Reconocimiento de gestos de mano utilizando Kinect

Yi Li

El equipo de ingeniería y ciencias de la computación

Universidad de Louisville

Louisville,  KY 40214, EE.UU.

Yi,li@louisville,edu.

   Reconocimiento de gestos Abstract-Hand  (HGR) es un importante

Tema de investigación porque algunas situaciones requieren communi- silenciosa con cationes de lenguajes de signos. HGR sistemas computacionales asistir en silencio la comunicación, y ayudar a la gente a aprender un lenguaje de signos. En este artículo, un novedoso método para contactar con menos HGR utilizando Microsoft Kinect para Xbox se describe, en tiempo real y un sistema implementado HGR. El sistema es capaz de detectar la presencia de gestos, para identificar los dedos, y a reconocer los significados de nueve movimientos en una pre-definido gesto popular escenario. El

Precisión  del sistema HGR es de 84% a 99% con single-

  Los gestos de la mano,  y desde el  90% hasta el 100% si ambas manos realizar el mismo movimiento al mismo tiempo. Porque el sensor de profundidad del Kinect es una cámara de infrarrojos, las condiciones de iluminación, deudores" colores de piel y ropa, y el fondo tienen poco impacto en el rendimiento de este sistema. La precisión y robustez hacen de este sistema un componente muy versátil que puede ser integrado en una amplia variedad de aplicaciones en la vida diaria.

Keywords-Human-interacción con el ordenador de mano, gesto recogni-

Ción, Kinect, identificación de dedos

I. Introducción

Siempre hay una necesidad de comunicar mediante el signo lan- medidores, como charlar con personas con discapacidad del habla y la audición. Además, existen situaciones en las que se prefiere la comunicación com- silenciosa; por ejemplo, durante una operación, un cirujano puede gesto a la enfermera para obtener asistencia. Es difícil para la mayoría de las personas que no están familiarizados con el lenguaje de señas para comunicarse sin un intérprete. Así, el software que transforma los símbolos en los lenguajes de signos en texto sin formato puede ayudar con la comunicación en tiempo real, y puede  también proporcionar formación interactiva a la gente a aprender un lenguaje de signos. Reconocimiento de gestos se ha convertido en un importante campo de investigación con el enfoque actual sobre el reconocimiento de la emoción interactiva y reconocimiento de gestos de mano.

Microsoft  Kinect  proporciona una forma sencilla y barata para la interacción del usuario en tiempo real. El controlador de software publicado por Microsoft llamado Kinect Software Development Kit (SDK) con Interfaces de programación de aplicaciones (API) dan acceso a flujos de datos raw del sensor así como seguimiento del esqueleto. Sin embargo, no hay ninguna mano datos específicos disponibles para el reconocimiento de gestos, aunque sí incluyen información de las articulaciones entre las manos y los brazos. Se ha hecho poco trabajo para Kinect para detectar los detalles a nivel de los dedos individuales.

En este artículo, un sistema HGR utilizando Microsoft Kinect para

Xbox ha sido implementado. El sistema detecta los gestos de la mano hecha por el usuario, los compara con los signos de la pre- definido gesto popular y muestra la coincidencia de significado y la imagen correspondiente en la pantalla. El sistema

978-1-4673-2008-5/12/$31.00 ©2012 IEEE

Representa un nuevo enfoque de la interacción humano-ordenador que utiliza nada pero las manos desnudas como los medios de entrada.

Se generan los datos de profundidad y se convierte en la imagen raw

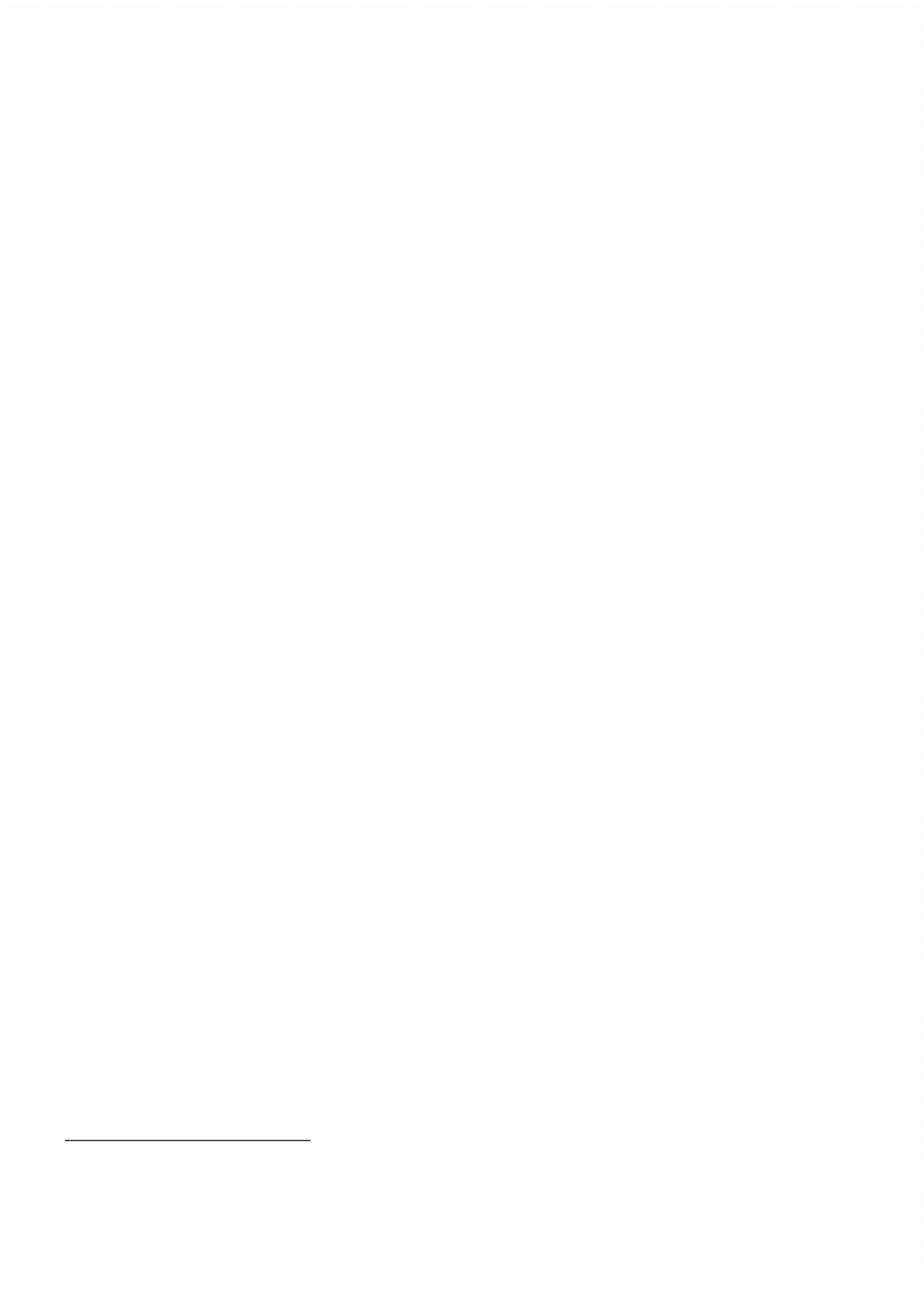
Datos   de Kinect Sensor por un marco de código abierto llamado OpenNI (Interacción Natural), con un controlador de código abierto llamado SensorKinect por PrimeSense, que hace Kinect para Xbox compatible con Microsoft Windows 7. Se utiliza la última versión del controlador de gráficos, el sistema está desarrollado en C# de Microsoft Visual Studio 2010. La precisión del sistema HGR es de 84% a 99% con un solo gesto de la mano, y desde el 90% hasta el 100% si ambas manos realizar el mismo movimiento al mismo tiempo. La identificación de todos los dedos individuales nunca se ha hecho en la literatura con la plataforma Kinect al mejor de mi conocimiento. Porque el sensor de profundidad del Kinect es una cámara de infrarrojos, las condiciones de iluminación, deudores" colores de piel y ropa, y el fondo tienen poco impacto en el rendimiento de este sistema. La precisión y robustez hacen de este sistema un componente muy versátil que puede ser integrado en una amplia variedad de aplicaciones en la vida diaria.

II. Litro maduros REV ER

Por lo general, un lenguaje de signos consta de tres partes: la ortografía, Word nivel finger- firmar el vocabulario, y no funciones manuales. Por lo tanto, HGR constituye la base en la traducción de lenguajes de signos. Tecnología reciente para el reconocimiento de gestos es incorporar la información objeto de distancias, o profundidades, normalmente capturada por una cámara 3D, o un conjunto de cámaras que producen imágenes 3D. En [1], Van den Bergh et al.  Utiliza un tiempo de vuelo para mejorar el reconocimiento de la cámara. Su sistema podría reconocer seis gestos: mano abierta, puño, apuntando hacia arriba, en forma de L, apuntando a la cámara, y el pulgar arriba. El reconocimiento basado en RGB mostraron una precisión de 99,5 4%, basada en la profundidad y el reconocimiento mostraron una precisión del 99.07%, mientras que la combinación de los dos métodos mostraron 99,5 4%. Esto sugiere que el reconocimiento basado en profundidad puede ser lo suficientemente buena como para formar la base de un sistema HGR especialmente en mano la segmentación, como la combinación de dos métodos no proporcionan una mejora significativa de la exactitud.

Numerosas investigaciones se han realizado con Kinect desde su lanzamiento en 2010. Sin embargo, sólo unos pocos sistemas fueron desarrollados para el HGR, y sólo unos pocos gestos fueron reconocidos. Yang et al. [2] propone un sistema de reconocimiento de gestos con información detallada proporcionada por Kinect, e implementado en una aplicación de reproductor de medios. La trayectoria de mano fue examinado por un

La función 3D vector y el gesto fue reconocido por una oculta



196

(A)  (B)  (C)  (D)

(E)  (F)  (G)  (H)

Fig. I. Seguimiento del contorno. Trazar el contorno de los píxeles azules: Vamos píxeles verdes representan el contorno detectado píxeles. El píxel con límite rojo denota el contorno actual píxel; el píxel con frontera verde indica el inicio de un barrio de píxel comprobando, a saber, el último píxel contorno excepto para el punto de partida. La flecha negra indica la ruta en el sentido de las agujas del reloj de la vecindad comprobación; la flecha discontinua muestra la ruta que el algoritmo no necesita para terminar, porque el próximo contorno pixel ya se ha detectado.

Modelo de Markov (HMM). Este sistema demostró la applica- bilidad de utilizar Kinect para reconocimiento de gestos sin contacto en una interfaz de usuario (UI). El sistema no pudo encontrar las puntas de los dedos; por lo que se limitaba a reconocer únicamente movimiento gestos como onda, y muévalo hacia arriba/abajo, izquierda/derecha, y hacia adelante/atrás. Un método para hacer un seguimiento de dedos y los centros de las palmas utilizando Kinect fue presentado por Raheja et al.  [3].   Cuando los dedos se extendieron, la exactitud de la detección de dedos fue casi

100%, y que los centros de palma fue de alrededor del 90%. Sin embargo, este método no intento de reconocimiento de gestos. Él  et al.  [4] propone un enfoque basado en el uso de los datos de profundidad proporcionada por Kinect para detectar las puntas de los dedos. Después de mano fueron encontrados, el ratón haciendo clic en moción fue reconocido y probado en el popular juego Angry Bird; es decir, reconocieron sólo un gesto.

III. Jefes de Delegación se reunieron

El sistema del presente trabajo consta de tres principales com- ponentes: detección de manos, dedos, identificación y reconocimiento de gestos. Este sistema está construido sobre la incandescente NUl [5], proyecto que está disponible de forma gratuita en línea. OpenNI marco se utiliza para extraer los datos de profundidad del Kinect Sensor.

A. Detección de manos

El  primer  paso  es separar las manos del fondo. Profundidad de umbrales se configuran manualmente con antelación para especificar el intervalo de profundidad donde los gestos deben aparecer para ser reconocidos. El rango de detección,  zRange,    entre 0,5m y 0,8m; hasta dos manos en este rango puede ser detectado. Los píxeles con profundidades fuera zRange son ignorados en el resto del proceso de detección de movimiento. De ahora en adelante, un píxel de mano también pueden ser contempladas como un punto. Estos píxeles de mano se proyecta a un espacio 2D para

Análisis posteriores. Distancia entre dos píxeles(XL P1,Y1)

Y  P2(X2,Y2) se define como:

D(p1,p2)  = -Xd2 + (Y2-yd2. (1) El   algoritmo k-means clustering se aplica a todas las particiones

Los píxeles a dos grupos. K-means clustering es un método

A la  partición   n  observaciones  en  k grupos,  G1, G2,... , Gk; cada observación se identifica con la media más cercana, fLi(x,y), que se calcula como la media de puntos de Gi. K-means clustering dentro de clúster minimiza la suma de cuadrados:

K

ArgminL L Ilpj(x,y)-/Li(X,Y) 11 2 . (2)

C  I=l Pj (x,y)CE;

En el sistema actual, el valor de  k es  2. El sistema funciona en tiempo real, y K-means clustering continuamente se aplican siempre que se produce un cambio en el origen de datos de entrada. Al comienzo de cada iteración, cada clúster vacía se inicializa con un punto aleatorio situado dentro del  zRange como    media. Después de K-means converge, los puntos pertenecientes a cada mano están agrupados. Si la distancia entre los dos centroides de manos es menor que un valor predefinido, los dos grupos se fusionan en uno.

A continuación, el casco convexo de manos son calculadas. El algoritmo de análisis de Graham [6] se utiliza para calcular el polígono convexo de la mano detecta clusters.  Contornos de mano son detectados por un algoritmo de rastreo Moore-Neighbor modificada. Después los medios K- clústeres, un grupo de puntos de la mano son encontrados y almacenados. La figura 1 muestra un ejemplo de aplicación del algoritmo para detectar el contorno.

B.   Identificación de dedos

Sea  N(a)  serán los ocho puntos de vecinos de un píxel  . Vamos

P  denotan el contorno actual pixel. Deje  q  denotan la partida de pixel actual comprobación de barrio. Vamos  C designan el conjunto

Detectado puntos del contorno, que es inicializado para ser el conjunto vacío.

Algoritmo de seguimiento del contorno:

1) de arriba a abajo y de izquierda a derecha, escanear todos los píxeles en la pantalla hasta que un píxel   se encuentra un punto de mano, que se define como el punto de partida.

2) establecer el contorno actual pixel P a  8.  Establecer  el punto de partida

Píxel  de  barrio comprobando q    sea el punto a la inmediatamente al norte de  s.

3)  Inserte p a  c, y  calcular el barrio N(p).

(A)  (B)

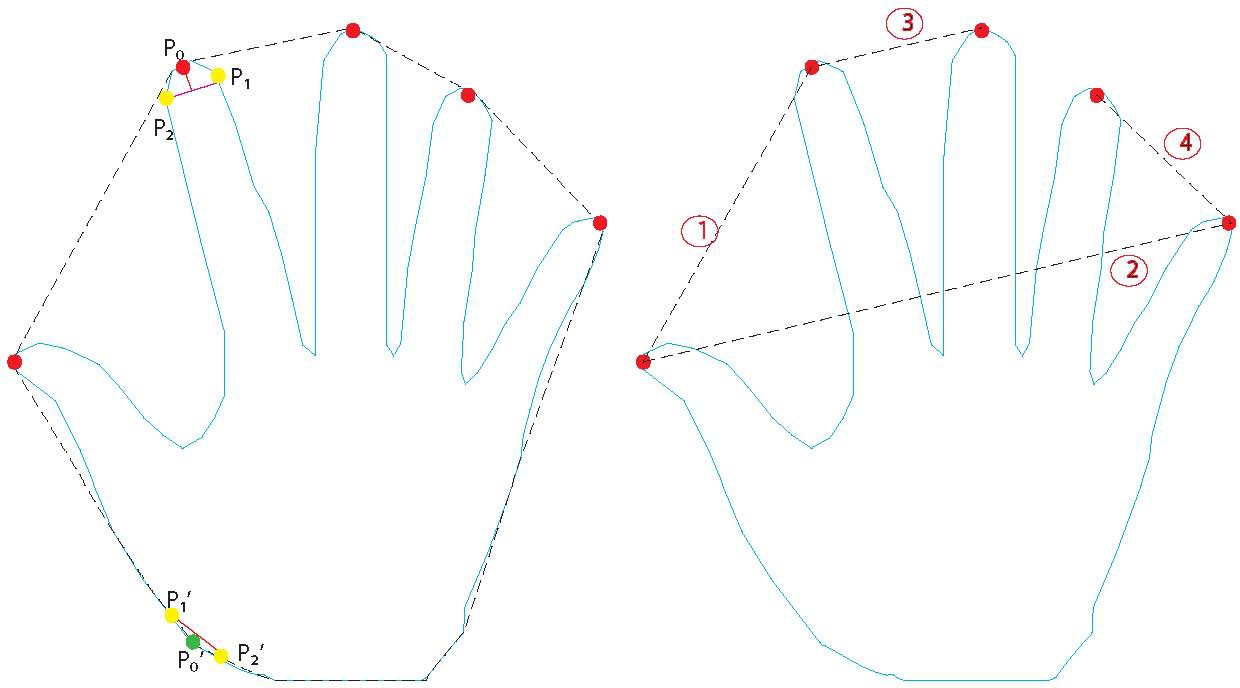
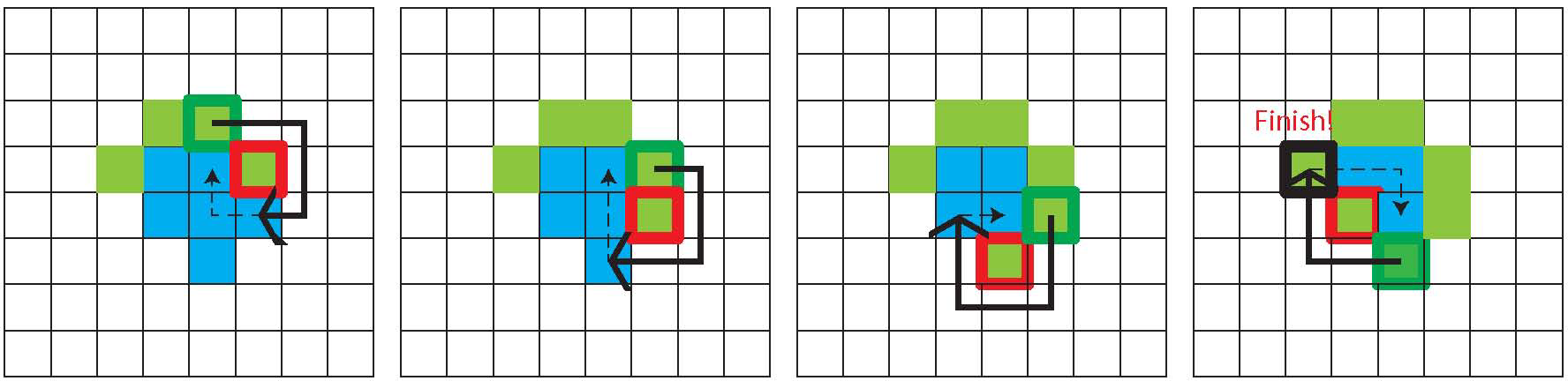
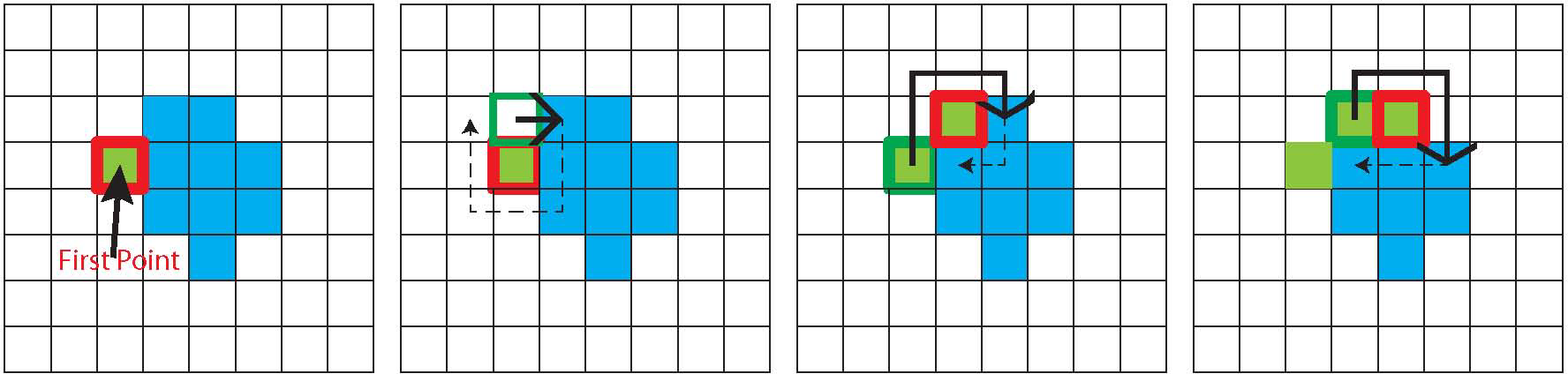
Fig. 2. 2(a) Detección de Fingertip. Los puntos rojos son reales las puntas de los dedos. y el punto verde no es la punta de un dedo. Los puntos amarillos se usan para comprobar la alineación de puntos three-. La distancia entre el Po  y la línea hecha por  PI   y P2.

Aparentemente es mayor que el de  p y la línea hecha por  p y p;. Por lo tanto

Un  umbral distancia podría ser utilizado para diferenciar los puntos y non- fingertip

Puntos fingertip.

2(b) Identificación de los dedos.



197

(A)  (B) Fig. 3. Los resultados de la identificación de los dedos.

4) Comenzar desde  q,  recorrer  el barrio N(p) en el sentido de las agujas del reloj hasta que una mano pixel  r es  encontrado.

5) Set  q  a    p y   a  ser nuevo contorno pixel  r. A continuación, repita el paso 3 hasta que el punto de partida  es  alcanzado de nuevo, o el número de puntos detectados excede un valor máximo.

Después de la mano de contornos son detectados, los centros de las palmas se calculan como los centros de la inscripción de los círculos de la mano de los contornos. Las posiciones de palm centros se utilizan para calcular las direcciones de los dedos, que se describen en la siguiente sección. Los centros de los círculos de inscribir son mucho más estables que los centroides de clusters de mano, porque éstos varían mucho al abrir/cerrar las manos o doblar los dedos.

Las puntas de los dedos son detectados por controlar los tres puntos align- ment relación. Candidato puntos son los puntos que se encuentran tanto en el polígono convexo y el contorno de la mano. El método que se muestra en la Figura 2(a).

Algoritmo de alineación de tres puntos:

1) sea  C el conjunto de todos los candidatos que se encuentran en ambos puntos el polígono convexo y el contorno de la mano.

2) Para cada punto Po en  C, tome los dos puntos PI y P2

Fig. 4. Ejemplo de reconocimiento. En primer lugar, el sistema determina que el número de dedos en este gesto es de dos. A continuación, los nombres de los "dedos" y "el dedo índice, dedo medio" se encuentran. El clasificador de la tercera capa calcula el ángulo de dos vectores de dedos están dentro de un rango de 30 grados a 60 grados. Así, el gesto es reconocida como la "victoria" en el gesto popular escenario, mientras que en el caso de números es reconocido como el "número dos" a través de un proceso similar.

El dedo índice, ya que la distancia entre ellos es la mayor de entre todos los vecinos de dedos.

2) El dedo meñique es identificado como el más lejano lejos del dedo pulgar, mientras el dedo medio es identi- establecidos como la más cercana del dedo índice.

3)  El resto es el dedo anular.

El  proceso de detección e identificación de los dedos de las manos se realizan cada vez que cambia el origen de datos de entrada. Si el mismo objeto todavía existe en el siguiente fotograma con una transformación en comparación con el fotograma anterior, todas las propiedades de este objeto está correlacionado con el bastidor antiguo. Los resultados de la identificación de los dedos se muestran en la figura 3.

C.   Reconocimiento de gestos

Una vez que   los   dedos se identifican con sus nombres, estamos listos para reconocimiento de gestos. Los gestos son pasa por tres capas de clasificadores: Dedo Dedo collect- contando, nombre

Ing y vector coincidentes. El ángulo  entre dos vectores

V;.(XIl YI) y  V); y2(X21) se calcula como:

VI·V;

Las dos direcciones opuestas a lo largo del contorno del Po.

3)  Encontrar el punto medio de PI y P2 y calcular la distancia

A = arccos

IIv;.IIV;II·

( 3)

Entre este punto y  Po. Si la distancia es mayor que un valor específico, los tres puntos no están alineados, es decir, lejos de ser colineales. Entonces este  Po es identificado como la punta de un dedo. De lo contrario, volver al paso 2 y comprobar el siguiente punto de candidatos.

Después de dedos son detectados, sus vectores de dirección son fáciles de calcular restando la posición de la palma

Center  Pc(xc, yc)  (que  se   calcula  en la   sección  III-A):

Si = (x - Y - xc, Yc). Los  vectores están apuntando desde la palma

Centro para los dedos.

 Los nombres de los dedos se determinan de acuerdo a sus distancias relativas, que requieren que el usuario abra Palm y extender los dedos. Cuatro procedimientos para identificar los nombres de todos los dedos se muestran en la Figura 2(b). La identificación de todos los dedos individuales nunca se ha hecho en la literatura con la plataforma Kinect al mejor de mi conocimiento.

Algoritmo de identificación de dedo:

1) La primera y más sencilla es identificar el pulgar y

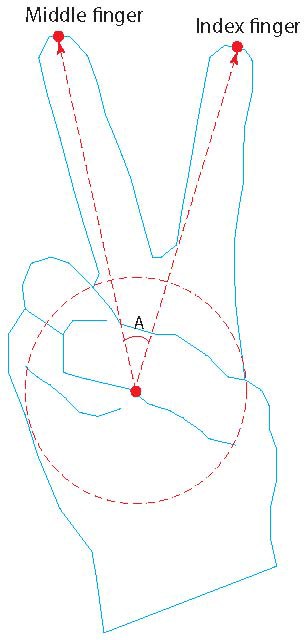
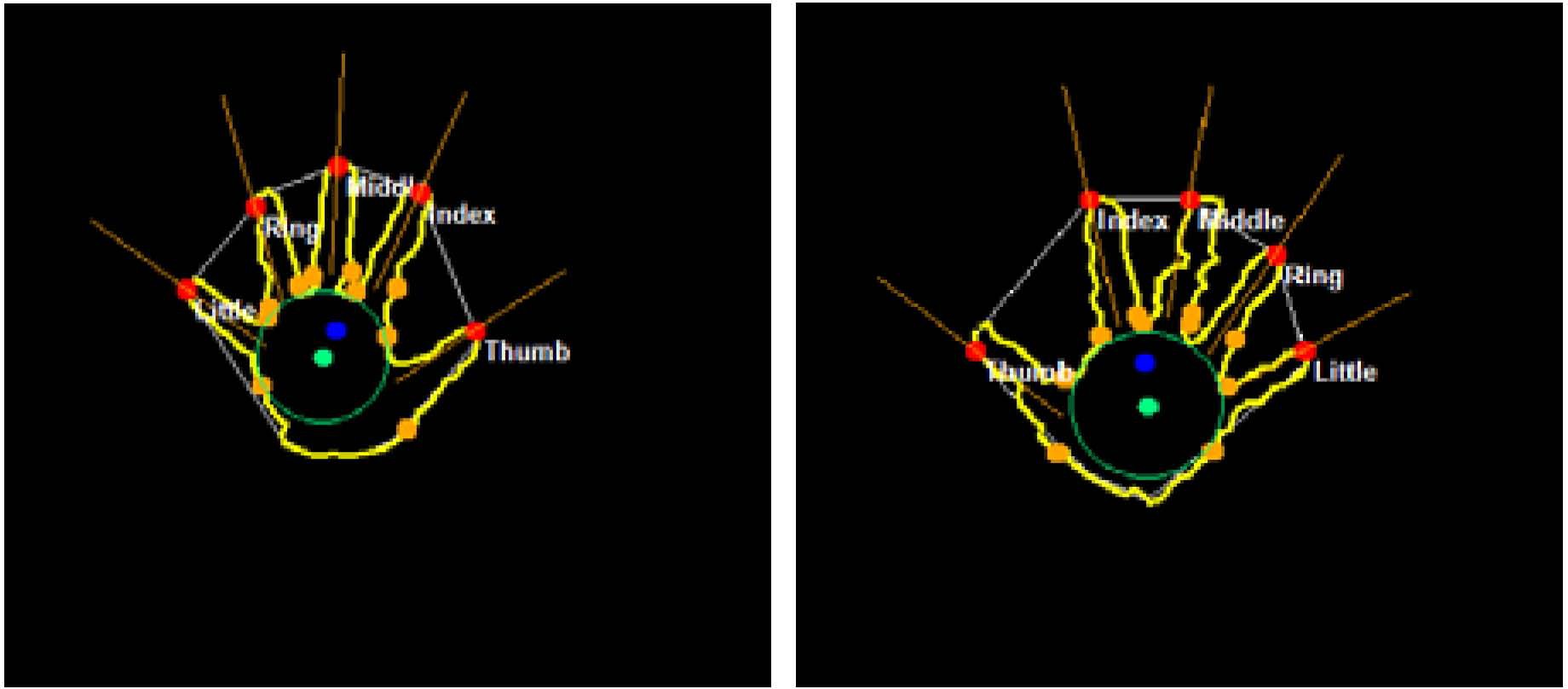
Tres capas de clasificadores:

1) dedo contando clasificador: Los gestos son clasificados por primera vez por el número de dedos extendidos, y luego enviado a la correspondiente segunda capa de clasificadores.

2) Nombre del dedo de recolección: Los gestos son clasificadas por los nombres de los dedos extendidos. Si el combina- ción de los dedos extendidos en un gesto es único entre todos los gestos, termina el proceso de reconocimiento, y el significado de este gesto se muestra; de lo contrario, el gesto es enviado a la correspondiente tercera capa de clasificadores.

3) Vector coincidente: Los vectores de dirección de todos los dedos extendidos son medidos, y todos los ángulos igualitario entre los dedos extendidos son calculados. El significado de este gesto es ahora clasificado según estos ángulos. Entonces, el significado del gesto se asigna y se visualiza en la pantalla.

Un ejemplo de reconocer la "victoria" se ilustra en la figura 4.



198

(A)  (B)  (C)

(D)  (E)  (F)

(G)  (H)  (I)

Fig. 5. Los resultados del reconocimiento. 5(a) es la del "Gesto" con el propósito de calibración; 5(b) es la "victoria"; 5(c) es "Bueno"; 5(d) es la letra "L";  5(e) es la letra "Y"; 5(t) es "I love you"; 5(g) es "Pulgar arriba"; 5(h) es "pulgar abajo"

;  5(i) es el   gesto de Star Trek, que significa "larga vida y prosperidad".

Tabla I HGR EXACTITUD

Ción  oscilan entre 0,5m a 0,8m. El algoritmo k-means clustering se utiliza para obtener dos grupos de píxeles de mano; si la distancia entre los dos medios de clúster está dentro de un umbral, los dos grupos se fusionan, y sólo una parte es detectada. A continuación, el algoritmo de análisis de Graham se utiliza para determinar el casco convexo de manos; un algoritmo de seguimiento del contorno se utiliza para detectar los contornos de la mano. Posteriormente, el candidato alcance son recogidas por calcular los píxeles que son comunes tanto en el polígono convexo y el contorno de la mano, desde el cual se detectan los dedos aplicando el algoritmo de alineación de tres puntos. Ger Fin- nombres están determinados de acuerdo a sus distancias relativas, y un vector de dirección está asignada a cada dedo. Se reconocen los gestos pasándolos por un conjunto de clasificadores que contiene tres capas: dedo contando clasificador, dedo recogiendo y vector nombre coincidente. Un escenario está construido para ser reconocido: Popular gesto, que consta de nueve movimientos populares. El reconocimiento tiene una precisión de al menos 8 a 4% para una mano, mientras que aumenta a más del 90% cuando ambas manos están haciendo los mismos gestos.

El  sistema actual ofrece un nuevo enfoque para HGR en

El  nivel  de  los dedos individuales. Abre posibilidades para HGR

Los gestos

Precisión (%).

Los gestos

Precisión (%).

Pulgar abajo

88,25

Star Trek

84

Sistemas para ser aplicada en muchas formas prácticas en la vida real.

Algunas orientaciones para la labor futura para mejorar la precisión del reconocimiento, para agregar más escenarios como finger- ortografía del alfabeto y los números, y deportes árbitro ges-

Tures. Además, mediante la incorporación de modelos ocultos de Markov, el

El     gesto popular  escenario  consta  de   nueve gestos: " Inicio gesto" (mano abierta), "Pulgar Pulgar arriba", "abajo", "Victoria", "Bueno", "I love you", la letra "L", la letra "S" y "larga vida y prosperidad", el HGR resultados del gesto Popular se muestran en la Figura 5,

D,  resultados

Para probar este sistema, cuatro personas están invitados a realizar cada una de las nueve gestos para 100 veces: 50 veces con la mano izquierda y 50 veces con la mano derecha. La precisión media de cada gesto se calculan y se muestran en la tabla I. Para " Inicio Gesto", la precisión es de casi el 100%, y los nombres de los dedos de las manos están perfectamente identificados. El mejor de los casos es la letra "L" con una precisión del 91%, mientras que en el peor de los casos es "larga vida y prosperidad" con una precisión del 84%. Debido a la dificultad del   gesto de Star Trek, sólo dos de las cuatro personas son capaces de realizar "larga vida y prosperidad". Además, si el firmante utiliza las dos manos para hacer los mismos gestos al mismo tiempo, la exactitud es significativamente aumentado a más del 90% para todos los nueve gestos.

IV. Conclusión Y TRABAJO FUTURO

En este artículo, un HGR sistema basado en Microsoft Kinect para Xbox es introducido. El sistema está motivado por la im- portancia de la comunicación en tiempo real en situaciones concretas, como charlar con personas con discapacidad del habla y la audición. Es capaz de trabajar en la oscuridad, y puede ser fácilmente trasplantadas a otras aplicaciones.

En el sistema desarrollado, se distinguen de las manos

Antecedentes por la información de profundidad; concretamente, la detec-

El sistema puede permitir el reconocimiento continuo de gestos que forman palabras o frases, y la restringida definición escenarios pueden ampliarse a contextos de discurso. Con el reciente lanzamiento de Microsoft Kinect para Windows 7, los sensores son más poderosos que el Kinect para Xbox, que lo hace más adecuado para HGR. Se espera que más aplicaciones con excelentes resultados de reconocimiento será desarrollada.

ACKNOW LEDGMENT

El   autor  desea agradecer al Dr. Ouyang Ming, su asesor, para proporcionar los dispositivos y guiarla por todo el camino a lo largo de esta investigación, que está parcialmente apoyada por FAA conceder 11-G-OlO a Ming Ouyang.

Referencias

[I]   M.  Van den Bergh y L.  Van Gool, "combinación RGB y cámaras ToF para 3D en tiempo real, interacción gestual de mano" en  las aplicaciones de visión por computador (WACV) Taller  sobre IEEE 2011, 2011, págs. 66 -72.

[2]   c. Yang,  Y.  Lang, l. Beh,   D. Hans, y H. Ko, "reconocimiento de gestos

Utilizando mano basado en profundidad para el seguimiento de la aplicación del controlador sin contacto", en  la electrónica de consumo (ICCE), 2012 Conferencia internacional IEEE sobre,  2012, págs. 297 -298.

[3] l. L. Raheja, A. Chaudhary y K. Singal, "Seguimiento de dedos y centros de Palm con Kinect", 2011 Tercera Conferencia Internacional sobre Inteligencia Computacional de simulación de modelado, págs.  248-252, 20 II.

[4]   G.-F. Él, S.-K. Kang, W.-c. Canción y S. T. pulmón, "gesto en tiempo real

Reconocimiento con profundidad 3D, cámara" en  ingeniería de software y servicio

Ciencia (ICSESS), IEEE 2011 2ª Conferencia Internacional de 2011, págs.

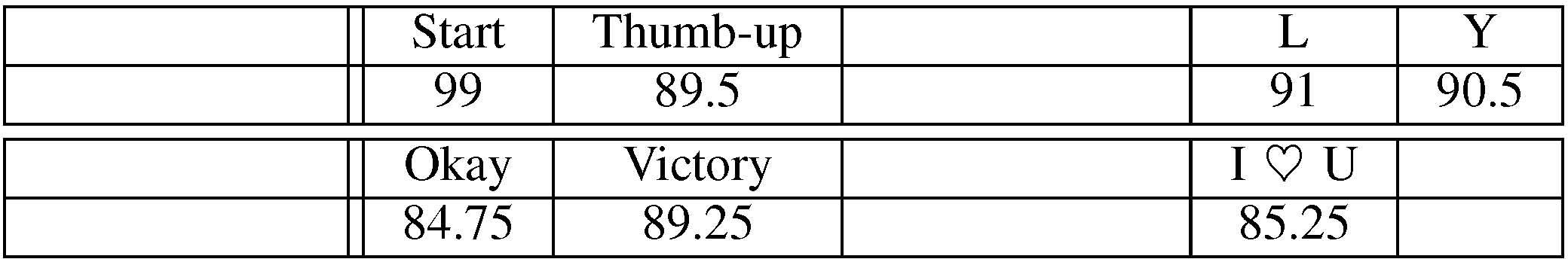
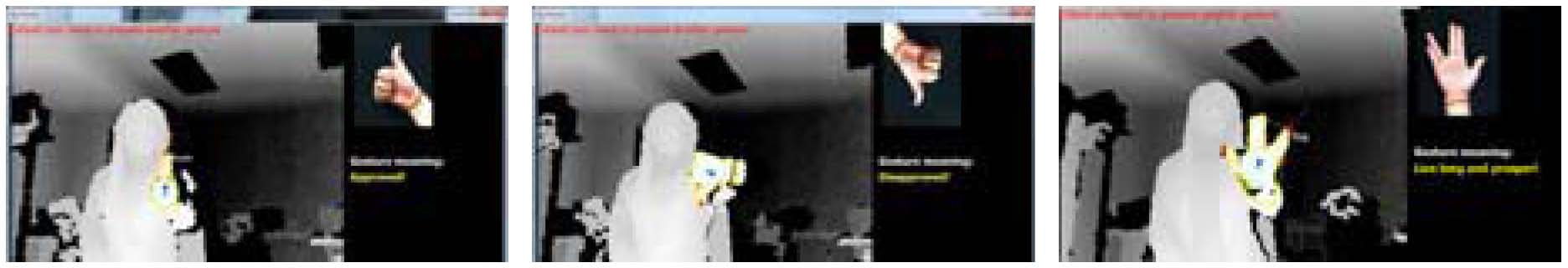
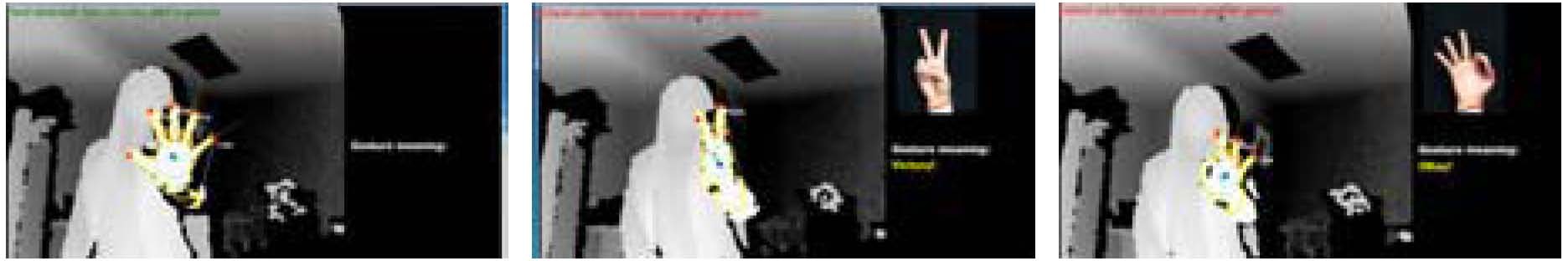
187 a 190.

[5] S. Stegmueller, "Seguimiento de la mano y de los dedos con Kinect la profundidad de datos", 2011, [http://candescentnui.codeplex.com.](http://candescentnui.codeplex.com/)

[6] R. L. Graham, "Un algoritmo eficaz para determinar el polígono convexo

De un conjunto planar finito", cartas de procesamiento de la información, vol.  1, no. 4, págs.

132-133, 1972.



199