Estrategia para Detección de la Mano y Reconocimiento de Gestos con los Dedos

M.C. Aarón Junior Rocha Rocha[1], M.C. Ana Celia Segundo Sevilla[2] y
M.A. y M.C.T.C. Julieta Raquel Hernández Vidales[3]

***Resumen*—La detección de la mano y el reconocimiento de gestos es un tema ampliamente estudiado en los sistemas de Interacción Humano-Computadora (IHC). El uso de diversos y novedosos dispositivos (como cámaras de tiempo de vuelo, *time of flight)* provee a los investigadores una forma fácil y rápida de obtener información de una escena generando soluciones diversas cada vez más eficientes. En este artículo se presenta una estrategia de procesamiento del mapa de profundidad, obtenido mediante Kinect de Microsoft, para realizar las tareas de detección de la mano y el reconocimiento de gestos con los dedos. El enfoque que aquí se presenta tiene la ventaja de ser intuitivo, tener un bajo costo computacional, ser escalable y que puede reconocer más de una mano. Así mismo, se presenta un conjunto de pruebas realizadas controlando el puntero del *mouse* y sus funciones básicas.**
***Palabras clave*—IHC, detección de la mano, reconocimiento de gestos, *Kinect*, mapa de profundidad.**

Introducción

La Interacción Humano Computadora busca hacer más fácil, intuitivo y cómodo el uso de las computadoras en general. Estudia el diseño y desarrollo de nuevas interfaces tanto de *hardware* como de *software* que permitan y mejoren la interacción del usuario con la computadora. En este contexto, Kinect de Microsoft (2015) se ha convertido en un dispositivo ampliamente utilizado en esta área, pues provee a los investigadores y desarrolladores de una gran cantidad de información espacial de los objetos en una escena real. Este dispositivo ha permitido desarrollar múltiples sistemas de videojuegos educativos y de entretenimiento (Gonzáles, 2012), rehabilitación física (Muñoz *et al*, 2013 y Moreno *et al*, 2013), control robótico y de cómputo (Sucar *et al*, 2014 y Meza y Muñoz, 2014), Realidad Aumentada (Moreno *et al*, 2013 y Leal *et al*, 2013), etc. Realizar la interacción mediante un dispositivo de este tipo puede requerir diversas tareas como, encontrar e identificar a un usuario, reconocer las partes de su cuerpo, reconocimiento de los gestos (las indicaciones) que realiza, entre otros. Es de especial interés para esta investigación la tarea de detectar la mano y reconocer gestos que se realizan con los dedos, en particular el gesto de tocar.

Estado del arte

En el estado del arte se puede encontrar diversos trabajos enfocados tanto en la detección de la mano como el reconocimiento de gestos. En el primer caso algunos autores (Frati y Prattichizzo, 2011 y Park, 2008) realizan la detección mediante técnicas como la búsqueda del casco convexo (Eilberg, 2004) a una porción segmentada de la imagen. De este modo pueden encontrar la punta de los dedos y el centro de la mano. Sin embargo, la información de profundidad que se puede obtener de un solo punto perteneciente a un dedo no es consistente a lo largo de una secuencia de imágenes, aún estando inmóvil.

Otros autores (Rudomin *et al*, 2014 y Ryan, 2012) realizan una búsqueda del contorno de la mano basado en sus características morfológicas y, posteriormente, realizan un análisis mediante el algoritmo de *curvatura-k*, para encontrar los dedos de la mano. Sin embargo, el Kinect, basado en sus principios de funcionamiento, resulta impreciso pues los haces de luz que utiliza tienden a perderse o verse afectados por distintas condiciones físicas del ambiente en que opera, generando lecturas erróneas. Comúnmente, dichas lecturas erróneas generan huecos o protuberancias en los objetos, pixeles muertos o falsos positivos que distorsionan y entorpecen la aplicación de algoritmos como *curvatura-k.*

El trabajo presentado en este documento tiene como objetivo abordar varias de las problemáticas ya mencionadas, como la imprecisión del sensor, la detección de los dedos cuando existen lecturas erróneas, así como generar una técnica de reconocimiento de gestos diferente a lo reportado hasta el momento en el estado del arte.

[1] El M.C. Aarón Junior Rocha Rocha es profesor asignatura "A" del Instituto Tecnológico de Estudios Superiores de Zamora, Zamora, Michoacán. mcaaron87@gmail.com (**autor corresponsal**)
[2] La M.C. Ana Celia Segundo Sevilla es profesora titular "A" del Instituto Tecnológico de Estudios Superiores de Zamora, Zamora, Michoacán. chell081@hotmail.com
[3] La M.A. y M.C.T.C Julieta Raquel Hernández Vidales es profesora titular "A" del Instituto Tecnológico de Estudios Superiores de Zamora, Zamora, Michoacán. julietahv1970@yahoo.com.mx

ISBN 978-1-939982-17-9

Técnica propuesta

La estrategia propuesta en este documento está constituida de dos fases: la primera fase consiste en detectar la mano y su ubicación en la escena encontrando los dedos de esa mano. Este proceso permite detectar más de una mano, sin embargo sólo permite encontrar 4 de los 5 dedos de la mano (se omite el pulgar). La segunda fase consiste en determinar el gesto que se está realizando.

Fase 1. Detección de la mano

Para la primera fase se parte del mapa de profundidad proporcionado por el Kinect y se realizan 5 pasos fundamentales.

Paso 1. Se recorre la matriz de profundidad para encontrar el punto más cercano al sensor, *pz*, pues se considera que el usuario se colocará de frente al sensor y levantará la mano hasta mostrar la palma al sensor Kinect. De esta forma se asume que la mano del usuario será el "objeto" más cercano al sensor, como se ilustra en la Figura 1*(a)* y *(b)*. Note que en este enfoque no se consideran otros objetos que pudieran estar en la escena.

(a) (b)

Figura 1. (a)Ejemplifica la posición que se asume que tomará el usuario en relación al sensor Kinect. (b)Se muestra el punto más cercano al sensor que, en este caso, corresponde al dedo medio de la mano del usuario.

Paso 2. Se segmenta el mapa de profundidad y se extraen únicamente los puntos que se encuentran en un rango de [*z*-100, *z+100]*, donde *z* es el valor sobre el eje Z del punto *pz. E*sto servirá para permitir que haya un rango de movilidad para los dedos, tanto hacia adelante como hacia atrás. Además, esta segmentación reduce el costo computacional de la estrategia, pues delimita la cantidad de puntos que se procesarán en cada paso. Este proceso resulta en una imagen binaria cuyos valores son 1 si esta dentro del rango y 0 en caso contrario, como se ilustra en la Figura 2. Para facilitar la explicación de los pasos posteriores se llamará a los 1s puntos válidos y a los 0s puntos inválidos.

Como ya se mencionó, este tipo de sensores presentan diferentes tipos de problemas, entre ellos, la imprecisión en la lectura de los datos. Los datos presentan imperfecciones como huecos y protuberancias, como se observa en la Figura 2.

Figura 2. Resultado de aplicar el proceso de segmentación a la imagen para extraer las manos. Los puntos válidos se representan en blanco y los puntos inválidos en negro.

En particular en este enfoque, las imperfecciones que más le afectan son los puntos aislados y los puntos muertos, que son los puntos válidos rodeados de puntos no válidos y viceversa, respectivamente. El motivo se explicará más adelante. Para disminuir la cantidad de estos puntos muertos se utilizará un proceso de dilatación y erosión de la

imagen como lo explica Pajares (2008). La dilatación elimina una porción importante de los puntos muertos y la erosión permite eliminar algunas pequeñas protuberancias como líneas delgadas y puntos aislados. Los elementos estructurales utilizados para los procesos de dilatación y erosión de la imagen se muestran en la Figura 3(a) y (b).

1	1	1
1	• 1	1
1	1	1

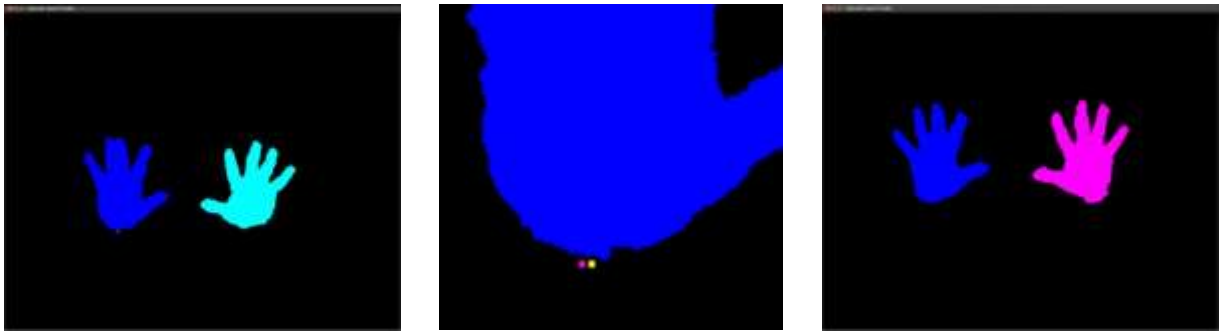
(a)

• 1	1	1

(b)

Figura 3. (a) Elemento estructural para el proceso de Dilatación. (b) Elemento estructural para el proceso de Erosión

Paso 3. Se analizan los puntos restantes para formar grupos mediante vecindades, es decir, se generan grupos con los puntos contiguos que aparecen en el mapa. De esta forma es posible encontrar múltiples objetos en la escena, en el rango de proximidad al punto más cercano al sensor. Estos grupos u objetos pueden ser una o varias manos, como se observa en la Figura 4*(c)*.



(a) (b) (c)

Figura 4. (a)Identificación de falsos grupos u objetos en la imagen. (b) Acercamiento a falsos grupos de pixeles. (c)Resultado de identificar los grupos de puntos por vecindades de pixeles continuos.

Como ya se mencionó la matriz de profundidad generada por el sensor contiene una cantidad significante de ruido de varios tipos. El proceso de dilatación y erosión de la imagen corrige pequeños errores como puntos aislados o puntos muertos. Sin embargo, tras el proceso de agrupación se encuentran grupos que son muy pequeños para ser objetos significativos. Por ejemplo, en la Figura 4*(a)* se muestra una captura tomada con el Kinect y en la Figura 4*(b)* se observa un acercamiento a dos grupos de puntos aislados al resto de la mano. Estos errores se presentan a lo largo de una secuencia de video en distintas posiciones y de diversos tamaños. Una solución sencilla para remover estos errores es convertir los puntos válidos en no válidos de los grupos cuyo tamaño sea menor a un cierto umbral, λ. El tamaño puede ser obtenido durante el proceso de agrupamiento.

Paso 4. Se segmenta cada grupo de puntos restantes y se escanean de forma individual con el objetivo de encontrar y diferenciar cada uno de los dedos de la mano. Dicho escaneo se realiza recorriendo la matriz de profundidad de arriba hacia abajo y de izquierda a derecha. Se distinguen dos casos durante el escaneo: al encontrar un nuevo dedo y al continuar escaneando un dedo ya encontrado. En la Figura 5 se ilustra el proceso de detección de un primer dedo de la mano.

En el primer caso, cuando se encuentra el primer punto válido sobre un renglón, y_i, este se etiqueta como *dn*, donde *n* es el número de dedo encontrado hasta el momento (1 para este ejemplo), y que es el origen para propagar dicha etiqueta hacia otros puntos válidos vecinos. A partir de este punto origen, se etiquetan como *d1* los puntos válidos a su derecha, hasta encontrar un punto no válido. Después, a partir de la posición del punto origen pero sobre y_{i+1} , se hace el mismo etiquetado, ahora tanto a la izquierda como a la derecha. Esto se puede interpretar como una propagación de la etiqueta de los puntos hacia sus vecinos. El proceso de etiquetado se realiza cada vez que se encuentra un nuevo dedo (punto válido sin etiquetar) como se observa en la Figura 5*(a), (b)* y *(c)*.

En el segundo caso, durante el escaneo se encuentra un punto válido ya etiquetado, perteneciente a alguno de los dedos. Este punto será el nuevo origen y el procedimiento a seguir es muy similar al del primer caso. Se etiquetan los puntos válidos no etiquetados que se encuentren debajo del origen, tanto a la izquierda como a la derecha como se ilustra en la Figura 5*(d), (e)* y *(f)*. Los puntos a la derecha del origen, ya están etiquetados.

Al final de procesar la imagen con ambos casos de escaneo se obtendrán 4 de los 5 dedos de la mano (del índice al meñique). Para obtener el pulgar se requiere un paso adicional que no se aborda en este documento.

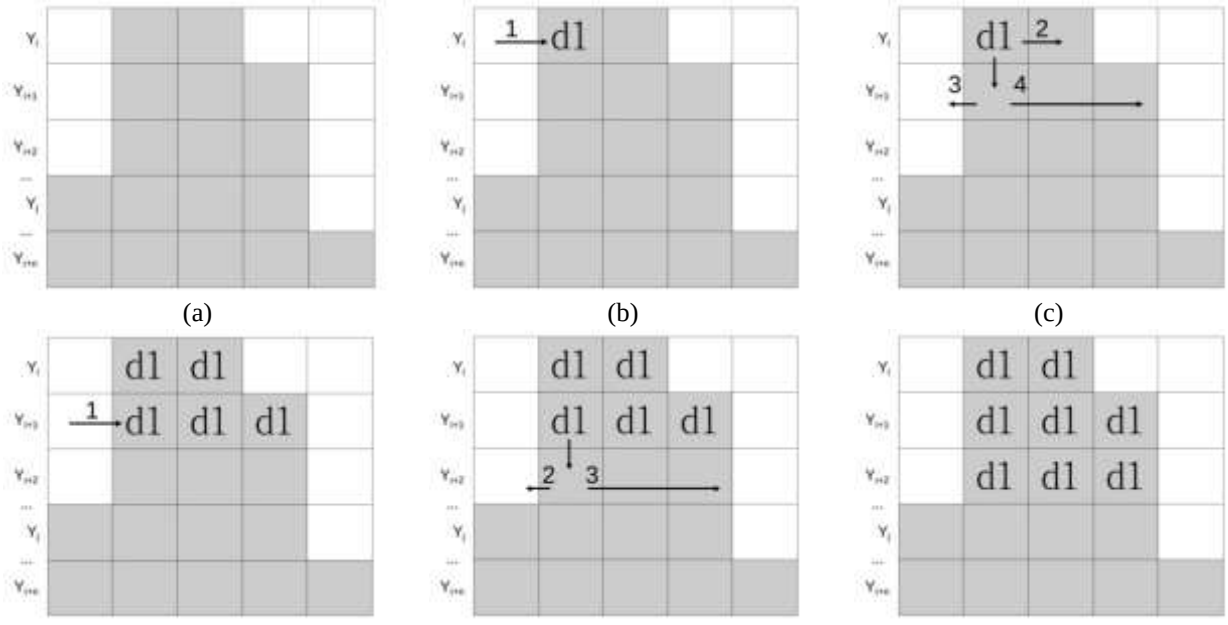


Figura 5. (a) Imagen a original a escanear. (b) Se inicia el escaneo en y_i, al encontrar un punto válido se etiqueta como d1. (c) A partir del punto recién encontrado se etiquetan todos los puntos válidos a su derecha hasta el siguiente punto no válido, así como los que se encuentran a la izquierda y a la derecha debajo de dicho punto origen. (d) Imagen a escanear a partir de y_{i+1} con puntos válidos ya etiquetados. (e) Se etiquetan los puntos que están debajo del origen tanto a la izquierda como a la derecha. (f) Resultado del proceso de etiquetado del segundo caso.

Paso 5. Finalmente se procede a realizar una identificación de los dedos basados en su ubicación relativa. Dado que la posición de la mano debe ser vertical, los dedos se ubican uno al lado de otro, como ya se había explicado. Por lo tanto, se podrán identificar según el orden en el que aparecen de izquierda a derecha o viceversa. Si se reconocen dos manos se considera primero el orden en el que cada mano aparece para determinar si es la izquierda o derecha. El orden de aparición de los dedos debe invertirse de acuerdo a la mano a la que pertenecen.

Fase 2. Reconocimiento de gestos

Este trabajo se centra en un tipo de gestos muy específico, que consiste en simular toques con los dedos como se realizarían sobre una superficie táctil (un cristal o una pantalla). El gesto de tocar se divide en tres subtipos: tocar, mantener y soltar. Se comienza con la premisa de que el usuario iniciará la interacción mostrando la palma de la mano al sensor. A partir de ahí se realiza el proceso de detección de la mano que ya se explicó. Posteriormente, se debe analizar la información de los dedos para determinar si se ha realizado un gesto.

Un gesto consiste en acercar uno o más dedos hacia el sensor, para simular un toque. En principio, se podría afirmar que lo que se necesita es conocer la posición de la punta de los dedos sobre el eje *Z*, sin embargo, debido a la inexactitud del Kinect no es posible obtener una lectura consistente. En otras palabras, el valor en *Z* de un mismo punto *(x,y)* del mapa de profundidad varía de un *frame* a otro, aún cuando los objetos permanezcan inmóviles. Por este motivo, se calcula el valor en el eje *Z* de cada dedo, dn_z, mediante una media aritmética del valor *z de* todos los puntos que lo conforman.

Con el objetivo de determinar si un dedo ha realizado el gesto de "tocar" se utiliza la función *t*, definida por (1).

$$t(dn) = \begin{cases} true & ssi \quad (dn_z - \bar{x}) < (\sigma * -1) \\ false & en\ caso\ contrario \end{cases} \qquad (1)$$

donde $\bar{x}$ es la media aritmética de los valores de todos los dedos de la mano; y σ es la desviación estándar de los

mismos valores. Esto representa que el valor un dedo debe sobresalir del resto de los valores para ser considerado un toque. De esta forma se tiene la posibilidad de realizar toques con uno o dos dedos, directamente con esta estrategia, o con más dedos si se utiliza información adicional como el valor de la palma de la mano o la posiciones previas de los dedos, lo cual no se abordará en este trabajo.

Caso de prueba

Para probar esta estrategia se definió un caso simple de Interacción Hombre-Computadora (HCI). El objetivo de la prueba consiste en controlar el movimiento y las funciones de clic izquierdo y derecho del mouse en un entorno de escritorio real.

El control del movimiento del puntero se habilita cuando se detecta la mano y los dedos. Al momento, se despliega un área, a la que se identifica como *Pad*, el cual representa el área de desplazamiento del mouse y que tiene dos objetivos: 1) permitir al usuario ubicar su posición en la pantalla respecto al mundo real; 2) delimitar la longitud de los movimientos necesarios para llegar de un extremo de la pantalla a otro.

Para controlar el puntero, se utiliza como referencia el centro del área de la mano. Luego, se mapea la posición de dicho punto en el área del *Pad* respecto a la pantalla de la computadora.

Mientras el *Pad* permanece habilitado, el sistema se mantiene a la espera de reconocer el gesto de tocar realizado con cualquiera de los dedos, índice o medio, de la mano derecha, controlando la función de clic izquierdo y derecho, respectivamente. Cuando se reconoce el gesto, el sistema lleva acabo el evento «Presionar» del mouse, y al dejar de reconocer el gesto realiza el evento «Soltar», esto ocurre para ambos dedos.

Comentarios finales

Resumen de resultados

El experimento realizado se dividió en dos casos, el primero realizando los gestos con la mano inmóvil y el segundo con la mano en movimiento. En ambos casos se contabilizó si los tres subtipos del gesto se realizan correctamente. La tasa de reconocimiento de los gestos con la mano inmóvil se presenta en la tabla 1(a). De los resultados se observa que los gestos son reconocidos correctamente en la mayoría de los ensayos con el dedo índice; sin embargo, con el dedo medio no se tienen resultados favorables. Por este motivo se realiza la misma prueba ahora con el dedo anular, como alternativa al dedo medio, con lo cual la tasa de reconocimiento mejora notoriamente. Se atribuye el deficiente reconocimiento del dedo medio con esta técnica, a sus características morfológicas, pues este dedo tiene una mayor longitud que los otros dos mencionados. Al colocar la mano de frente al sensor, en una posición inclinada, el dedo medio tiene valores de mayor cercanía al sensor, respecto a los otros dedos, aun sin realizar el gesto. Por este motivo se obtienen numerosos falsos-positivos. Los resultados del segundo caso se presentan en la tabla 1(b) y muestran que la tasa de reconocimiento sigue siendo alta con los dedos índice y anular. Al igual que en el primer caso, el reconocimiento con el dedo medio es deficiente.

	Mano inmóvil		
	Tocar	**Mantener**	**Soltar**
Índice	95%	95%	95%
Medio	35%	45%	25%
Anular	90%	95%	90%

a)

Mano en movimiento		
Tocar	**Mantener**	**Soltar**
90%	90%	90%
15%	25%	20%
90%	85%	85%

b)

Tabla 1. a) Tasa de reconocimiento de gestos con mano inmóvil. b) Tasa de reconocimiento de gestos con mano en movimiento.

Conclusiones

En este documento se presentó una estrategia para detección de la mano y reconocimiento de gestos que resulta intuitiva, de bajo costo computacional y escalable. Se presentaron los resultados de las pruebas realizadas las cuales señalan la eficacia de la estrategia. Esta investigación permitirá mejorar la interacción mediante sistemas de tipo kinestésico como lo es el Kinect de Microsoft. Además, este trabajo da lugar a múltiples líneas de trabajo futuro que permitirán extender y mejorar lo que aquí se presentó. Por ejemplo, mejorar el proceso de detección de la mano, para permitir encontrar todos los dedos en diferentes orientaciones de la mano; descartar la suposición de que la mano controladora será el punto más cercano al sensor; reconocer gestos de todos los dedos e incluso con varios de forma simultánea. También se prevé la posibilidad de realizar un proceso de aprendizaje para reconocer secuencias de

ISBN 978-1-939982-17-9

gestos comunes o frecuentes y que permita disminuir los errores de detección al aprender distintos perfiles de usuario.

Referencias

Eilberg, E. "Convex hull algorithms," *Student Scholarship*, 2004.

Frati, V., y Prattichizzo, D. "Using Kinect for hand traking and rendering in wearable haptics," *World Haptics Conference (WHC) 2011 IEEE*, vol., no., pp.317,321, 21-24 June 2011

González, V. "Advant y Advant-ed: plataforma para el entrenamiento cognitivo y físico con Kinect," 2012.

Leal, J. A., Altamirano. L., y González, Jesús. "Occlusion Handling in Video-Based Augmented Reality Using the Kinect Sensor for Indoor Registration." *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*. Springer Berlin Heidelberg, 2013. 447-454.

Meza, L. E., y Muñoz, P. A. "Programación de órdenes a un robot Lego mediante interfaz natural," 2014.

Microsoft. "Kinect for Windows", recuperado el 12/08/2015, http://www.microsoft.com/en-us/kinectforwindows/

Moreno, F., Ojeda, J., Ramírez, E., Mena, C., Rodríguez, O., Rangel, J., y Álvarez, S. "Un framework para la rehabilitación física en Miembros superiores con Realidad Virtual." *Primera Conferencia Nacional de Computacion, Informatica y Sistemas. Universidad Central de Venezuela*, 2013.

Muñoz, J E., Henao, O. A., y López, J.F. "Sistema de Rehabilitación basado en el Uso de Análisis Biomecánico y Videojuegos mediante el Sensor Kinect," 2013.

Pajares, G., y de la Cruz, J. *Visión por computador: Imágenes digitales y aplicaciones*, México: Alfaomega, Ra-Ma, 2008.

Park, H. "A method for controlling mouse movement using a real-time camera," *Brown University, Providence, RI, USA, Department of computer science*, 2008.

Rudomin, I., Ramírez, J., y Arzate, C. "Método robusto para detectar dedos usando profundidad," *Research in Computer Science*, Vol. 64, 2014.

Ryan, D. "Finger and gesture recognition with microsoft kinect." 2012.

Sucar, L. E., Morales. E., Palacios-Alonso, M., Heyer, P., Vázquez, I., Carillo, D., Ruiz, E., Reyes, E., Mosso, A., Enríquez-Caldera, R., y Tobon, J. "Markovito's Team Description RoboCup@ Home 2014," 2014.