

KINECT INTERACTION SYSTEM APPLIED TO PROCESS MANIPULATION

SISTEMA DE INTERACCIÓN CON KINECT APLICADO A MANIPULACIÓN DE PROCESOS

**MSc. Camilo Pardo-Beainy*, MSc. Edgar Gutiérrez-Cáceres*,
Ing. Daniel Pardo*, Ing. Milton Medina*
MSc. Fabián Jiménez****

***Universidad Santo Tomás**

Facultad de Ingeniería Electrónica, Grupo de Investigación GIDINT,
Tunja, Boyacá Colombia.

Tel.: 57-8-7440404, Ext. 5610

E-mail: camilo.pardo@usantoto.edu.co, eagutierrezca@gmail.com,
pardo_1104@hotmail.com, mer1156@hotmail.com

****Universidad Pedagógica y Tecnológica de Colombia**

Escuela de Ingeniería Electrónica. Ave. Central del Norte 39-115, Tunja, Boyacá,
Colombia. Teléfono: (+578) 7405626 Ext: 2565. E-mail: fabian.jimenez02@uptc.edu.co

Abstract: This paper resumes main aspects related with a construction of an interface characterized by employing a RGB and Depth image of a Kinect in order to follow movements in the environment with specific characteristics. The system works as an interface in where the user controls different processes and applications using motion commands, Kinect reads into this data and the program is responsible for filtering, selecting and evaluating the information in order to perform an action.

Keywords: Kinect, Digital Image Processing, Depth Map, Process Manipulation, Tracking Process.

Resumen: Este artículo resume los principales aspectos relacionados con la construcción de una interfaz caracterizada por emplear una imagen RGB y de profundidad de un Kinect para seguir movimientos en un entorno con características específicas. El sistema funciona como una interfaz en la que el usuario controla diferentes procesos y aplicaciones utilizando comandos de movimiento, Kinect lee los datos y el programa es responsable de filtrar, seleccionar y evaluar la información para realizar una acción.

Palabras clave: Kinect, procesamiento digital de imágenes, Mapa de Profundidad, Manipulación de Procesos, Proceso de seguimiento.

1. INTRODUCCION

En el área de la automatización de procesos y en el sector industrial, la comunicación existente entre hombre y maquina (interfaz usuario-maquina), representa uno de los ejes de más alto interés, en cuanto a control y manipulación de procesos se refiere; la comunicación en estos procesos exige que el usuario pueda manipular e

interactuar fácilmente con la máquina (López et al., 2013; Niño et al., 2020), en donde la interfaz brinde las herramientas necesarias para poder efectuar cualquier acción o cambio sobre el sistema (Pardo-Beainy et al., 2014).

En el presente trabajo, se detallan las actividades realizadas con el sensor Kinect orientadas a determinar las capacidades técnicas y de

funcionamiento del sensor, con el fin de poder diseñar una interfaz para hacer seguimiento de movimientos de determinados objetos o extremidades del cuerpo con el objetivo de proporcionar una interfaz sencilla, precisa y atractiva, en donde el usuario pueda efectuar comandos de acción sobre un sistema determinado.

Para el sistema se contemplaron dos fases importantes, en la primera se empleó un programa de verificación y reconocimiento del kinect realizado en LabVIEW en donde se disponen los subVI's necesarios para captar las imágenes y mapas de profundidad proporcionados por el sensor. En la segunda fase se contempla el filtrado, la detección de coordenadas por medio de mapas de profundidad y seguimiento de elementos de forma circular en el rango de trabajo.

2. CARACTERÍSTICAS DEL SENSOR KINECT

El Kinect se ha convertido en un importante sensor 3D y recibe mucha atención debido a su rápido sistema de reconocimiento de la posición humana. El bajo costo, rentabilidad y velocidad de captura de las medidas de este sensor, hace del Kinect un dispositivo de reconocimiento en Robótica, reconstrucción de escenas 3D y reconocimiento de objetos(Vivas et al., 2017).

2.1 Kinect como dispositivo de Sensado 3D

Este dispositivo se caracteriza por ser un equipo alargado conectado a una especie de base, esto con el fin de disponerlo en una posición horizontal. El dispositivo incorpora:

- Una cámara RGB.
- Un Proyector infrarrojo (que funciona como cámara de profundidad).
- Un micrófono multi-array bidireccional.

En conjunto, estos elementos permiten capturar imágenes y movimientos de los cuerpos captados, además de ofrecer reconocimiento facial y aceptar comandos de voz (Ivorra Martínez, 2015).



Fig. 1. Características físicas del sensor (Eduardo et al., 2014).

2.2 Imagen IR

La cámara de IR, es empleada para observar y decodificar los patrones de proyección IR para triangular la escena 3D. La cámara exhibe distorsiones radiales y tangenciales(Smisek et al., 2011).

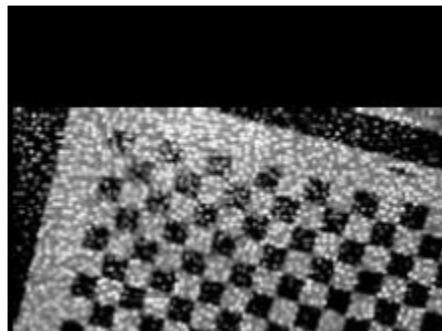


Fig. 2. Imagen IR proporcionado por el Kinect. (Štrbac, Markovi, et al., 2012)

2.3 Imagen RGB

La cámara RGB de Kinect realiza 30 capturas por segundo según las especificaciones. Es decir, 1 frame cada 30 ms utilizando la resolución (formato) estándar del sensor: 640×480 RGB.



Fig. 3. Imagen RGB proporcionada por el Kinect. (Štrbac, Markovi, et al., 2012)

2.4 Imagen de Profundidad

Uno de los aspectos más importantes del Sensor Kinect es la escena de profundidad. La imagen de profundidad es construida mediante la triangulación entre la imagen IR y el proyector. Las figuras 3 y 4 presentan un ejemplo de un sistema de sensado RGB y de profundidad respectivamente utilizando el sensor Kinect (Štrbac, Markoviü, et al., 2012).

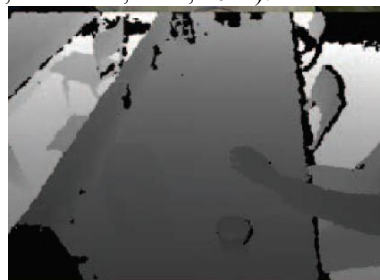


Fig. 4. Imagen de profundidad proporcionada por el Kinect. (Štrbac, Markovi, et al., 2012)

El sensor de Kinect adquiere imágenes de video con un sensor CMOS de colores a una frecuencia de 30 Hz, en colores RGB de 32-bits y resolución VGA de 640×480 píxeles. El canal de video monocromo CMOS es de 16-bit, resolución QVGA de 320×240 píxeles con hasta 65,536 niveles de sensibilidad (Štrbac, Markovič, et al., 2012).

Para calcular distancias entre un cuerpo y el sensor, el sensor emite un haz láser infrarrojo que proyecta un patrón de puntos sobre los cuerpos cuya distancia se determina. Una cámara infrarroja capta este patrón y por hardware calcula la profundidad de cada punto. El rango de profundidad del sensor de Kinect está entre 0.4m y 4 m(Mathe et al., 2012).

La cámara de Kinect funciona con hardware y software propios para el reconocimiento de imagen.

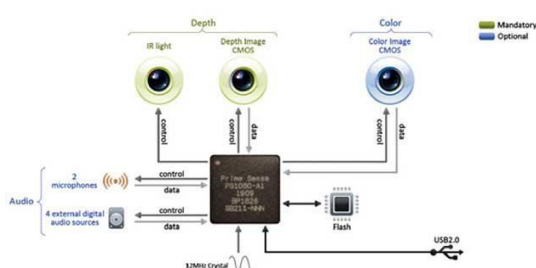


Fig. 5. Hardware y componentes.
(Mathe et al., 2012)

La cámara tiene dos funciones principales:

- Generar un mapa en 3D de la imagen que tiene en su campo visual.
- Reconocer humanos en movimiento entre los objetos de la imagen a partir de diferentes segmentos de las articulaciones del cuerpo y un esquema en escala de grises(Mathe et al., 2012).

En un amplio campo visual con objetos, la cámara Kinect trata de reconocer a qué distancia están los distintos objetos, distinguiendo movimientos en tiempo real.

El sensor Kinect puede llegar a distinguir la profundidad de cada objeto con una resolución de 1 centímetro y las estimaciones de la altura y anchura con una exactitud de aproximadamente 3 milímetros. El hardware está compuesto por la cámara y el proyector de luz infrarroja (ver figura 6), añadido al firmware y a un procesador que utiliza algoritmos para procesar las imágenes tridimensionales (Mathe et al., 2012).

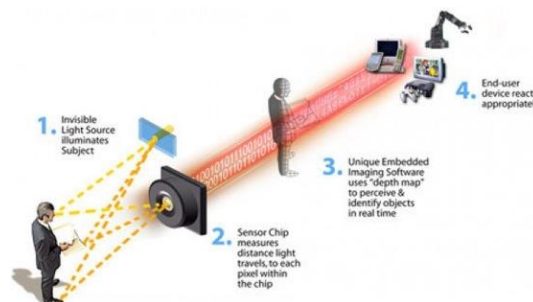


Fig. 6. Proceso de adquisición de datos.
(Mathe et al., 2012)

3. ALGORITMO PROPUESTO

3.1 Verificación y reconocimiento del dispositivo kinect en LabVIEW.

Para poder trabajar Kinect y LabVIEW, se hace necesario adquirir los drivers del sensor para que pueda ser reconocido por el equipo. En la figura 7 y la tabla I, se detallan los drivers necesarios para instalar el dispositivo:

Name	Date modified	Type	Size
amd64	07/06/2012 17:11	File folder	
ia64	07/06/2012 17:11	File folder	
x86	07/06/2012 17:11	File folder	
Xbox_NUI_Audio.cat	12/11/2010 22:23	Security Catalog	1 KB
Xbox_NUI_Audio.inf	12/11/2010 22:23	Setup Information	8 KB
Xbox_NUI_Camera.cat	12/11/2010 22:23	Security Catalog	1 KB
Xbox_NUI_Camera.inf	12/11/2010 22:23	Setup Information	8 KB
Xbox_NUI_Motor.cat	12/11/2010 22:20	Security Catalog	1 KB
Xbox_NUI_Motor.inf	12/11/2010 22:20	Setup Information	8 KB

Fig. 7. Drivers del sensor.

Tabla 1: Driver's de Instalacion

Drivers	Dispositivo
Xbox_NUI_Audio	Audio
Xbox_NUI_Camera	Video
Xbox_NUI_Motor	Motor (Kinect)

Posteriormente, es necesario adquirir los subVIS que permiten adquirir los datos proporcionados por el Kinect en LabVIEW, a continuación se detallan los Vis más importantes del proyecto:

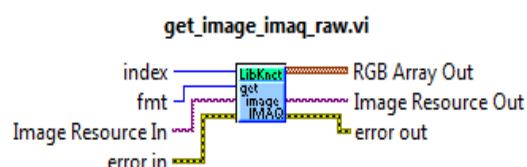


Fig. 8. SubVI de acquisition de imagen RGB.

Get_image_imaq, proporciona la imagen RGB entregada por el Kinect además de un arreglo de datos de la imagen adquirida.

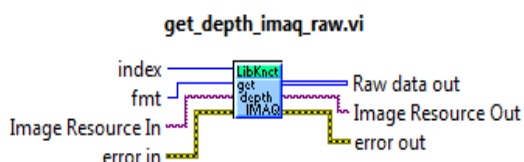


Fig. 9. SubVI de adquisición de imagen de profundidad.

Get_depth_img, proporciona un mapa de profundidad en donde se detalla los objetos más cercanos y lejanos al sensor.

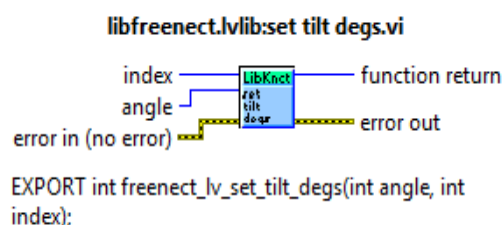


Fig. 10. SubVI de configuración de ángulo de vista del Kinect.

El Set tilt degs, permite ajustar el Ángulo del servomotor que permite posicionar al Kinect al gusto del usuario.

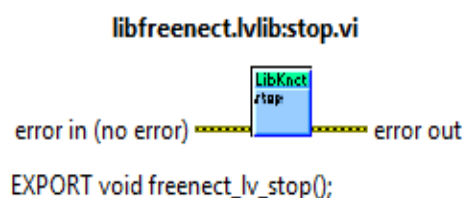


Fig. 11. SubVI de parada de datos.

Stop, permite detener la ejecución de cada uno de los subVIs empleados anteriormente.

Para poder verificar el funcionamiento de cada uno de los subVIs, se diseñó un programa (titulado "Point Cloud") caracterizado por adquirir y visualizar todos los datos que puede entregar el Kinect, el cual se detalla en la figura 12.

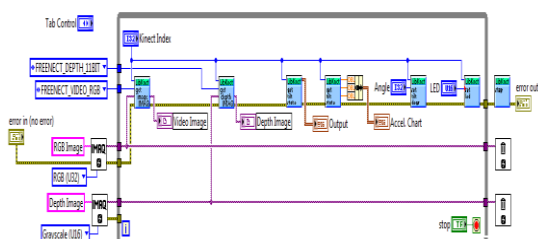


Fig. 12. Programa de verificación y prueba del Kinect.

Este programa proporciona tanto la imagen RGB como el mapa de profundidad que se requiere para el proyecto, además de otras funciones como el control del Ángulo del Kinect, las características del acelerómetro que incorpora el sensor, entre otras. En figura 13 se detalla la Imagen RGB y la de profundidad:

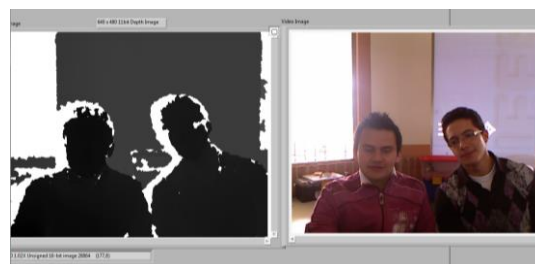


Fig. 13. Imágenes (RGB y Depth) proporcionada por el Kinect.

3.2 Filtrado y detección de coordenadas por medio de mapas de profundidad

Teniendo el mapa de profundidad, se aprovecha el arreglo de datos generado por el SubVI de profundidad, para determinar los elementos más cercanos y más lejanos de la imagen, de tal forma que se puedan definir rangos de trabajo en donde se contemplen solo los datos de información de la profundidad requerida (Bevly, 2012).

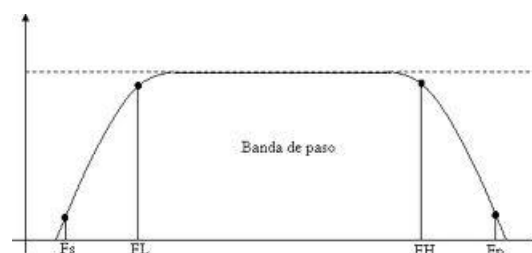


Fig. 14. Rango de trabajo para la adquisición de imagen.

En la figura 14, se puede detallar el rango mencionado, solo los objetos que se encuentren dentro serán detectados. Para la selección de rangos, se contemplaron los puntos en donde el sensor, en conjunto con la interfaz, obtenía mejores mapas de profundidad y se realizaba una mejor detección; el mapa de profundidad se caracteriza por brindar un valor numérico proporcional a la distancia que hay entre los objetos más cercanos y más lejanos al sensor.

El rango de profundidad de medición se encuentra entre 1,2 m y 3,5 m; lo que implica un valor numérico de profundidad en el programa entre 10000 y 70000. Por tal motivo, el rango de trabajo definido, comprende los valores presentados en la tabla 2.

Tabla 2. Valores de distancia y valor numérico en el programa para la profundidad en el rango de trabajo.

Distancia Min. Del Rango	Distancia Max. Del Rango	Valor Min. De Profundidad.	Valor Max. De Profundidad.
1,5 m.	2,0 m.	20000	25000

Para los objetos que se encuentren fuera dicho rango se mostrará en color blanco la pantalla y los que están dentro del rango, se presentan en color negro, tal como se aprecia en la figura 15.

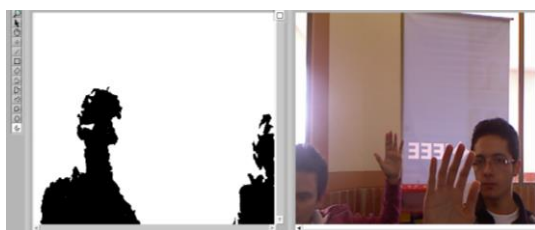


Fig. 15. Rango de trabajo funcional

Con dicho parámetro definido, el siguiente paso consiste en el reconocimiento de un objeto en dicho rango, por tal motivo se incorporó una etapa de filtrado y detección a la imagen, en donde se aplica un filtrado de tipo laplaciano para suavizar los bordes (en la figura 16 se pueden detallar los efectos de este tipo de filtro) y contornos de la imagen, luego se diseñó una etapa en donde se detectan todos los objetos que tengan forma similar a la de un círculo (Štrbac, Marković, et al., 2012; Um et al., 2011).

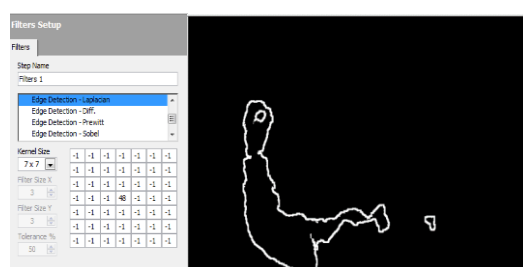


Fig. 16. Filtro Laplaciano empleado.

El programa se encarga de hacer seguimiento a todos los objetos con dicha forma y obtener la coordenada de dicho círculo, como se puede apreciar en la figura 17; con el valor de la coordenada se obtiene un valor numérico que facilita el acondicionamiento de los datos, además de brindar al usuario la posibilidad de poder ubicarse en la interfaz e interactuar con el sistema.

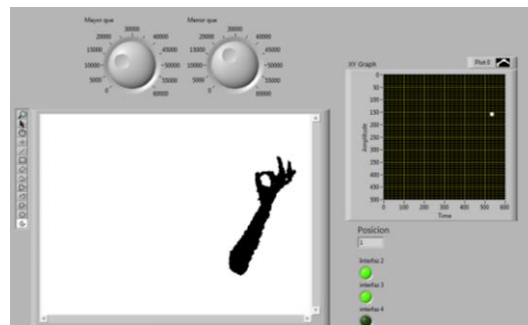


Fig. 17. Filtro aplicado sobre el entorno.

Teniendo la coordenada, se realizan una serie de arreglos y condiciones con el fin de dividir la pantalla en varias secciones (inicialmente en cuatro partes), esto con el fin de comparar el valor de la coordenada y el de la posición del objeto en la pantalla; si la coordenada cumple con las condiciones en la división de la pantalla, se activará un led indicador, que representará el proceso con el que se está interactuando.

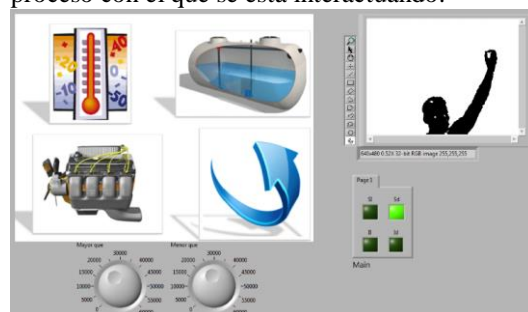


Fig. 18. División de cuadrantes para trabajar sobre el entorno.

En la figura 19, se detalla una imagen en donde se encuentran 4 leds indicadores, cada led representa cada una de las cuatro esquinas de la pantalla (superior izquierda, superior derecha, inferior izquierda, inferior derecha), si el círculo se encuentra dentro de esas regiones, el led correspondiente a esa región se activará; en ese orden de ideas, con que el usuario ubique un círculo en uno de los cuatro cuadrantes, la interfaz inmediatamente realizará una acción sobre un sistema, ya sea desde activar una electroválvula, hasta una alarma.



Fig. 19. Leds indicadores de los cuadrantes del sistema para controlar los procesos.

3.3 Aplicaciones

El sistema está abierto para manipular varios procesos y/o aplicaciones; aun así, el sistema está diseñado principalmente para manipular procesos como temperatura, nivel y velocidad, por lo que se está controlando una planta simuladora de fluidos industriales; teniendo las ventajas de:

- Poder manipular aplicaciones por medio de visión artificial sustituyendo las interfaces convencionales.
- Vincular nuevos sensores y dispositivos de tecnologías convergentes a procesos de nivel industrial.
- Manipular varios procesos a través de una interfaz sencilla, cómoda y dinámica para el usuario.

Entre algunas de las otras aplicaciones que se pueden vincular a la interfaz desarrollada, se pueden mencionar Interfaces de control en Robótica, Domótica, Seguridad, Medicina, Psicología, entre otras.

4. CONCLUSIONES

Uno de los avances tecnológicos más importantes de los últimos tiempos, es el empleo del tratamiento digital de imágenes en diferentes áreas (como la medicina, la instrumentación Industrial, etc.), en pro del desarrollo tecnológico y apuntando a interfaces más sencillas y cómodas al usuario, por lo que la interfaz diseñada, brinda la facilidad de interactuar con uno o varios procesos haciendo uso de la visión artificial, en donde el sistema detecta objetos con formas circulares en un espacio determinado de la pantalla, con el fin de emular una especie de cursor que seleccione cada proceso.

A partir del desarrollo del programa de aplicación de prueba "Point Cloud" se dio inicio al proceso de familiarización con el Sensor Kinect con el fin de lograr evaluar el funcionamiento general del mismo, para obtener la imagen RGB, el mapa de profundidad, el ángulo de control del sensor, entre otros aspectos.

Los subVIS para adquirir tanto la imagen RGB, el control de ángulo del motor como el mapa de profundidad, permitieron utilizar el sensor Kinect con el software LabVIEW, facilitando vincular la interfaz a un sistema de comunicación bidireccional y otros proyectos desarrollados a lo largo de las pruebas.

Las propiedades y características del sensor, facilitaron el procesamiento en el desarrollo del programa, la adquisición de imágenes en tiempos relativamente pequeños y la manipulación de cada una de las aplicaciones.

REFERENCIAS

- Bevly, D. (2012). Using the Microsoft Kinect for 3D Map Building and Teleoperation. *Transform*, 1054–1061.
- Chiang, I., & Chen, S. (2012). Institutionalized Older Adults with Wheelchairs. *Time*. <https://doi.org/10.1109/DIGITEL.2012.69>
- Eduardo, J., Salvatore, J., Osio, M., & Morales. (2014). Detección de objetos utilizando el sensor Kinect. *12 Th Latin American and Caribbean Conference for Engineering and Technology*, 1–10.
- Ivorra Martínez, E. (2015). *Desarrollo de técnicas de visión hiperespectral y tridimensional para el sector agroalimentario* [Universidad Politécnica de Valencia]. [https://riunet.upv.es/bitstream/handle/10251/48541/Ivorra - Desarrollo de técnicas de visión hiperespectral y tridimensional para el sector agroali....pdf?sequence=1](https://riunet.upv.es/bitstream/handle/10251/48541/Ivorra_-_Desarrollo_de_técnicas_de_visión_hiperespectral_y_tridimensional_para_el_sector_agroali....pdf?sequence=1)
- López, D. A., Hernández, J. E., & Peña Cortes, C. A. (2013). Advances in the control of bipedal platforms using the system Kinect. *Revista Colombiana de Tecnologías de Avanzada*, 2, 117–124. <https://doi.org/https://doi.org/10.24054/16927257.v22.n22.2013.419>
- Mathe, L., Samban, D., & Gómez, G. (2012). Estudio del funcionamiento del sensor Kinect y aplicaciones para bioingeniería. *Interface*, 1–6.
- Niño, C., Castro, S., & Medina, B. (2020). Characterization for location in the capture of video applied to artificial vision techniques in the people detection. *Revista Colombiana De Tecnologías De Avanzada (Rcta)*, 38(38), 121–143.
- Pardo-Beainy, C., Gutiérrez, E., Jiménez López, F. R., & Sosa, L. (2014). Drawing robot controlled through the kinect sensor. *Revista Colombiana de Tecnologías de Avanzada*, 2(24), 43–48.
- Smisek, J., Jancosek, M., & Pajdla, T. (2011). 3D with Kinect. *Image (Rochester, N.Y.)*, 1154–1160.
- Štrbac, M., Markovi, M., & Popovi, D. B. (2012). Kinect in Neurorehabilitation: Computer Vision System for Real Time Hand and Object Detection and Distance Estimation. *11th Symposium on Neural Network Applications in Electrical Engineering, NEUREL-2012*, 127–132.

- Štrbac, M., Markoviü, M., & Popoviü, D. B. (2012). Kinect in Neurorehabilitation: Computer Vision. *Image (Rochester, N.Y.)*.
- Um, D., Ryu, D., & Kal, M. (2011). *Multiple Intensity Differentiation for 3-D Surface Reconstruction With Mono-Vision Infrared*. 11(12), 3352–3358.
- Vivas, J. E., Imbachi, S. C., & Florez, J. F. (2017). Assessment of estimation techniques optics depth measures. *Revista Colombiana De Tecnologias De Avanzada (Rcta)*, 1(29), 138–143.