando una segmentación binaria mediante un análisis de intensidad de píxeles, los que superen el límite se transforman a blanco y los que no a negro. El límite es la intensidad media de los píxeles y un umbral.

Desenfoque Mediano: Para reducir ruido en la imagen binaria se aplica un filtro de desenfoque mediano

Dilatación: Para disminuir regiones interiores se aplica dilatación que amplía la dimensión de las regiones, con un elemento matricial estructurante de tamaño 3

Erosión: La dilatación engrosa detalles del gesto, por tanto se realiza una erosión para acentuar detalles del gesto con elemento matricial estructurante de tamaño 15

c) Obtención del Contorno de la Mano

Para acelerar cálculos sin afectar al contorno primero se realiza reducción doble de la dimensión de la matriz.

Luego se detecta todos los contornos presentes en la imagen y finalmente se selecciona aquel de mayores dimensiones el cual corresponde al gesto y se almacena en una lista de puntos

Finalmente se transforman las coordenadas de los puntos del contorno a su correspondiente posición dentro de la matriz en la imagen original en la misma proporción que se redujo

d) Obtención de Proporciones y puntos de interés

Rectángulo que contiene la mano, los dedos y la palma: Se obtiene la zona rectangular en cuyo interior se encuentra el gesto de la mano, que se obtiene a partir del contorno calculando el rectángulo de mayor área.

El rectángulo de los dedos se obtiene a partir del rectángulo de la mano aplicando las proporciones Áureas.

Sea p la proporción entonces:

Falangeta = $1p$

Falangina Medio = $1,3p$

Falange = $2,3p$

Metacarpio = $3,6p$

Sus límites son:

superior = límite superior Rectángulo mano

derecho = límite derecho Rectángulo mano

izquierdo = límite izquierdo Rectángulo mano

inferior = $0.56097 \times$ límite inferior Rectángulo mano

El rectángulo de la palma se obtiene de igual forma aplicando las proporciones Áureas.

Sus límites son:

superior = $0.56097 \times$ límite inferior Rectángulo mano

derecho = límite derecho Rectángulo mano

izquierdo = límite izquierdo Rectángulo mano

inferior = límite inferior Rectángulo mano

Búsqueda de puntos candidatos a cima y meseta: A partir del contorno y el rectángulo de la mano se recorren todos los puntos, verificando su orientación y cambios de dirección, hallando puntos máximos y puntos mínimos. El procedimiento anterior no garantiza obtener todos los candidatos, para hallar otros puntos candidatos, se busca en el proceso anterior puntos contiguos en el mismo plano.

Selección de puntos: Se debe descartar falsos candidatos, en base a una distancia insuperable por la estructura de la mano, se analiza la distancia entre puntos vecinos hallando el punto central, se emplea la distancia entre dos puntos. Finalmente se compara el primer y segundo grupo de puntos candidatos y mediante un proceso de cálculo de distancia se descartan los puntos vecinos.

Análisis del punto cima del pulgar

Es probable que falte detectar puntos del dedo pulgar por su forma, para completar en base a proporción áurea de la mano se define que:

= 1 Mano abierta

Si $\text{alto} / \text{ancho de la mano} < 1$ Mano cerrada

> 1 mano con pulgar recogido

e) Interpretación del gesto

Modelo de la mano: Se construye un modelo simplificado de la mano empleando el enfoque de sistemas analizando la estructura, relaciones y proporción áurea como sistema.

Una mano tiene palma y 5 dedos que tienen falangeta, falangina, falange y metacarpo, genera gestos y tiene proporción áurea, los dedos tienen 3 estados

Dos dedos contiguos pueden estar juntos o separados, para determinarlo se calcula el ángulo entre ambos a partir del triángulo que forman con función arco coseno, si este es menor a $\pi/16$ los dedos están unidos, si no, están separados.

Un gesto está definido por el estado de los dedos y la unión o separación entre sus dedos

Codificación de los gestos: Se define el modelo:

$G = \{d1 \ d2 \ d3 \ d4 \ d5 \ e1 \ e2 \ e3 \ e4\}$

Donde:

G = Gesto

$d1 \ d2 \ d3 \ d4 \ y \ d5$ = posturas de meñique, anular, medio, índice y pulgar, con sus estados $d_i = \{0, 1, 2\}$ donde 0 extendido 1 doblado y 2 contraído

$e1 \ e2 \ e3 \ y \ e4$: Espacios entre: meñique y anular, anular y medio, medio e índice e índice y pulgar, con sus estados $e_i = \{0, 1\}$ donde 0 sin separación y 1 con separación

A cada combinación de estados le corresponde un solo gesto, los gestos codificados y su significado son almacenados previamente

Reconstrucción de postura e interpretación: Para cada punto cima de cada dedo, se aplica la proporción áurea desde la base hasta la cima, detectando el estado de cada dedo contraído, doblado o extendido, luego verificamos si cada par de dedos contiguos están unidos o separados, luego se codifica ese gesto.

Finalmente se hace una búsqueda del gesto codificado en los gestos almacenados y se recupera su significado.

3. RESULTADOS

Desarrollado el método se creó un prototipo para pruebas empleando las herramientas OpenCV, C#, Blender, Unity, Mixamo y MakeHuman

En hardware se empleó un equipo portátil de 2,1 GHz, video incorporado de 2GB, 8 GB RAM, disco 1TB y cámara VGA de 0.307MP.

Se implementó varias rutinas de código para todas las etapas, teniendo los siguientes resultados:

Obtención de imágenes: Se crearon varias rutinas para obtener las imágenes y obtener la ROI obteniendo el fondo y los gestos a partir del ROI

Desenfoque Gaussiano: Para ello se crearon rutinas de código y se empleó la función GaussianBlur de OpenCV.

Conversión a escala de Grises y eliminación de fondo: Para esta etapa se crearon varias rutinas de código, para convertir a escala de gris se empleó la función cvtColor y para la sustracción de fondo se empleó la función absdiff de OpenCV.

Binarización de la Imagen: Para la binarización del gesto se creó rutinas de código y se empleó threshold de OpenCV.

Desenfoque Mediano, erosión y dilatación de la imagen Binaria: Se crearon varias rutinas de código y se aplicó las funciones MedianBlur, erode y dilate de OpenCv.

Detección del Contorno: Para la reducción de matriz y detección del contorno se crearon rutinas de código y se empleó pyrDown y findContour de OpenCV.

Rectángulos de la mano: Se crearon rutinas de código para obtener los rectángulos de la mano, dedos y palma.

Detección de puntos límites: Se crearon rutinas de código para determinar todos los puntos candidatos, eliminar aquellos contiguos y para seleccionar los puntos límites.

Sección áurea para detectar y reconocer el gesto: Se crearon diversas rutinas de código para aplicar las proporciones áureas, para calcular el ángulo entre dedos, reconstruir el gesto, codificar y almacenar previamente los gestos, comparar e interpretar el gesto.

Para probar el método se realizó experimentos con 5 personas de ambos sexos, diferentes edades entre 5 a 41 años, con diferentes manos, fondos, iluminación y gestos.

Pruebas en Aplicaciones

Se desarrollaron 5 aplicaciones para probar la utilidad del método, con funciones específicas: control de objetos, control de cámara, manejo de avatar en entornos virtuales y aplicaciones de velocidad.

Se observó en los usuarios que: sienten alta satisfacción al usar una aplicación sin tocarla, es necesario una explicación previa para su uso, el manejo es más fácil para algunos usuarios y que no todos pueden realizar el mismo gesto

Tiempo y promedio de Reconocimiento

A partir de los experimentos se obtuvo los siguientes resultados:

% gestos reconocidos= 92,857%

% gestos no reconocidos= 7,143%

Tiempo promedio de reconocimiento 11,77 milésimas de segundo.

4. DISCUSIÓN

Principales diferencias con otros métodos: No se emplea algoritmos de machine learning, algoritmos evolutivos, redes bayesianas, máquinas de soporte vectorial, lógica difusa, algoritmos de agrupamiento, polígono convexo, cálculo de defectos, momentos Hu y otros, debido principalmente a su costosa implementación computacional, aunque estos pueden aplicarse con facilidad al método propuesto

No se emplea dispositivos especiales como: guantes, brazaletes, sensores, cámaras RGB, dispositivos Kinect, leap motion, etc, debido a que no están disponibles para la mayoría de los usuarios, son costosos o casi inaccesibles en ciertos lugares

Se aplica de forma novedosa el cálculo de proporciones áureas de la mano, las series de fibonacci, lucas y litler y el enfoque de Sistemas

Se emplean varias estrategias de optimización como: ROI, reducción de matrices, reducción de puntos de polígonos, puntos límites, etc. con el fin de obtener respuestas en tiempo real en aplicaciones y , se emplea una cámara monocular sin calibración previa accesible y disponible a todos los usuarios

Se realizan pruebas exitosas en programas con escenarios tridimensionales y que requieren alta velocidad de respuesta.

Comparación del rendimiento con otros métodos: Los porcentajes de reconocimiento y sus tiempos de otros métodos son:

77% Villa, Valencia& Berrio (2018)

85.08+-15.21% y 3+-1 ms Chung (2018)o

99% Sandipgiri(2019)

90,7% y 29.38 ms Motoche(2018)

96.3% Yang (2016)

91.7% Ronchetii(2016)

92% Rojas(2021)

93.16% Berru (2019)

68% Ortiz(2020)

99% Maqueda(2018)

82,15% Nasri(2021)

En base a la comparación con otros métodos se puede apreciar que el rendimiento del método es bastante satisfactorio en tiempo y porcentaje de reconocimiento, también muchos de los trabajos revisados no presentan el tiempo promedio de reconocimiento y no presentan experimentos en aplicaciones a diferencia del método-

5. CONCLUSIONES

El método permite el reconocimiento de gestos de la mano a partir de imágenes bidimensionales contribuyendo a la interacción humano computador

Aporta una solución innovadora aplicando la proporción áurea para la reconstrucción del gesto de la mano.

Aporta de forma innovadora con el uso de las series de Littler, Lucas y Fibonnaci para determinar la postura de los dedos y el reconocimiento del gesto de la mano.

Aporta de forma novedosa con la detección de puntos límites para determinar la posición de los límites de los dedos de la mano.

Permite a partir de una imagen, estimar y recuperar información de la mano, sin importar el tamaño de la mano.

Aporta con un modelo de almacenamiento y reconocimiento de gestos de la mano, que se modelo empleando el enfoque de sistemas.

Aporta con varias estrategias de aceleración en diferentes etapas del método como la reducción de matrices, detección reducida de puntos de contorno y otras estrategias que aceleran el reconocimiento, buscando un trabajo en tiempo real.

Identifica y describe los métodos de reconocimiento de gestos de la mano que permiten la interacción humano computador, sus características y estado actual.

Aporta a la solución del reconocimiento de gestos de la mano a partir de una imagen adquirida por una cámara monocular sin calibración previa.

Logra una alternativa altamente útil y atractiva para aplicarlo a sistemas de realidad virtual, sistemas de realidad aumentada, videojuegos, sistemas para personas con capacidades diferentes, etc.

Brinda un aporte altamente práctico, al proporcionar un prototipo construido sobre un engine altamente utilizable como es unity3d, el cual permite construir y obtener productos sobre diversas plataformas como Linux, Windows, Android, ios, Xbox, etc.

El rendimiento del método propuesto es bastante satisfactorio en tiempo y porcentaje de reconocimiento en comparación con los métodos comparados

6. REFERENCIAS

Andrade, F. J. (2016). Un enfoque inteligente para el reconocimiento de gestos manuales, Buenos Aires, Universidad Nacional del Centro.

Ansari, Farah Jamal(2017),Hand Gesture Recognition using fusion of SIFT and HoG with SVM as a Classifier, Septiembre 2017, International Journal of Engineering Technology Science and Research, Vol. 4, Pág. 913-922

Ariel, M. (2010). Seguimiento 3D de la mano, 2010

Arranz, F., Liu, Q., & Lopez, J. (2012). Interacción persona computador basado en el reconocimiento visual de manos. Madrid: Universidad Complutense de Madrid.

Arteaga Araceli Casans(2001) ,Aspectos Estéticos De La Divina Proporción, Universidad Complutense de Madrid, Tesis Doctoral, ISBN: 84-669-1867-1 , Madrid, España

Barros Rengifo Nicandro Rafael (2014), Algoritmos de Reconocimiento de Gestos de La Mano Basado en Tecnología Kinect, Universidad Técnica Estatal De Quevedo, Quevedo-Los Rios-Ecuador

Berru Novoa, Bryan Jos (2019) ,Reconocimiento de gestos estáticos del abecedario de la lengua de señas peruana utilizando cámaras de baja resolución, Universidad Peruana de Ciencias Aplicadas, Lima

Chacón Rosero Mario Andrés (2014), Secuencia Fibonacci y Productos: Análisis del comportamiento del consumidor frente al número áureo, Universidad San Francisco de Quito, Quito, Ecuador.

Chung Lui Edison Alejandro (2018), Modelo de Reconocimiento en tiempo real de gestos de la mano utilizando técnicas de Deep Learning y señales electromiográficas, Escuela Politécnica Nacional, Quito

Cruz Bautista Antonio Guadalupe(2019) Software de Realidad Virtual para la manipulación de datos LIDAR mediante el uso de Interfaz Natural de Usuario, Tesis doctoral en Tecnología Avanzada, Instituto Politécnico Nacional, Querétaro, 2019

Díaz Cortez Jorge(2017), Estética y Belleza, Proporción Áurea, Academia Chilena de Medicina y Cirugía Estética

García Cortés, D. C. (2014). Reconocimiento de Gestos de Manos como Mecanismo de Interacción Humano-Computador. Bogotá

FOTOGRAFÍAS DE LA INVESTIGACIÓN



Figura 1: Proceso de Reconocimiento de un gesto

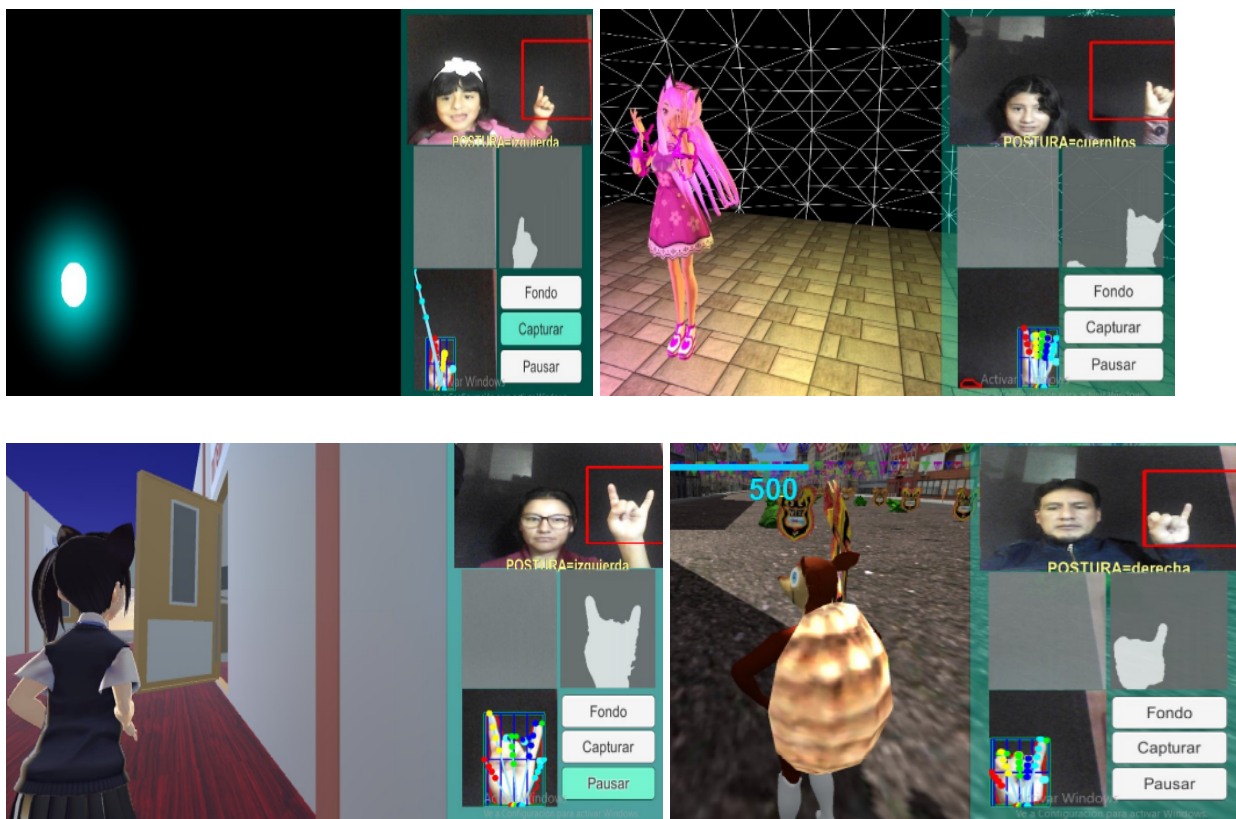




Figura 2: Pruebas en Aplicaciones

SOBRE EL AUTOR

Doctor en Ciencias de la Computación, Máster en Ciencias de la Computación con mención en Seguridad Informática y Software Libre, Máster en Educación Superior, cuenta con Diplomados en Preparación Evaluación y Gestión de Proyectos, Formación Basada por Competencias y Metodología de la Investigación Científica.

Ha sido director de la carrera Ingeniería Informática y actualmente es Presidente de la Sociedad Científica de docentes de la Universidad Nacional "Siglo XX", Coordinador del Instituto de Investigación y Desarrollo de Aplicaciones Informáticas IIDA, Director de la Revista Científica "Ciencia y Tecnología Informática", director de las publicaciones Memorias SOCID, Memorias Ciencia y Tecnología Informática y Docente Universitario Titular en la carrera Ingeniería Informática de la Universidad Nacional "Siglo XX", docente de pre y postgrado, organiza varios eventos académicos, sus principales líneas de investigación: Visión Artificial, Realidad Aumentada, Virtual y Extendida, Interacción Humano Computador, Desarrollo de Juegos, Seguridad Informática, Programación Competitiva y Software Libre. Publicó varios artículos y libros, recibió varios reconocimientos y premios.



Figura 3: Fotografía de presentación de la ponencia