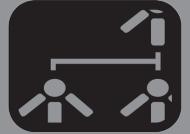


IDENTIFICACIÓN DE LA PUNTA DE LOS DEDOS DE LA MANO EN UN PLANO 2D BASADO EN KINECT

IDENTIFICATION OF THE TOE OF THE HAND IN A 2D
PLANE BASED KINECT



AUTOR

SAMIR CASTAÑO RIVERA
Magíster en Software Libre
*Universidad de Córdoba
Docente
Grupo SOCRATES
samircastano@yahoo.com
COLOMBIA

AUTOR

DANIEL JOSÉ SALAS ÁLVAREZ
Magíster en Informática
*Universidad de Córdoba
Docente
Grupo SOCRATES
danielsalas@correo.unicordoba.edu.co
COLOMBIA

AUTOR

RICARDO JOSÉ MONTES RODRÍGUEZ
Ingeniero de sistemas
*Universidad de Córdoba
Investigador Auxiliar
Grupo SOCRATES
ricardojosemontesrodriguez@gmail.com
COLOMBIA

INSTITUCIÓN

UNIVERSIDAD DE CORDOBA
UNICOR
Universidad Pública
Carrera 6 No. 76-103
Correo@correo.unicordoba.edu.co
COLOMBIA

RECEPCIÓN: Abril 23 de 2013

ACEPTACIÓN: Mayo 16 de 2013

TEMÁTICA: Ingeniería de software y Marco de trabajo.

TIPO DE ARTÍCULO: Artículo de Investigación Científica y Tecnológica.

RESUMEN ANALÍTICO

Este trabajo de investigación tiene como propósito presentar un sistema de reconocimiento de acciones o gestos realizados mediante la mano con el fin de mejorar la interacción maquina-hombre, basados en la tecnología Kinect, mediante el diseño de un modelado matemático y la implementación de un algoritmo computacional.

El algoritmo implementado permitió el reconocimiento de la ubicación de una mano y dedos dentro de un plano 2D utilizando la tecnología Kinect. Es importante resaltar la importancia de los resultados logrados debido a la mejora de la experiencia de interacción gestual por parte del usuario y a la posibilidad de reconocimiento de gestos como abrir o cerrar la mano.

PALABRAS CLAVES: Sistema de Reconocimiento, Computación Basada en Gestos, Kinect, Reconocimiento, Interface natural de usuario.

ANALYTICAL SUMMARY

This research work aims to present a recognition system of actions or gestures made by hand in order to improve the man-machine interaction based on Kinect technology, by designing mathematical modeling and implementation of an algorithm computer.

The implemented algorithm allowed the recognition of the location of a hand and fingers into a 2D plane using Kinect technology. It is important to highlight the importance of the results achieved due to improved interaction experience for the user gesture and gesture recognition ability as opening or closing the hand.

KEYWORDS: Kinect, Gesture-Based Computing, finger recognition, natural user interface.

INTRODUCCIÓN.

El informe HORIZON [1] del año 2011, marca como tendencia tecnológica, la Computación Basada en gestos, la cual presenta al cuerpo humano como enfoque principal de interacción con distintos dispositivos electrónicos de tal forma que partes del cuerpo humano, tales como las manos, pies u otros sirvan como medios de acción. Es de anotar que este tipo de interacción nace como una evolución a las interfaces de líneas de comando y graficas de usuario.

Estas tendencias futuristas pensadas años atrás, se denominan NUI [2] (interfaces de usuario natural) y hoy son una realidad del mundo tecnológico, que sorprenden por su versatilidad y potencialidades para el entretenimiento y la educación. Siendo este último uno de los sectores que puede aprovechar su simplicidad y naturalidad para facilitar el aprendizaje en niños y jóvenes [3]. Además, ser utilizado como dispositivos de entrada de gran interacción, dada las limitaciones de

los dispositivos de entradas tradicionales, tales como el teclado y mouse. En la actualidad una de las tecnologías que impulsa esta nueva forma de comunicación hombre-máquina es la denominada Tecnología KINECT [4] basada en un sensor de profundidad, cámara RGB entre otros componentes, desarrollada por la empresa Microsoft y lanzada al mercado norteamericano en noviembre del 2010. Recientemente ha salido al mercado LEAP MOTION DEVICE CONTROL [5], que es un sistema de reconocimiento de gestos y cuyo propósito principal es controlar todo tipo de sistema informático con los gestos de la mano sin utilizar la superficie táctil. No obstante, es necesario conocer el impacto de esta última tecnología en el mercado mundial.

Entre las funcionalidades de mayor relevancia de la tecnología KINECT se encuentra la capacidad de reconocimiento de articulaciones del cuerpo humano [6], lo cual es posible al obtener la estructura del esqueleto, sin embargo, una de las mayores dificultades que se tiene cuando se trabaja con esta tecnología es

que no permite el reconocimiento de la ubicación de los dedos de la mano, este hecho genera limitaciones que impiden una mayor versatilidad de esta herramienta, por lo tanto, este trabajo de investigación tiene el propósito de presentar un sistema de reconocimiento que incluye modelamiento, algoritmo y software para determinar de forma efectiva el centro de la palma de la mano y el reconocimiento de la forma de un dedo a partir de la técnica DTW (Alineamiento Temporal Dinámico) dentro de un plano 2D, como medio para mejorar la interacción gestual basada en KINECT.

Este documento se encuentra organizado de la siguiente forma, en primer lugar, esta introducción, posteriormente se muestran los estudios previos, trabajos e investigaciones relacionadas con el reconocimiento de manos, dedos y centro de la palma, en la tercera parte se presenta, los resultados, la cual incluye el modelo matemático y su aplicación mediante el planteamiento de un algoritmo y su validación mediante pruebas experimentales para analizar el grado de efectividad del algoritmo.

1. ESTADO DEL ARTE.

La interacción hombre-máquina, es un área de conocimiento que se deriva de las ciencias computacionales y se articula con las ciencias conductuales, se encarga de la planeación, diseño e interacción de los usuarios con los computadores y otros tipos de dispositivos electrónicos.

Una aplicación de la interacciones hombre-máquina, son las denominadas NUI (Natural User Interfaces), la cual incluyen la computación basada en gestos, que hoy se hace realidad gracias a los sistemas de reconocimiento representados en tecnologías, como los sensores Microsoft Kinect y Leap Motion Device Control, entre otros, las cuales permiten que los usuarios utilicen las manos y partes del cuerpos para interactuar con el sistema, permitiendo una gran versatilidad y uso efectivo de estas herramientas en las áreas de las educación [7] [8] [9], salud [10] [11] [12] [13] y el entretenimiento [14] [15], entre otras.

Los estudios recientes de los sistemas de interacción gestual [16] [17] [18] muestran las grandes potencialidades de estas tecnologías, sin embargo, persisten limitaciones, motivo por el cual es importante el diseño de nuevos algoritmos y modelos computacionales dotados de mayor optimización que permita mejorar la interacción de los usuarios con este tipo de sensores.

La propuesta basada en los datos de profundidad [19] toma una imagen obtenida a través del sensor KINECT y la implementación de procesos tales como

la segmentación de píxeles, obtención del contorno, trayectoria de filtros y envolvente convexa [20] para determinar la mano y la ubicación de los dedos.

En la revisión de literatura se destacan las propuestas [21] [22], las cuales se basan en los datos obtenidos a través de una cámara RGB y la implementación de procesos como la segmentación de píxeles, binarización, obtención del contorno y la envolvente convexa para determinar la mano y posibles dedos.

Entre las soluciones revisadas en la literatura se destaca trabajos que tiene como eje central el reconocimiento del contorno basado en la segmentación de una área de estudio[19] y los datos de profundidad obtenidos mediante el sensor, además de la determinación experimental de un umbral de distancia para reconocer o no la forma de la mano y por ende ubicar la punta de los posibles dedos basado en la identificación de la curvatura de una curva o en algunos casos denominado el algoritmo de k-curvatura. De igual manera los trabajos presentados [23] [24] [25] utilizan la técnica de búsqueda de la curvatura para determinar las posibles puntas de los dedos.

Es de anotar que las propuestas [19] [21] [23] presentan una serie de pasos o procesos en común tales como:

- Segmentación.
- Binarización.
- Detección Contorno.
- Casco Convexo

Este trabajo realiza contribuciones esenciales en el primer eje, es decir, se centra en utilizar los datos de la imagen de profundidad, la ubicación de la mano soportada en la librería OpenNI y la aplicación de propuestas en investigaciones anteriores, tales como segmentación, binarización, algoritmo de k-curvatura. El segundo eje presenta al centro de la palma como un mecanismo para reducir la cantidad de puntos a explorar como posibles dedos y la aplicación del algoritmo de reconocimiento de patrones DTW (Dynamic Time Wrapping) para validar la forma o patrón correspondiente a un dedo.

2. RESULTADOS.

2.1 MODELO MATEMÁTICO PARA DETERMINAR LA UBICACIÓN DEL CENTRO DE LA PALMA DE LA MANO.

A continuación se describe el modelo utilizado para determinar el centro de la palma de la mano, siguiendo de manera sistemática los siguientes pasos:

Sea:

CP: coordenadas (x,y) del contorno imagen binarizada.

CI: coordenadas (x,y) internas al contorno de la imagen binarizada.

A = $\{(x,y): (x,y) \in CP\}$, min: menor valor de un conjunto finito de valores.

B = $\{(x, y): (x,y) \in CI\}$, max: máximo valor de un conjunto finito de valores.

Si $(x,y) \in B \Rightarrow d((x,y), A) = \min \{d(x, y), (z,w): (z,w) \in A\}$

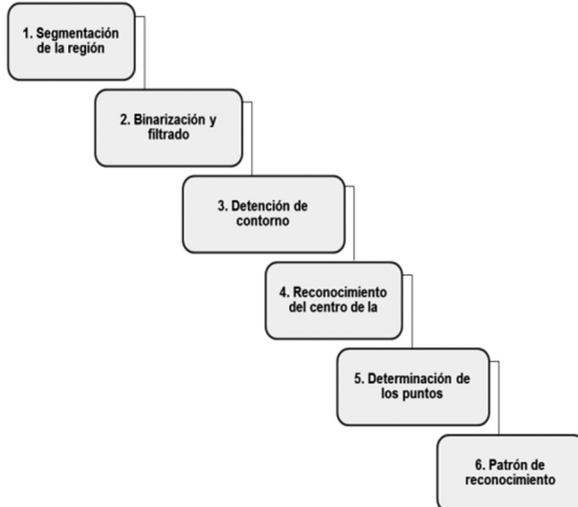
El centro de la palma de la mano viene dado por:

P = $\max\{d(x, y), A\}: (x,y) \in B\}$.

2.2 PROCESO ALGORITMICO

El proceso algoritmo para el reconocimiento se llevó a cabo los siguientes pasos fundamentales:

FIGURA 1. Algoritmo de reconocimiento.

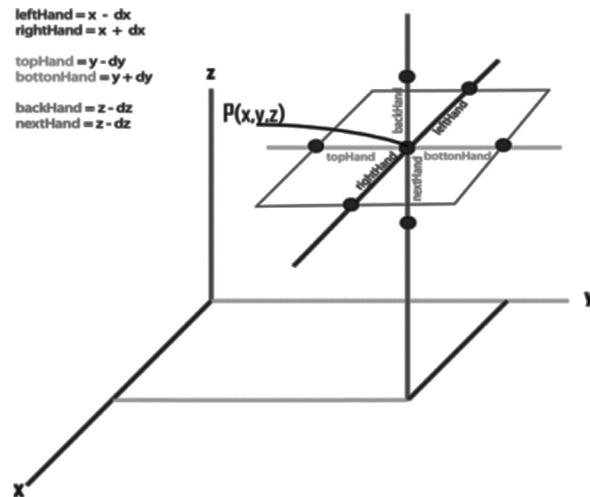


2.3 SEGMENTACIÓN DE LA REGIÓN CORRESPONDIENTE.

La segmentación de una imagen de profundidad permite identificar o aislar grupos de píxeles que constituyen un objeto mediante términos como continuidad, proximidad y similitud, para su posterior análisis.

Una de las técnicas de segmentación es la denominada segmentación por regiones, en donde se busca determinar un grupo de píxeles que corresponden a un objeto y la región a la cual pertenecen. Para determinar la región y el grupo de píxeles que corresponde a la mano, utilizamos la imagen de profundidad obtenida a través del sensor KINECT, una coordenada $P(x, y, z)$ obtenida mediante la librería OpenNI la cual determina el punto central de nuestra área de estudio así como la inclusión de 3 valores dx, dy y dz obtenidos experimentalmente, referentes a distancias próximas al punto P.

FIGURA 2. Área de la mano delimitada por el punto $P(x, y, z)$.



2.4 BINARIZACIÓN Y FILTRADO.

La binarización es un proceso por el cual se toma una imagen y se obtienen dos (2) grupos de píxeles que cumplen cierto criterio o condición, todo esto con el objetivo de lograr una nueva imagen que contenga dos posibles valores por lo general blanco y negro y de este modo lograr facilitar la identificación de objetos en una imagen.

Como se puede observar en la Figura. 2, claramente se ilustra claramente la existencia de dos grupos, un grupo de píxeles de color blanco que corresponde a la mano y un grupo de color negro que no, sin embargo, no basta con realizar este paso para obtener una imagen óptima, para ello es necesario la aplicación de un filtro de paso bajo que suavice la imagen y de este modo reduzca los ruidos generados en los bordes de la misma.

FIGURA 3. Segmentación y binarización de la imagen de la mano.



2.5. DETECCIÓN DEL CONTORNO.

El proceso de detección de contorno en una imagen permite el aislamiento de puntos para facilitar la identificación o características de un objeto.

Para realizar este proceso se utilizó la imagen binarizada y la librería OPENCV con el objeto de obtener la lista de puntos que pertenecen al contorno. En la Figura. 3 se puede observar el contorno de la mano.

FIGURA 4. Detección del contorno utilizando la librería OPENCV.



2.6 RECONOCIMIENTO DE LA MANO Y DEDOS.

La ubicación de los dedos se determinó aplicando dos premisas:

- Búsqueda de puntos convexos [26] dentro del contorno de la mano basado en el centro de la palma de la mano.
- Descarte de puntos candidatos de posibles dedos.

2.7 DETERMINACIÓN DE PUNTOS CONVEXOS.

Sea $P(x, y)$ el centro de la palma de la mano obtenida mediante el modelo matemático propuesto y P_c la lista de puntos (x, y) del contorno de la mano. La búsqueda de puntos convexos (posibles dedos) se realiza teniendo en cuenta aquellos puntos que están dentro del contorno (P_c) y no pertenece al borde limitado por la circunferencia que tiene como centro la palma de la mano, de esta manera se puede reducir el número de puntos analizar como posible dedo.

Para determinar la ubicación de un punto candidato se especifica la siguiente expresión:

Sea d una función con dos parámetros p_1 y p_2 para calcular la distancia euclidiana entre los puntos.

Sea $PC = (P_{cx}[i], P_{cy}[i])$ punto candidato a dedo.

Sea $PA = (P_{cx}[i-1], P_{cy}[i-1])$ y $PP = (P_{cx}[i+1], P_{cy}[i+1])$ punto anterior y punto posterior al punto candidato P .

Sea d_e , distancia determinada experimental.

Si $(d((P_{cx}[i-1], P_{cy}[i-1]), PC) = d_e$ y $d((P_{cx}[i+1], P_{cy}[i+1]), PC) = d_e$ \Rightarrow

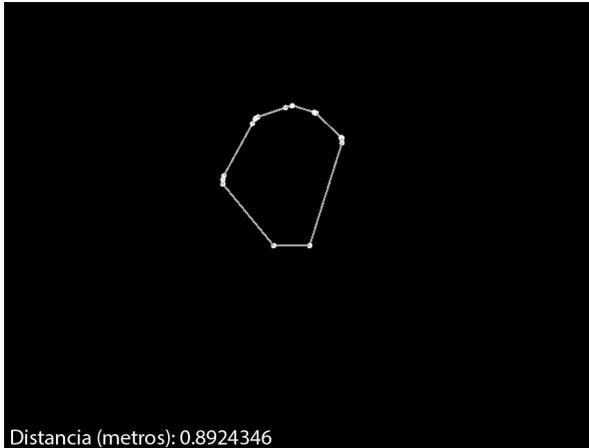
Entonces PC es un punto candidato a dedo.

Es de anotar que los 3 puntos PC, PA, PP no deben estar alienados.

2.8 PATRÓN DE RECONOCIMIENTO.

Las investigaciones [19] [21] se enfocan en la técnica casco convexo para determinar la punta de los dedos mediante los vértices que encierran el casco convexo, esto implica que en algunos casos se determinen más de 5 vértices por la propia deformidad del contorno de una mano en términos de los datos obtenidos mediante la imagen de profundidad Figura 5.

FIGURA 5. Casco Convexo.



Esta propuesta busca disminuir el grado de error presente en esta técnica y mejorar la ubicación de la punta de los dedos y se basa en tomar cada punto candidato obtenido mediante la determinación de puntos convexos y la búsqueda de un número específico de puntos posteriores y anteriores a este, todo en relación a los datos especificados dentro de un patrón representado en la Figura 5.

La determinación de si un punto candidato se considera como punta de un dedo, se realiza mediante dos fases:

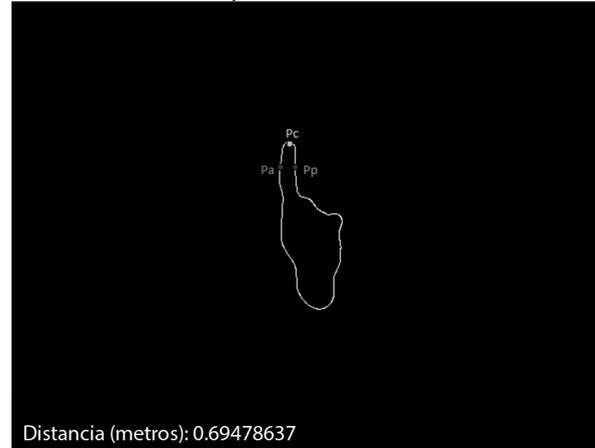
- 1) Trazado de puntos
- 2) Reconocimiento de Forma.

Trazado de puntos: Permite unir los puntos anteriores y posteriores al punto candidato y armar una forma geométrica.

Donde $P(c)$ indica el punto candidato a punta de dedo. $P(a)$ indica el inicio de puntos anteriores a $P(c)$. $P(p)$ indica el punto final de la lista de puntos posteriores. En la

Figura 6 se observa que el patrón de un dedo se obtiene mediante la unión de todos los puntos a partir de $P(a)$, $P(c)$ hasta llegar $P(p)$.

FIGURA 6. Puntos que conforma un dedo.



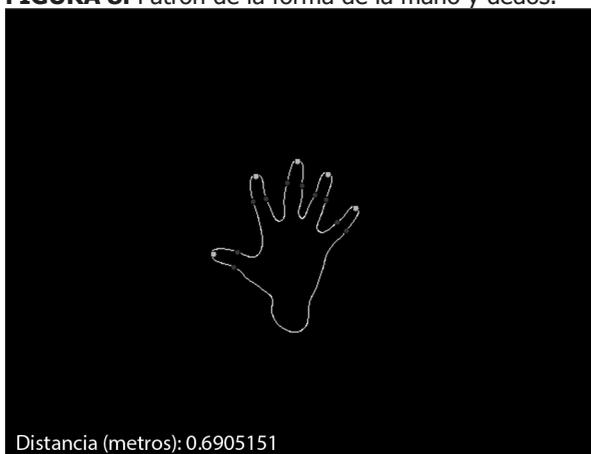
Reconocimiento de forma: para realizar este aspecto se utilizó la técnica DTW (Dinámico Time Wrapping) para verificar el grado de similitud entre los puntos definidos dentro de un patrón específico (ver Figura 7) y los puntos obtenidos mediante del trazado de puntos (fase 1).

La aplicación de la técnica se basa en tomar la forma de un dedo y obtener el grado de similitud entre todas las formas de un dedo de la mano Figura 8. El algoritmo de reconocimiento toma aquellos patrones cuyo grado de similitud oscile entre 0.8 y 1.0 como punta de un dedo.

FIGURA 7. Patron de la forma de un dedo.



FIGURA 8. Patron de la forma de la mano y dedos.



3. PRUEBAS

En esta etapa se realizaron las pruebas relacionadas con la efectividad del algoritmo para el reconocimiento de la mano y la ubicación de la punta de los dedos con relación a distancias obtenidas en metros (m) hacia el sensor Kinect.

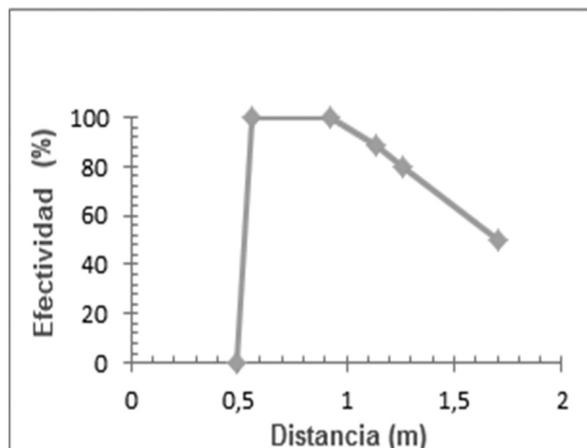
A continuación, se presentan los resultados alcanzados:

TABLA 1. Pruebas de la distancia funcional de la mano al sensor.

Distancia metros al sensor	Efectividad %
0,49	0
0,56	100
0,92	100
1,14	89
1,26	80
1,70	50

Al realizar una gráfica con la información obtenida se puede observar las distancias que concurren en un mejor funcionamiento del algoritmo y su tendencia de efectividad con relación a la distancia del sensor.

FIGURA 9. Resultado de las pruebas distancia efectividad.



De acuerdo a los resultados obtenidos y relacionado en la Figura 9 se observas que existe un rango de distancias entre 0,56 (m) hasta 0,92 (m) las cuales se pueden denominar distancias optimas en donde el grado de efectividad es de un porcentaje del 100, esto debido a factores tales como la luminosidad del sitio, área de visión del sensor y claridad de la imagen de profundidad. De igual forma existen una grado de disminución en la efectividad que se ve reflejado entre un rango de distancia entre 1,14 (m) hasta 1,26 (m) estos datos reflejan la limitación del sensor por aspectos como intensidad y claridad de pixeles de la imagen y por el ende la capacidad de acierto del algoritmo en determinar la ubicación de los posibles dedos de la mano en la Fig. 10, Fig. 11 y Fig. 12 se muestran algunas pruebas realizadas a ciertas distancias.

Figura 10. Ejecucion del algoritmo a 0,49m no tiene vision de la mano no tiene efectividad el algoritmo.



FIGURA 11. Ejecución del algoritmo a 0,88m el algoritmo tiene una efectividad del 100% detecta los 5 dedos de la mano.



FIGURA 12. Ejecución del algoritmo a 1,74m el algoritmo tiene una efectividad del 50% por lo que no es optimo al detectar los dedos de la mano.



4. CONCLUSIONES.

En la actualidad distintos enfoques establecen aspectos como la segmentación, binarización y búsqueda del casco convexo como medios para determinar la ubicación dedos de la mano, a partir de esto tomando como referencias aspectos establecidos en distintas investigaciones nuestra intención es dar un aporte en tomar la ubicación del centro de la palma de la mano, la aplicación de la técnica de reconocimiento de secuencias DTW para demostrar la importancia de esta como herramienta en la tarea de ubicar los posibles dedos de una mano basado en los datos de profundidad obtenidos mediante el sensor kinect.

Es preciso anotar que debido a factores como luminosidad y distorsión de contornos, entre otros, es notable el grado de disminución en el reconocimiento de la mano y por ende la incorrecta ubicación de los dedos y centro de la palma. De igual forma es importante mencionar que con base al modelo propuesto es posible llevar a cabo una identificación óptima de la mano y la ubicación de cada dedo en la misma dentro de un plano de 2 dimensiones dentro de cierto rango de distancia especificados en la Tabla 1, permitiendo a futuro reconocer en tiempo real acciones gestuales como abrir, cerrar y levantar un dedo de la mano. Sin embargo, una de las mayores debilidades del algoritmo es la incapacidad de identificar con certeza si es el dedo pulgar o menique es el que se encuentra levantado sin importar la posición de la mano. A continuación se presentan algunas de las ventajas y desventajas del algoritmo:

Ventajas

- Reconocimiento de la mano, forma y ubicación.
- Identificación de posibles dedos.
- Capacidad de identificación en tiempo real de acciones como levantar un dedo.
- Capacidad de reconocer gestos como abrir o cerrar la mano.

Desventajas

- Alta dependencia de la luminosidad.
- Incapacidad de identificar con exactitud que dedo es (pulgar o menique).

Los trabajos futuros se enfocaran en la aplicación de técnicas en la identificación de patrones que permitan mejorar y estabilizar el reconocimiento de gestos en tiempo real dentro de un plano 3D, con el objeto de llegar a tener mayor preciso en cada movimiento presente en la mano y los dedos que la conforman.

5. REFERENCIAS.

- [1] HORIZON REPORT (2011), The 2011 Horizon Report is a collaboration between the new media Consortium and EDUCAUSE learning Initiative an EDUCAUSE Program.
- [2] MICROSOFT, NUI: Natural User Interfaces (online), cited (29 de may., 2013). <http://research.microsoft.com/en-us/collaboration/focus/nui/default.aspx>

- [3] BALDASSARRI, Sandra, CERESO, Eva, MARCO, Javier. (2007): Desarrollo de interfaces naturales para aplicaciones educativas dirigidas a niños.
- [4] Lowe Scott, Microsoft Kinect Review (online) nov 2012 (cited 29 de mayo. (2013). <http://www.ign.com/articles/2010/11/04/microsoft-kinect-review>
- [5] TERDIMAN, D. (2013). Leap Motion: 3D hands-free motion control, unbound.
- [6] BUXTON, B. (2010). CES 2010: NUI with Bill Buxton. Microsoft Research.
- [7] ANACLETO, J., SILVESTRE, R., SANTANA, B., and FELS, S. (2012). Therapist-centred design of NUI based therapies in a neurological care hospital. 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2318–2323. doi:10.1109/ICSMC.2012.6378087
- [8] CHIANG, I.-T., TSAI, J.-C., and CHEN, S.-T. (2012). Using Xbox 360 Kinect Games on Enhancing Visual Performance Skills on Institutionalized Older Adults with Wheelchairs. 2012 IEEE Fourth International Conference On Digital Game And Intelligent Toy Enhanced Learning, 263–267. doi:10.1109/DIGITEL.2012.69
- [9] MUNSON, J., and PASQUAI, P. (2012). The Promise and the Perils, (June), 89–91.
- [10] PARAJULI, M., and SHARMA, D. (2012). Senior health monitoring using Kinect. 2012 Fourth International Conference on Communications and Electronics (ICCE), 309–312. doi:10.1109/CCE.2012.6315918
- [11] REGO, P. A., MOREIRA, P. M., and REIS, L. P. (n.d.). Natural User Interfaces in Serious Games for Rehabilitation.
- [12] SHMINAN, A. S., TAMURA, T., and HUANG, R. (2012). Student Awareness Model based on student affective response and generic profiles. 2012 IEEE International Conference on Information Science and Technology, 675–681. doi:10.1109/ICIST.2012.6221731
- [13] THAKKAR, V., SHAH, A., THAKKAR, M., JOSHI, A., and MENDJOGE, N. (2012). Learning Math Using Gesture. International Conference on Education and e-Learning Innovations, 1–3. doi:10.1109/ICEELI.2012.6360617
- [14] ZHANG, L., HSIEH, J.-C., TING, T.-T., HUANG, Y.-C., HO, Y.-C., and KU, L.-K. (2012). A Kinect based Golf Swing Score and Grade System using GMM and SVM. 2012 5th International Congress on Image and Signal Processing, (Cisp), 711–715. doi:10.1109/CISP.2012.6469827
- [15] ZHANG, S., HE, W., YU, Q., and ZHENG, X. (2012). Low-cost interactive whiteboard using the Kinect. 2012 International Conference on Image Analysis and Signal Processing, 1–5. doi:10.1109/IASP.2012.6424988
- [16] PICKERING, C. (2010). Research Study of Hand Gesture Recognition Technologies and Applications for Human Vehicle Interaction
- [17] ALSHEAKHALI HAND, M. (2011). Gesture Recognition System
- [18] CHAKRABORTY, P. (2008). Hand Gesture Recognition: A Comparative Study
- [19] FRATI, V., and PRATTICCHIZZO, D. (2011). Using Kinect for hand tracking and rendering in wearable haptics. 2011 IEEE World Haptics Conference, 317–321. doi:10.1109/WHC.2011.5945505
- [20] EILBERG, E. (2004). Convex Hull Algorithm.
- [21] HOJOON, P. (2008). A Method for Controlling Mouse Movement using a Real-Time Camera.
- [22] RAHEJA, J. L., CHAUDHARY, A., and SINGAL, K. (2011). Tracking of Fingertips and Centers of Palm Using KINECT. 2011 Third International Conference on Computational Intelligence, Modelling & Simulation, 248–252. doi:10.1109/CIMSim.2011.51
- [23] JAMES, D. (2012). Finger and Gesture Recognition with Microsoft Kinect.
- [24] THIAGO, R. and PELLEGRINO, S. (2010). An Analysis of Features for Hand-Gesture Classification
- [25] TANAKA, J. et al. (2012). Hand Controller: Image Manipulation Interface Using Fingertips and Palm Tracking with Kinect Depth Data.
- [26] CHETVERIKOV, D. (1999). Detection of High Curvature Points in Planar Curves. <http://visual.ipan.sztaki.hu/corner/node7.html>, visitado en diciembre de 2012.