Robusto Part-Based Mano Gesto

Reconocimiento usando Kinect Sensor

Ren Zhou, Junsong Yuan*, miembro del IEEE*, Jingjing Meng*, miembro del IEEE*, y Zhengyou Zhang*, Fellow, IEEE*

***Resumen-*El recientemente desarrollado sensores de profundidad, por ejemplo, la Kinect Sensor, han brindado nuevas oportunidades para la interacción humano-ordenador (HCI). Aunque se han hecho grandes progresos aprovechando la Kinect Sensor, por ejemplo, en el seguimiento del cuerpo humano, reconocimiento facial y reconocimiento de la acción humana, sólido reconocimiento de gestos de mano sigue siendo un problema abierto. En comparación con todo el cuerpo humano, la mano es un pequeño objeto con articulaciones más complejas y más fácilmente afectados por errores de segmentación. Es pues un problema muy difícil reconocer los gestos de la mano. Este documento se centra en la creación de una sólida pieza de mano basados en el sistema de reconocimiento de gestos con Kinect Sensor. Para manejar las ruidosas formas mano obtenida de la Kinect Sensor, proponemos un nuevo dis- tancias, métricas de distancia del transportador Finger-Earth (FEMD) para medir la diferencia entre las formas a mano. Ya que sólo coincide con el dedo las piezas aunque no toda la mano, se pueden distinguir mejor los gestos de las manos de leves diferencias. Los extensos experimentos demuestran que nuestro sistema de reconocimiento de gestos de mano es exacta (un 93,2% de exactitud media de un desafiante 10-gesto dataset), ef- ficient (promedio de 0.0750 s por fotograma), resistente a mano las articulaciones, distorsiones y orientación o cambios de escala, y puede funcionar en entornos no controlados (desordenada antecedentes y condiciones de iluminación). La superioridad de nuestro sistema se demuestra además en dos aplicaciones HCI de la vida real.**

***Índice de términos-*Transportador Finger-Earth mano de distancia, reconocimiento de gestos, la interacción humano-ordenador, sistema Kinect.**

I. INTRODUCCIÓN

Y gesto de reconocimiento es de gran importancia para la  interacción humano-ordenador (HCI), a causa de su ampliar-

**H**

Sive aplicaciones en realidad virtual, el reconocimiento de la lengua de signos, y juegos de ordenador [4]. A pesar de mucho trabajo previo, tradicio- nales visión basada en métodos de reconocimiento de gestos de mano [5]-[7] aún están lejos de ser satisfactorios para las aplicaciones de la vida real. Debido a la naturaleza de sensores ópticos, la calidad de las imágenes capturadas es sensible a las condiciones de iluminación y desordenada de fondo, así métodos basados en sensores ópticos suelen ser incapaces de detectar y rastrear las manos vigorosamente, que afecta en gran medida al rendimiento de reconocimiento de gestos de la mano.

Para habilitar una mano más sólido reconocimiento de gestos, uno eficaces es utilizar otros sensores para capturar el gesto de mano

Manuscrito recibió el 15 de julio de 2012, revisada en octubre 17, 2012; aceptado oc- Octubre 22, 2012. Fecha de publicación 25 de febrero de 2013; la fecha de versión actual el 15 de julio de 2013. Una versión preliminar de este trabajo apareció en la ACM Conferencia Inter- nacionales sobre Multimedia [1]. Este trabajo fue apoyado en parte por la Nanyang Ayudante Cátedra SUG M4080134 y regalo de Microsoft Research grant. El editor asociado coordinando la revisión de este manuscrito y ap- demostrando su publicación fue el Prof. Eckehard G. Steinbach.

Z. Ren, J. Yuan, y J. Meng son con la Universidad Tecnológica de Nanyang, Singapur.

Z. Zhang es con Microsoft Research, Redmond, WA 98052 USA. Las versiones en color de una o más de las cifras que figuran en este documento están disponibles en línea

En [http://ieeexplore.ieee.org.](http://ieeexplore.ieee.org/)

Digital Object Identifier 10.1109/TMM.2013.2246148

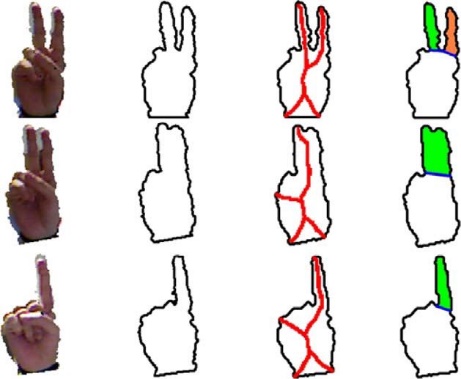


Fig. 1. Las dos primeras columnas ilustran tres casos difíciles para la ges- ture reconocimiento usando Kinect Sensor, donde las primeras dos manos tienen el mismo gesto, mientras que la tercera se confunden el reconocimiento. Utilizando el esqueleto repre- sentation se muestra en rojo en la tercera columna [2], los dos últimos gestos conducen a esqueletos muy similares, con lo que el esqueleto basado en el algoritmo de coincidencia [3] clasifica como el mismo gesto. En la última columna, la parte basa representaciones están ilustrados. Mediante la propuesta de métricas de distancia, *distancia del transportador Finger-Earth*, podemos clasificar las primeras dos manos como el mismo gesto y manejar las ruidosas formas mano obtenida por Kinect Sensor.

Y el movimiento, por ejemplo, mediante el guante de datos [8]. A diferencia de sen- sors óptico, estos sensores son normalmente más fiables y no se ven afectados por las condiciones de iluminación o desordenada fondos. Sin embargo, se requiere que el usuario nosar un guante de datos y, a veces requiere la calibración, es incómodo de usar y puede obstaculizar la articulación natural del gesto de la mano. Además, esos datos los guantes son generalmente más caros que los sensores ópticos, por ejemplo, en cámaras. Como resultado de ello, no se trata de una población muylar forma de reconocimiento de gestos de mano.

Gracias al reciente desarrollo de profundidad barato cam-

Eras, por ejemplo, la Kinect Sensor [9], nuevas oportunidades para la ges- ture reconocimiento emerge. En lugar de llevar un guante de datos utilizando la Kinect oxígen otambién puede detectar  y segmentar las manos ro- bustly, lo que proporciona una base válida para el reconocimiento de gestos. A pesar de los muchoséxitos en la aplicación decentes r la Kinect Sensor para articulado face reconocimiento [10], el seguimiento del cuerpo humano [11] yreconocimiento de acció n humana [12], es todavía un problema abierto para utilizar Kinect para hy reconocimiento de gestos. Debido a la baja resolución de la Kinect mapa de profundidad, normalmente, de sólo 640 480, aunque trabaja iremosa la vía un objeto grande, por ejemplo, el cuerpo humano, es difícil detectar y segmentar un objeto pequeño de una imagen con esta resolución, por ejemplo, una mano humana que ocupa un espacio muy pequeñortion po de la imagen con articulaciones más complejas. En talcaso, la segmentación de la mano a menudo es inexacta, por lo tanto, may afectar significativamente el paso de reconocimiento.

A mustrate el problema anterior, la primera columna de la Fig. 1

Muestratres ejemplos. Puede verse que los contornos (en la segunda columna) tienen importantes distorsiones locales además de suponer variaciones. Debido a la baja resolución y la inexactitud de la

1520-9210/$31.00 © 2013 IEEE

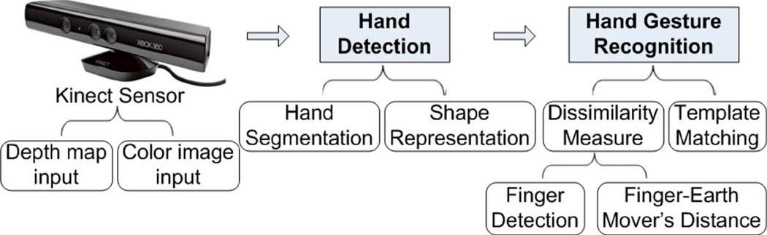


Fig. 2. El marco de nuestra parte de manos a base de sistema de reconocimiento de gestos.

Kinect Sensor, los dos dedos de la segunda mano son indistin- guishable como están cerca uno del otro. Desafortunadamente, los métodos de reconocimiento de formas clásicas, como por ejemplo correspondencia forma basada en algoritmos de coincidencia [13], [14] y el esqueleto de métodos de coincidencia [3], [15], no puede reconocer la forma contorno robustamente con graves distorsiones. Por ejemplo, como se muestra en la tercera columna de la Fig. 1, el rojo esqueletos de los últimos dos manos son muy similares. De ahí el esqueleto algoritmos de coincidencia clasificarlos como el mismo gesto [3].

Evidentemente, reconociendo esa mano ruidosos contornos es desafiante,

Especialmente si hay muchos gestos con la mano para reconocer. A fin de abordar este problema, proponemos una forma novedosa de distancia llamada métrica (Distancia del transportador Finger-Earth FEMD). FEMD está diseñado específicamente para las formas a mano. Ya que sólo coincide con los dedos mientras no la mano entera, pueden manejar mejor la ruidosa mano formas obtenidas por Kinect Sensor.

Fig. 2 muestra el marco de nuestra mano reconocimiento de gestos

Sistema. Utilizamos Kinect Sensor como el dispositivo de entrada que captura la imagen a color y su correspondiente mapa de profundidad. Con la ayuda de profundidad de cue, podemos detectar la mano del usuario enérgicamente a la desordenada antecedentes y condiciones de iluminación. Entonces, nosotros representamos la forma de mano por su dedo de piezas, que es detectado por la forma de la descomposición. Por último, la disimilitud  entre la mano obtenida de forma y cada gesto plantilla es medido por la distancia métrica, FEMD propuesto, para el reconocimiento de gestos.

Para evaluar nuestro método, debemos construir un nuevo dataset desafiante

(Contiene 1.000 casos recogidos en entornos no controlados). Las pruebas sobre este conjunto de datos demuestra que nuestro sistema de reconocimiento de gestos de mano no sólo funciona de forma precisa y eficaz (una media ac- curacy de 93,2% en 0.0750 s por bastidor), pero también es robusta para entornos no controlados y la mano gesto variaciones en la ori- entation, escala, articulación y forma distorsiones. Comparamos nuestro algoritmo con contextos de forma [13], y el esqueleto ruta ilarity sim- [3], en la sección IV-B4 y mostrar nuestra superioridad en la mano reconocimiento de gestos. Además, en la parte superior de nuestro gesto recon- nición algoritmo, construimos dos demostraciones de HCI de la vida real para ilustrar la efectividad de nuestro método en la sección V.

Las principales contribuciones de este trabajo son los siguientes:

•  Proponemos una pieza de mano basados en el sistema de reconocimiento de gestos, basada en una novela de métrica de distancia dedo Earth Mover dis- cia (FEMD). Es robusto a la orientación, la escala, la articula- ción de los cambios, así como las distorsiones de las formas a mano. A nuestro mejor conocimiento, esta es la primera parte de manos a base de sistema de reconocimiento de gestos con Kinect Sensor.

•   Debemos demostrar nuestra mano el algoritmo de reconocimiento de gestos en dos aplicaciones de HCI. El sistema propuesto funciona de forma precisa y eficaz en entornos no controlados. Es aplicable a otras aplicaciones de HCI.

II. REUFÓRICO WORK

Muchos visión basada en el reconocimiento de gestos de mano enfoques han sido propuestos en la literatura [16]-[18], véase [5]-[7] para obtener más comentarios. Visión basada en métodos de reconocimiento de gestos de mano pueden clasificarse en dos categorías. La primera cate- gory es la Estadística el aprendizaje basados en enfoques: por un gesto dinámico, tratándolo como el resultado de un proceso estocástico, la mano reconocimiento de gestos puede abordarse sobre la base de modelos estadísticos, como PCA, HMMs [16], [17] y más avanzadas de filtrado de partículas [19] y [20] algoritmos de condensación. La segunda categoría es la de los enfoques basados en reglas: Regla basada ap- proaches proponer un conjunto de reglas pre-codificado entre características de entrada, que son aplicables tanto para gestos dinámicos y estáticos gestos. Cuando se prueba un gesto de la mano, un conjunto de características que se extraen y se compara con las reglas codificadas, el gesto con la regla que mejor coincida con la entrada de prueba se ve como el gesto reconocidos [18].

Desafortunadamente, todos los métodos de reconocimiento de gestos de mano

Tienen restricciones sobre el usuario o el medio ambiente, lo que dificulta mucho su uso generalizado en aplicaciones reales. Por un lado, para inferir la pose de la palma y los ángulos de las articulaciones, muchos métodos use marcadores de colores para extraer las características de alto nivel, tales como la punta del dedo, joint ubicaciones o algunos puntos de anclaje en la palma [21]-[24]. Por otra parte, algunos métodos propuestos para representar a la región de mano por bordes o una elipse [25]-[27] usando el modelo de color de la piel. Sin embargo, un problema común de los métodos de estas dos categorías es la segmentación de la mano inexacta: ninguno de estos métodos funciona bien en entornos desordenado debido a la sensibilidad de los marcadores de colores y el color de la piel modelo para el fondo. Además, algunos estudios intentan primero completamente re- construir las superficies de mano 3D [8], [28]-[31]. Aunque el

 Los datos 3D proporciona información valiosa que puede manejar problemas como auto-oclusión, una precisa, robusta y en tiempo real 3D recon- struction todavía es muy difícil. Por otra parte, el alto costo computa- cional prohíbe su adopción generalizada.

Afortunadamente, el reciente desarrollo de   sensores de profundidad (por ejemplo, Kinect Sensor) proporciona una solución robusta para la segmentación de la mano. Sin embargo, debido a la baja resolución y la inexactitud del mapa de profundidad obtenidos, el contorno de la mano puede ser bastante ruidoso. Métodos de reconocimiento de formas clásicas no son robustas a graves distorsiones en las formas de mano. Por ejemplo, el contorno de reconocimiento basado en enfoques, como los momentos, no son sólidos cuando el contorno es contaminada por las distorsiones. Esqueleto basado en métodos de reconocimiento [15] también sufren distorsiones de contorno, porque incluso con poco ruido o ligeras variaciones en el contorno a menudo perturbar gravemente la topología de su representación del esqueleto. Bai *et al.*Propuso un método de poda esqueleto en [3], lo que hace que el esqueleto robusto para perfilar el ruido. Sin embargo, métodos basados en el esqueleto aún no puede lidiar con el problema de ambigüedad como se muestra en la Fig. 1, como la segunda y la tercera forma más esqueletos similares a los de la primera y de la segunda forma. En cuanto a la correspondencia basado en métodos de reconocimiento de formas tales como forma contextos [13] y la distancia interior [14], que no son eficaces en la solución de la ambigüedad en la Fig. 1 tampoco, porque las correspondencias de la segunda y las últimas manos tienen más contextos similares que el primero y el segundo.

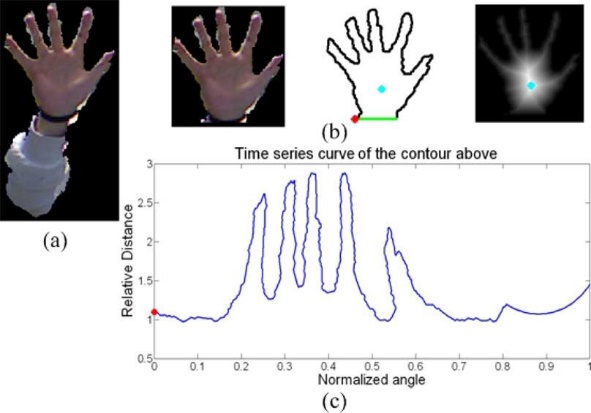


Fig. 3. Detección de manos. (A) El lado rugoso segmentada por umbral de profundidad; (b) una más precisa detección de mano con correa negra (la línea verde), el punto inicial (el punto rojo) y el punto central (el punto de cian); (c) la representación de la curva de series de tiempo.

III. PARTE-basándose Hy G RECOGNITION ESTURE

Ahora vamos a introducir nuestra parte de manos a base de sistema de reconocimiento de gestos. Fig. 2 ilustra el marco, que consta de dos módulos principales: detección de mano y mano de reconocimiento de gestos.

*A. Detección de manos*

Como se muestra en la Fig. 2, usamos Kinect Sensor como el dispositivo de entrada, que captura la imagen a color y el mapa de profundidad a 640 480. Generalmente la profundidad información derivada de Kinect Sensor es utilizable, pero no es muy preciso en los detalles.

Con el fin de segmentar la forma de mano, en primer lugar , localice la posición de la mano con el  seguimiento de la mano de Windows SDK Kinect fun- ción. Entonces, por definir el umbral de la posición de la mano con un cierto intervalo de profundidad, una mano áspera región puede ser obtenido, como se muestra en la Fig. 3(a). Segundo, se requiere que el usuario use un cinturón negro en la muñeca de la mano de gestos, a fin de seg- mento con más precisión la forma de la mano. Después de detectar los píxeles de color negro, utilizamos RANSAC para colocar una línea para localizar el cinturón negro, como se muestra en la Fig. 3(b). La forma de mano es generalmente de 100  100 pixels de resolución, posiblemente con graves distorsiones.

Tras detectar la forma de la mano, que representamos como una *curva de series de tiempo*, como se muestra en la Fig. 3(c). Esta forma de representación se ha utilizado con éxito para la clasificación y agrupamiento de formas [32]. La curva de series de tiempo registra la distancia relativa entre cada contorno de vértice y un punto central. Definimos el punto central como el punto con la máxima distancia distancia después de transformación en la forma (cian), como se muestra en la Fig. 3(b); y el punto inicial (el punto rojo) se define según la línea RANSAC detectados desde el cinturón negro (la línea verde).

En nuestra serie la representación,  el eje horizontal indica el ángulo entre cada contorno de vértice y el punto inicial rel- adores al punto central, normalizado por 360 . El eje vertical indica la distancia euclidiana entre los vértices del contorno y el punto central, normalizado por el radio del círculo dibujado de máxima. Como se muestra en la Fig. 3, la curva de series de tiempo captura las propiedades topológicas de Niza de la mano, como un dedo de piezas.

*B. Reconocimiento de gestos de mano*

Módulo de reconocimiento de gestos de la mano en la Fig. 2 es la parte más importante de nuestra parte de manos a base de sistema de reconocimiento de gestos. Con la mano forma y su curva de series de tiempo, presentamos cómo robustamente reconocer el gesto de la mano.

*1) la plantilla coincidente:*Utilizamos patrones recon- nición, es decir, el lado de entrada es reconocida como la clase con la que tiene la mínima distancia de disimilitud:

¿Dónde está la entrada a mano; es la plantilla de clase ;

Denota el transportador Finger-Earth propuesto la distancia entre el lado de entrada y cada plantilla. Ahora vamos a introducir el transportador Finger-Earth de distancia.

*2) Mover's Finger-Earth Distancia:*En [33], Rubner *et al.*Presentó una métrica general y flexible, llamado Earth Mover la distancia (EMD), para medir la distancia entre las firmas o histogramas. EMD es ampliamente utilizado en muchos problemas tales como la recuperación de imágenes basada en el contenido y el reconocimiento de patrones [34], [35].

EMD es una medida de la distancia entre dos probabilidad

Las distribuciones. Se nombra después de una analogía física que se dibuja desde el proceso de mover los montones de tierra repartidas por todo un conjunto de ubicaciones en otro conjunto de orificios en el mismo espacio. La ubicación del montón de tierra y el agujero denota la media de cada grupo en las firmas, el tamaño de cada montón de tierra o agujero es el peso del racimo, y la distancia entre un montón de tierra y un agujero es la cantidad de trabajo necesario para mover una unidad de tierra. Para utilizar este problema de transporte como una medida de distancia, es decir, una medida de disimilitud, uno busca la menos costosa de transporte el movimiento de tierra que requiere la menor cantidad de trabajo.

Referencias [36] y [37] aplicada a EMD shape matching y recuperación de contorno, que representa el contorno de un conjunto de las características descriptivas y calcula el conjunto de correspondencias con mínimos costes de EMD entre las características locales. Sin embargo, la EMD existentes basados en algoritmos de coincidencia de contorno tiene dos deficiencias cuando aplicado a mano reconocimiento de gestos:

•  Dos formas de mano difieren principalmente en características globales aunque no

Las características locales. Como se muestra en la Fig. 4(a) y (b), los dedos (características globales) son sus principales diferencias. Además, el gran número de características locales se ralentiza la velocidad de coincidencia de contorno. Por lo tanto, es mejor considerar características globales en la coincidencia del contorno.

•  EMD permite la coincidencia parcial, es decir, una firma y su

Subconjunto se consideran iguales en EMD medir: como en la Fig. 4(c) y (d), el EMD distancia de estas dos firmas es cero porque la firma en la Fig. 4(d) es un subconjunto de la Fig. 4(c). Sin embargo, en muchas situaciones la coincidencia parcial es ilógica, como en el caso de Fig. 4(a) y (b), donde el dedo en la Fig. 4(b) es un conjunto parcial de los dedos en la Fig. 4(a). Evidentemente, deben ser considerados diferentes.

Nuestra distancia del transportador Finger-Earth (FEMD) puede resolver estas dos deficiencias del contorno métodos coincidentes con EMD. Diferente del algoritmo basado en EMD que considera cada característica local como un clúster [36], representamos a la entrada a mano por características globales (clusters) con el dedo. Y añadimos penalidad sobre agujeros vacíos para aliviar las coincidencias parciales en funciones globales.

Formalmente, let     Ser el

Firma de primera mano con   Los clusters, donde  Es   el representante de clúster y Es el peso del clúster;

      Es la segunda mano firma con clústeres. Ahora mostramos cómo representar una curva de series de tiempo como una firma. Fig. 4(e) y (f) muestra la serie temporal

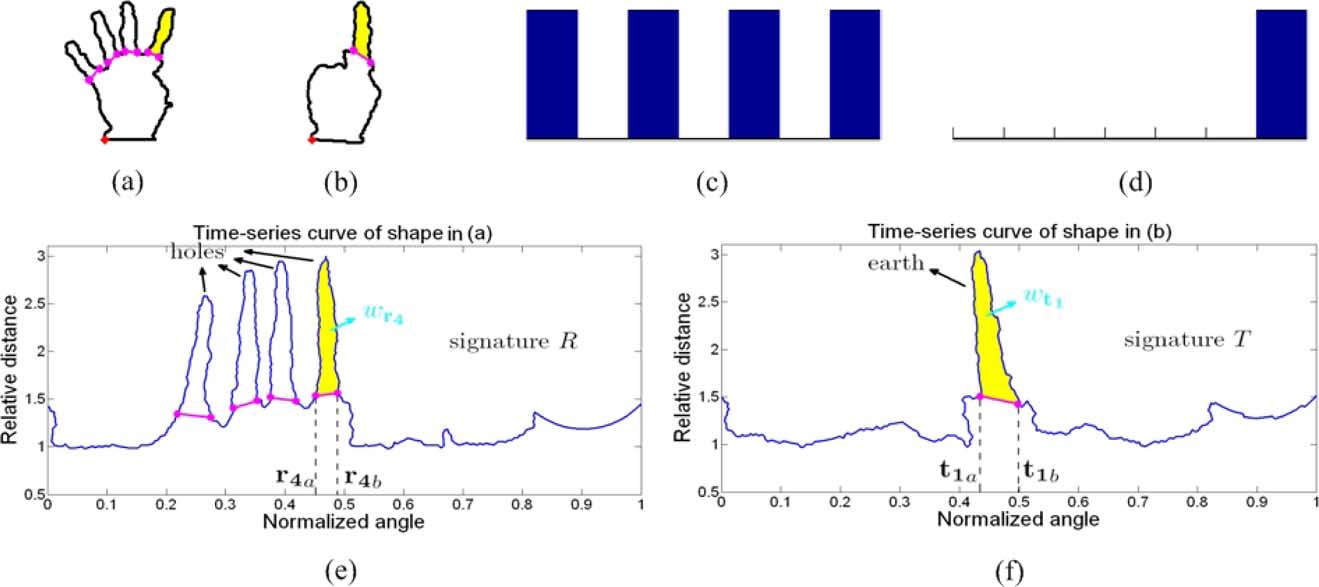


Fig. 4. (A) (b): Dos manos cuyas formas curvas de series de tiempo se muestran en (e) (f). (C) (d): dos firmas que coincidan parcialmente, cuyo costo de EMD es 0. (E) (f): ilustración de la firma representaciones de curvas de series de tiempo.



Fig. 5. Las piezas en color son los dedos detectadas por la propuesta de métodos de detección de dedos: (a) cerca de descomposición convexo, b) definir el umbral de la descomposición.

Curvas de las manos en la Fig. 4(a) y (b), respectivamente, donde cada dedo corresponde a un segmento de la curva. Definimos cada clúster de una firma como el dedo segmento de tiempo-se- ries curve:  el representante de cada grupo  se define como el intervalo de ángulo entre los extremos de cada segmento,

, Donde ; Y el peso de un

Clúster,  , se define como el área normalizado dentro del segmento de los dedos.

Es la planta matriz de distancias de la firma Y

En la sección IV-C. Como podemos ver, , son constantes para las dos firmas; para calcular la FEMD, debemos calcular el valor de  .   Se define por minimizar el trabajo necesario para mover todos los montones de tierra:

, ,

,

, Donde Es la distancia desde la tierra cluster A . Es

Definido como el mínimo intervalo de distancia para móviles , Totalmente a la superposición con  , Es decir:

Totalmente superposición con  , De lo contrario.

Por dos firmas,  y , su FEMD distancia se define como el menos trabajo necesario para mover los montones de tierra, además de la pena en el orificio vacío que no se llena con la tierra:

Donde es el factor de normalización, es el flujo de clúster a clúster , que constituye la matriz de flujo . El parámetro modula la importancia entre el primer y el segundo trimestres. Vamos a investigar los efectos de la

Seguimos la definición de la matriz de flujo en EMD, también tenemos la intención de encontrar el trabajo mínimo necesario para mover los montones de tierra. La primera restricción restringe el flujo de movimiento en una dirección: desde los montones de tierra a los orificios. La última restricción obliga a la máxima cantidad de tierra posible de ser movido. Vamos a demostrar la superioridad de la FEMD más EMD por coincidencia de contorno en la sección IV-C.

*C. Detección de dedos*

Para medir la distancia entre la mano FEMD formas, necesitamos para representar la forma de la mano como una firma con cada dedo como un grupo, es decir, detectar el dedo de la mano de piezas de forma. En la Fig. 5, proponemos dos métodos de detección de dedos para obtener piezas con el dedo de la mano formas. Ahora nos intro- duce estos dos algoritmos:

*1) Descomposición Near-Convex:*Observamos que los dedos tienen una propiedad geométrica común: se encuentran cerca de las piezas-convexos

La forma de la mano. Por lo tanto, ajustar la mínima cerca- convexo (descomposición MNCD) propuesto en [38], [39] para un método de detección de dedos, que se ilustra en la Fig. 5(a):

El objetivo del primer término en la función objetivo es reducir las partes redundantes que no son los dedos de las manos, y el segundo término es mejorar la naturalidad visual de la descomposición. El parámetro equilibra la influencia entre el primer y el segundo término.

Vamos a investigar los efectos de la  sección IV-C.

*2) definir el umbral de la Descomposición:*Aunque cerca-convexo de composición algoritmo puede detectar el dedo piezas con precisión, generalmente está complejamente formulado y no pueden ser resueltos en tiempo real. Por lo tanto proponemos una alternativa de métodos de detección de dedos que son más eficientes, denominado umbralizaciï descomposición, como se muestra en la Fig. 5(b).

Como se mencionó antes, la curva de series de tiempo revela una mano la información topológica. Como se muestra en la Fig. 5, cada dedo corresponde a un pico en la curva. Por lo tanto, podemos aplicar la información de la altura de la curva en forma de series temporales para descomponer los dedos. Específicamente, definimos un dedo como un segmento en la serie temporal curva, cuya altura es mayor que un umbral  . De esta manera, podemos detectar rápidamente los dedos. No obstante, elegir una buena altura umbral  es esencial. Vamos a investigar los efectos de la  sección IV-C.

IV.Valoraciones E

*A. Dataset*

Recogemos un gesto de mano nuevo dataset usando Kinect Sensor[(http: // eeeweba.ntu.edu.](http://eeeweba.ntu.edu.sg/)sg / computervision / Personas / inicio / Ren Zhou / HandGesture.htm). Nuestro conjunto de datos recopilados de

10 asignaturas, y contiene 10 gestos para el número 1 al 10. Cada objeto realiza 10 poses diferentes para el mismo gesto. Así, en total nuestro dataset tiene

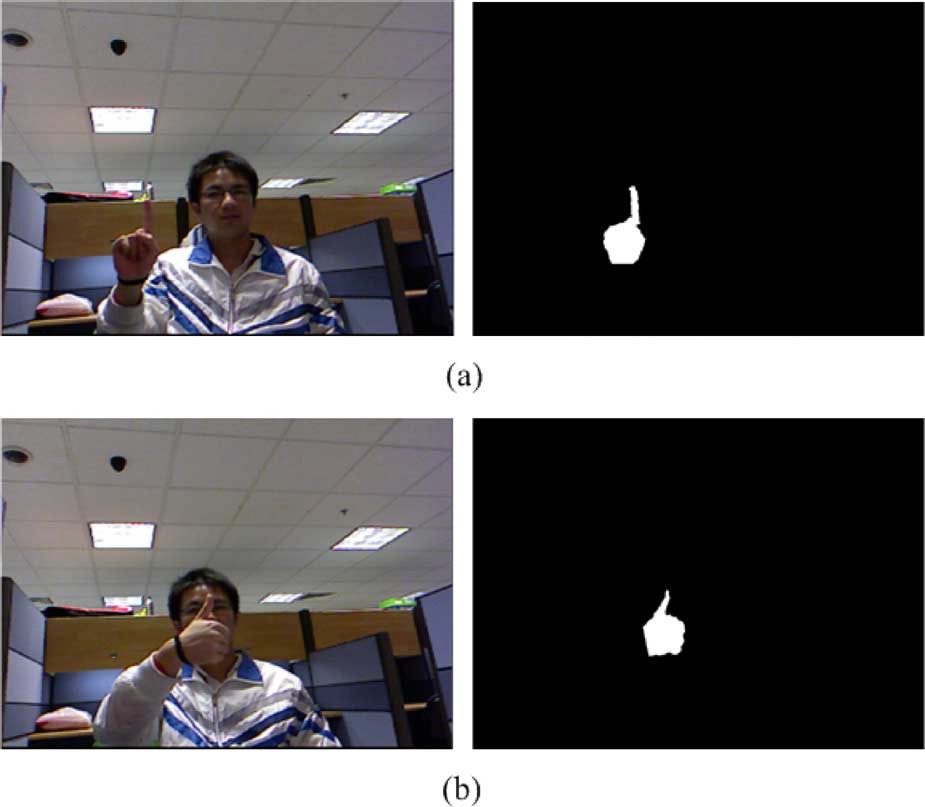
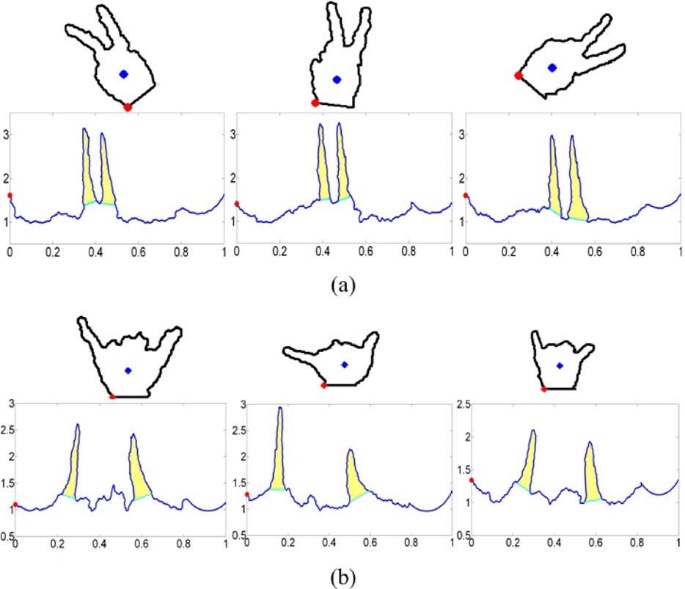


Fig. 6. Nuestro sistema es sólido al desorden en los fondos. (A) La mano que clut- trado por el fondo puede ser detectado con exactitud; (b) la mano que está lleno de cara puede ser detectada con precisión.

, Cada uno de los

Que consta de una imagen en color y el correspondiente mapa de profundidad. Nuestro dataset es un importante desafío en el dataset de la vida real, que se recoge en abarrote fondos. Además, de cada gesto, el tema plantea con variaciones en la mano la orientación, la escala, la articulación, etc.

*B. Evaluación del rendimiento*

Todos los experimentos se realizaron en un procesador Intel Core™2 Quad 2.66

GHz con 3 GB de RAM. Ahora podemos evaluar el rendimiento de nuestro sistema a partir de los siguientes aspectos:

*1) la robustez para abarrote Antecedentes:*Nuestra mano es robusto sistema de reconocimiento de gestos de lleno los fondos, porque la forma de mano se detecta mediante la información de profundidad, por lo tanto, los fondos pueden ser fácilmente eliminados. Fig. 6(a) muestra un ejemplo cuando la mano está saturado por el fondo, lo cual es difícil para otros métodos de reconocimiento de gestos de mano que utilizan marcadores de colores para detectar la mano. En la Fig. 6(b), muestra un caso difícil para el color de la piel de manos a base de métodos de reconocimiento de gestos, donde la mano está saturado por la cara del usuario.

Fig. 7. Nuestro método es robusto a la orientación y los cambios de escala. (A) Las manos

Con los cambios de orientación, y sus curvas de series de tiempo; (b) las manos con los cambios de escala y sus curvas de series de tiempo.

Sin embargo, nuestra mano la segmentación es muy precisos utilizando Kinect Sensor, como se muestra en la columna de la derecha de la Fig. 6.

*2) la  robustez para distorsiones y variaciones en la mano de ori-*

*Entation, escala, articulación:*En el ambiente de la vida real, una mano puede tener variaciones en la orientación, la escala y la articulación. Ser- lados, debido a la limitada resolución del mapa de profundidad, la mano formas siempre están distorsionados, o ambiguos. Sin embargo, podemos demostrar que la propuesta de disimilitud métricas de distancia, distancia del transportador Finger-Earth (FEMD), no sólo es resistente a los cambios de escala y orientación de la mano, pero también insensi- ble a las distorsiones y las articulaciones.

Fig. 7(a) muestra 3 manos con diferentes orientaciones. Como hemos

Puede ver, el punto inicial (el punto rojo en la figura) y el punto central (el punto azul) son relativamente fija en estas formas. Así, las series de tiempo de  las curvas de esas manos (segunda fila en la Fig. 7(a)) son similares, y las distancias son muy pequeñas. En

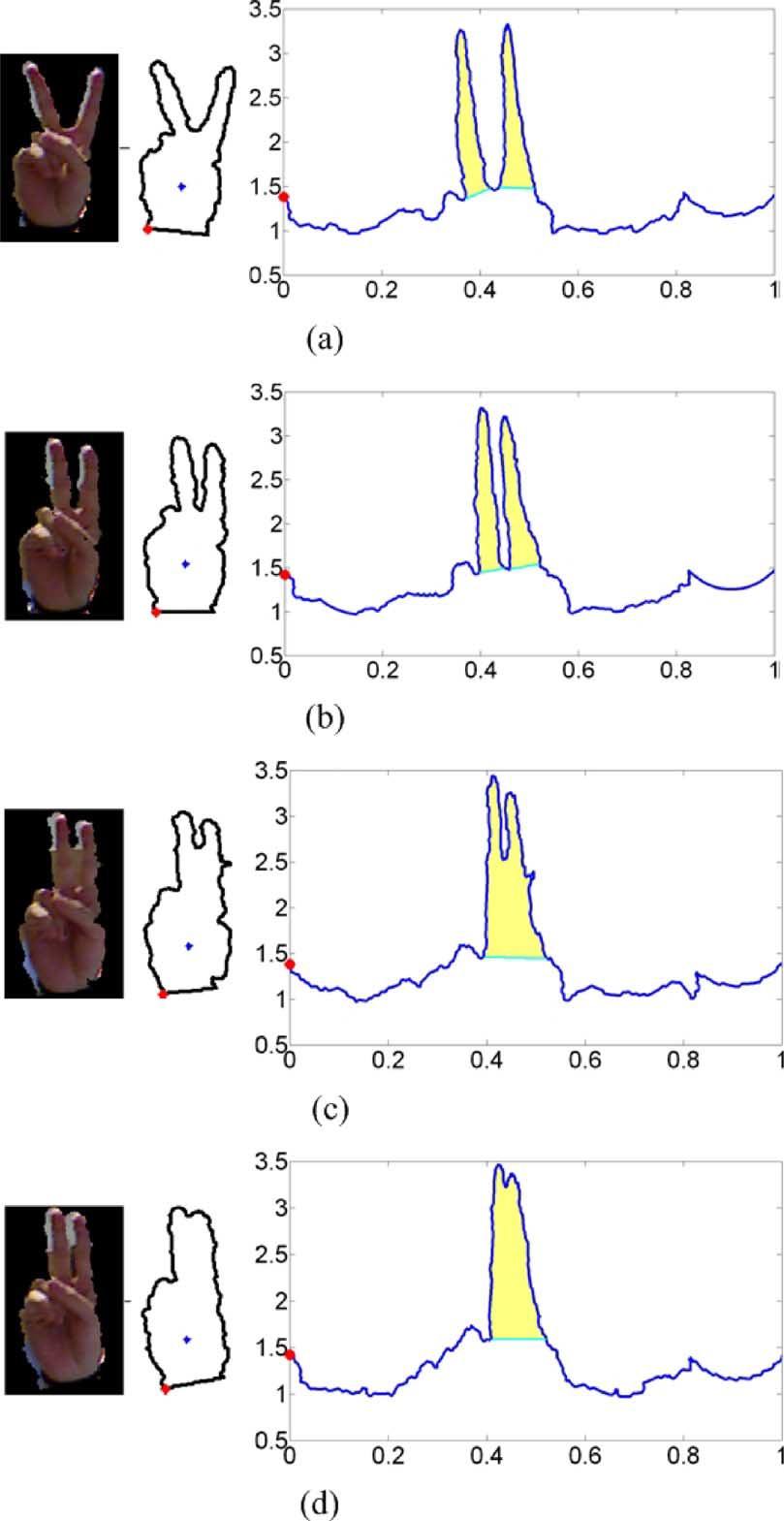


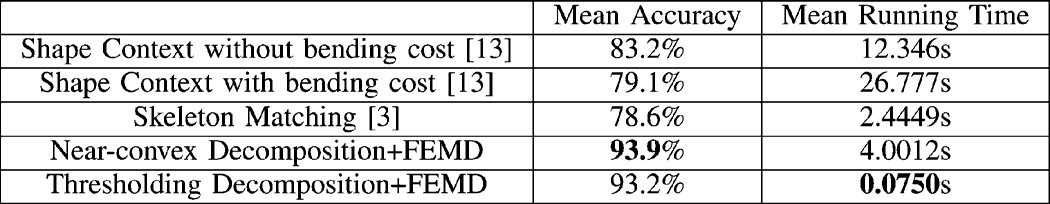
Fig. 8. Nuestro sistema es insensible a las distorsiones y la articulación.

Fig. 7(b), hay 3 manos de diferente tamaño. Porque el tiempo- serie curva y la distancia FEMD se normalizan, están correctamente reconocida como el mismo gesto. Por lo tanto podemos clude FEMD- que es robusto a la orientación y los cambios de escala.

Además, nuestro método de reconocimiento de gestos de mano es robusta a las articulaciones y distorsiones por lado imperfecto de la segmentación. Desde la propuesta de métricas de distancia FEMD utiliza funciones globales (dedos) para medir la disimilitud, distorsiones locales son tolerables. En cuanto a las articulaciones, Fig. 8 se muestra un ejemplo: la columna izquierda muestra 4 imágenes de mano del mismo gesto; la columna del medio muestra las correspondientes formas de mano; y la columna de la derecha muestra sus curvas de series de tiempo. Como podemos ver, la mano formas de Fig. 8(c) y (d) están fuertemente distorsionado. Sin embargo, como se muestra en la columna de la derecha de la Fig. 8, detectando el dedo (el amarillo de piezas re- giones), representamos a cada forma como una firma cuyos racimos son piezas con el dedo. En particular, las firmas de la Fig. 8(a) y (b) tienen 2 grupos:  ,  y las firmas de la Fig. 8(c) y (d) sólo tienen 1 cluster:  . En la sección III-B2, se puede estimar que  , y la tierra  , distancia   . Según la definición, sabemos que las distancias entre las FEMD . Por lo tanto, nuestro FEMD métrica es insensible a las distorsiones y las articulaciones.

Tabla I

E L M ACCURACY EAN y el M R  TIME UNNING EAN DE SHAPE CONTEXTS, SESQUELETO  MATCHING, Y  N uestra Métodos.N uestro PARTE-basándose  Hy G  RECOGNITION ESTURE  System  utilización  UTPERFORMS FEMD O T  S MATCHING RADITIONAL HAPE  UNLGORITHMS



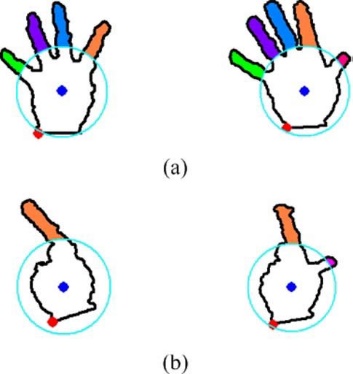


Fig. 9. Dos pares de confundir los gestos en el Experimento 1 (un) Gesto 4 y 5. (B) Gesto 1 y 8.

*3) la exactitud y eficiencia:*a fin de evaluar la exactitud y eficacia de nuestro sistema, se realizaron dos experimentos en el nuevo conjunto de datos. Experimento I utiliza umbralizaciï decompo- sición como se explica en la sección III-C2 para detectar el dedo piezas para medición y experimento FEMD II utiliza cerca de descomposición convexo, como se ilustra en la sección III-C1 para detección de dedos FEMD.

*Experimento I: Umbralizaciï Descomposición FEMD:*En ex-

Periment I, podemos fijar el umbral de altura Y LA FEMD

El parámetro .

Fig. 10 es la matriz de confusión del experimento I. La exactitud media es del 93,2%. Como se ve, las dos categorías más confuso gesto gesto son 5 y 4, 8 y 1 de gestos. Fig. 9 muestra dos casos confusos de estas categorías.

Porque el pulgar es más corto y pequeño, si descomponiendo las manos sólo por un umbral de altura, importantes regiones de dedo puede ser perdido en algunos casos. Como se muestra en la Fig. 9, los pulgares no están bien descompuesto. Como resultado, la FEMD distancias de estos dos casos son muy pequeñas, que confunden el reconocimiento.

Sin embargo, limita el umbral de la descomposición es rápida. Además, debido al pequeño número de características globales extraídos, FEMD opera ef- ficiently. El cuadro I indica el promedio de tiempo de funcionamiento de una mano recogni procedimiento- ción en el experimento I, 0.0750 s. Cabe señalar que nuestro algoritmo FEMD es principalmente reescrito y optimizado en C++, en lugar de simplemente utilizando Matlab como en nuestro trabajo anterior [1], que lleva a los mejores resultados de [1]. Como podemos ver, la fijación de umbrales de composición basada FEMD se ejecuta en tiempo real.

*Experimento II: Descomposición Near-Convex FEMD:* para descomponer con más precisión los dedos de las manos, debemos realizar otro experimento que detecta los dedos con cerca-convexo en descomposición. Aquí hemos solucionado el cercano-convexo de la composición  y el parámetro parámetro FEMD

. Fig. 12 muestra algunos resultados de detección de dedos de nuestro corto

Algoritmo de descomposición convexa. Como podemos ver, el dedo piezas

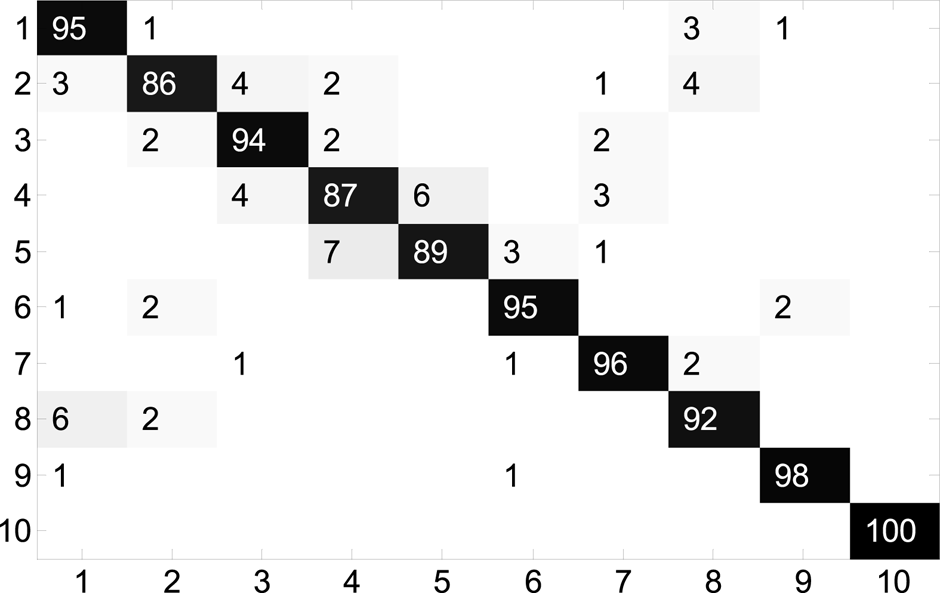


Fig. 10. La matriz de confusión del experimento I.

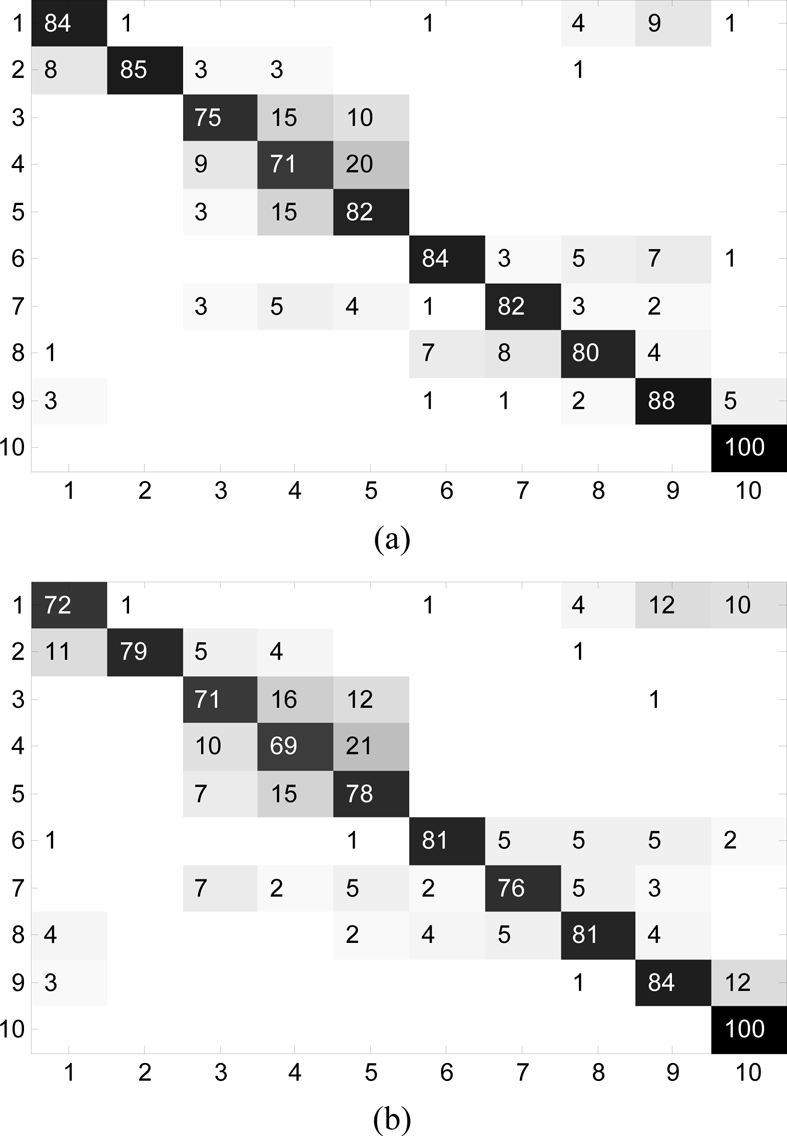


Fig. 13. La confusión de la matriz de reconocimiento de gestos de mano utilizando la forma Contexto [13]. (A) es el resultado del reconocimiento tomografía sin doblar el coste, y (b) es el resultado calculado con curvatura de costo.

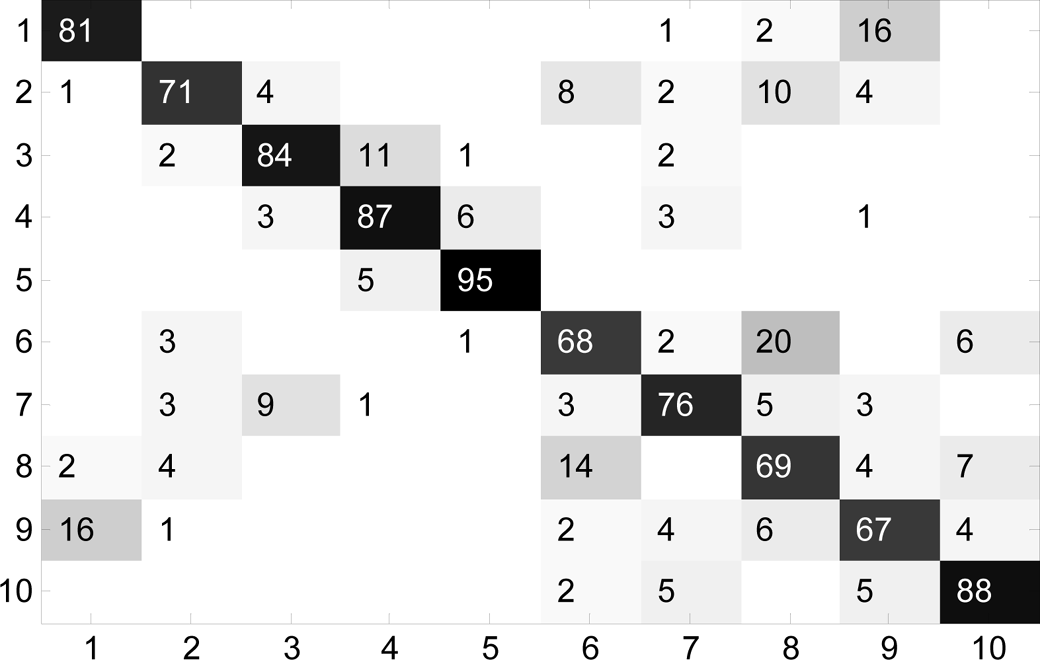
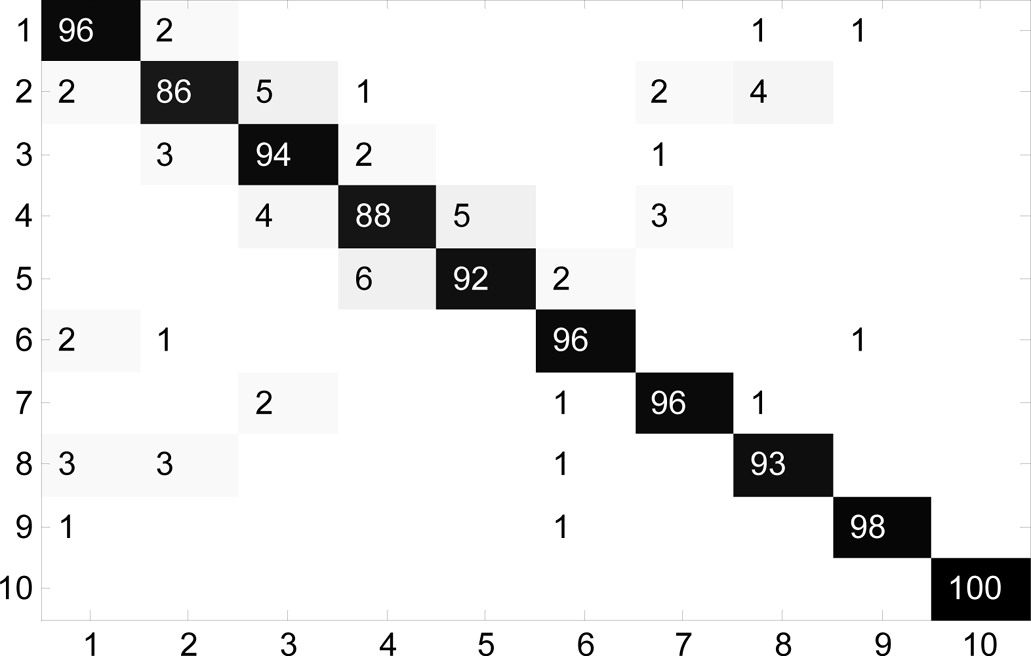
Fig. 11. La matriz de confusión del experimento II.



Fig. 12. Detección de dedos resultados del experimento II utilizando cerca-convexo decom- posición del algoritmo.

Los resultados de detección son más precisos que umbralizaciï decompo- sición.

Fig. 11 muestra la matriz de confusión del experimento II. Allí

No confundir seriamente las categorías. En la cuarta fila de la tabla I, la exactitud media y el promedio del tiempo de funcionamiento del experimento II se dan. La exactitud media de experimento II (93,9%) es mayor que la del experimento I (93,2%), debido a la descomposición de dedo más precisa. Por otra parte, la velocidad de 4.0012 experimento II (s) es más lento que el del experimento I (0.0750), debido al algoritmo de detección de dedos más complejos.

*4) la comparación con otros métodos:*FEMD está basado en una pieza

Mano métricas coincidentes. Vamos a comparar la correspondencia tradicional basado en el algoritmo de coincidencia, forma Contexto [13] y el esqueleto basado en el algoritmo de coincidencia, Ruta semejanza [3]. Su media de precisión y significa tiempos de ejecución figuran en

Fig. 14. La confusión de la matriz de reconocimiento de gestos de mano mediante un esquema de matching [3].

Tabla I. Nos pre-segmento forma la mano utilizando el mismo método como el nuestro en la sección III-A.

Fig. 13 ilustra la confusión matrices de forma Contexto [13]. Tanto de la Fig. 13(a) y (b), encontramos que la mayoría de con- fusión gesto clases 3, 4 y 5. La razón es que la aleta- gers son más fácilmente distorsionado en estas clases, haciéndolos en- distinguible, que hemos discutido antes en las Figs. 1 y

8. Fig. 15 muestra algunos casos confusos de la forma contexto donde

Las formas son localmente distorsionado.

De las dos primeras filas de la tabla I, observamos que, considerando el coste de transformación de flexión TPS empeora el reconocimiento

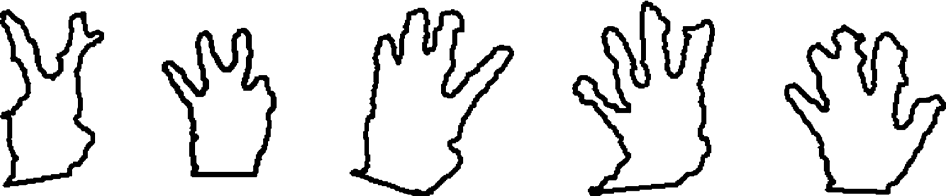


Fig. 15. Algunos casos confusos de la forma contexto [13], donde las formas son localmente distorsionado.

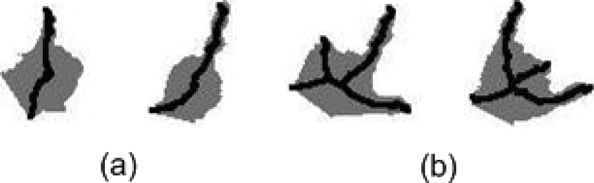


Fig. 16. Algunos casos confusos del esqueleto matching [3], donde muy diferentes formas similares de esqueletos.

Rendimiento en comparación con los contextos de forma calcula sin doblar el costo. La razón es que para ser invariante de rotación, forma contexto debe tratar el vector tangente en cada punto del eje positivo para el log-polar del bastidor del histograma. De cualquier modo, desde nuestra forma es binario, una pequeña variación en la forma podría causar cambios severos de los vectores tangentes en puntos de la forma. Por lo tanto añadir TPS doblando el costo empeora el rendimiento.

Fig. 14 muestra la matriz de confusión de esqueleto coincidentes. En primer lugar debemos podar el esqueleto ruidoso utilizando el método propuesto en [2] y coincidir con ellos mediante ruta similitud propuesto en [3]. En la figura, observamos que muchos gestos están gravemente confundida, como gesto de entre 1 y 9, 6 y 8 de gestos. La razón es que en esos casos, sus esqueletos son muy similares en cuanto a estructura mundial. Como se muestra en la Fig. 16, muy diferente de los gestos de la mano en (a) (b) tienen esqueletos muy similares. Por lo tanto esqueleto algoritmos de coincidencia son incapaces de diferenciar estas clases.

*C. El parámetro Sensibilidad*

En esta sección vamos a evaluar 3 parámetros importantes el umbral de altura  en descomposición umbralizaciï dedo detec- ción de método (Sección III-C2), el parámetro casi en descomposición convexo (Sección III-C1), y el parámetro FEMD formulación (Sección III-B2).

Los resultados se muestran en la Fig. 17. En umbralizaciï decompo- sición,  determina el radio del círculo en descomposición (véase Fig. 5(b)). Si es demasiado pequeño (es decir, ), los dedos no puede estar bien descompuesto; y si es demasiado grande (es decir, ), es- sential dedo regiones se perderán. Fig. 17(a) muestra que podemos obtener el mejor resultado si el ajuste  alrededor de 1,6. En Dedo detec- ción con cerca-convexo, descomposición, equilibra el impacto visual de la naturalidad y el número de piezas. Como se muestra en la Fig. 17(b), si sólo nos minimizar el término naturalidad visual (es decir,

), Obtendremos piezas ruidosas que afectan la mea FEMD- seguro. Además, la curva desciende rápidamente después  porque si mizing mini- el número de piezas demasiado ignorando la nat- uralness visual, podemos obtener las piezas que no son los dedos. En la medida FEMD,  modula la importancia entre los trabajos de movimiento de tierra  y la pena de orificio vacío . Fig. 17(c) muestra que si bien sólo considerando (es decir, ) o sólo considerando (es decir, ), FEMD no puede medir el co- rect disimilitud entre mano formas. Esta curva también justifica

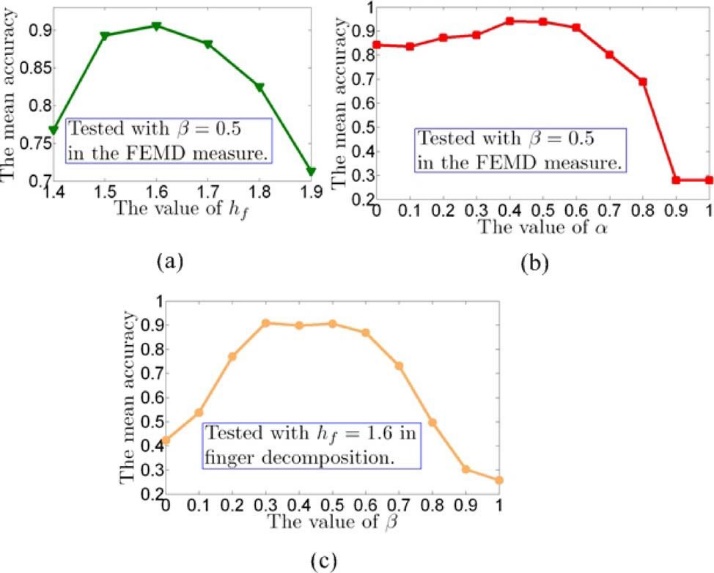


Fig. 17. El parámetro de sensibilidad en , y . Cuando , FEMD pasa a ser la métrica de EMD [33].

Que es mejor que la FEMD EMD (el caso especial cuando )

Para la medida de disimilitud entre mano formas.

V. Aplicaciones

Últimamente ha habido un gran énfasis en el Human-Computer Interaction (HCI) investigaciones para crear interfaces fáciles de utilizar, facilitando la comunicación y habilidades de manipulación natural de los seres humanos. Entre las diferentes partes del cuerpo humano, la mano es la herramienta de interacción más eficaz debido a su destreza. La adopción de la mano gesto como una interfaz HCI no sólo permitirá la des- empleo de una amplia gama de aplicaciones en la sofisticada com- poniendo entornos tales como sistemas de realidad virtual e inter- activa las plataformas de juegos, sino también en beneficio de nuestra vida cotidiana, tales como proporcionar ayudas para las personas con deficiencias auditivas, y mantener abso- laúd esterilidad en los entornos de atención de salud sin contacto con inter- se enfrenta a través de gestos [4].

Ahora nos proponemos utilizar el gesto de la mano como una interfaz

E introducir dos aplicaciones HCI de la vida real en la parte superior de la mano del sistema de reconocimiento de gestos: cálculo aritmético y Piedra-papel-tijeras juego. Cabe señalar que, en el sistema de demostración, realizamos  gestos de manos a base de bastidor recogni- ción mediante la métrica basada en distancia FEMD umbralizaciï descomposición método de detección de dedos, tomando en consideración la precisión y la eficiencia. Como se muestra en la tabla I, se ejecuta en tiempo real y logra ofrecer una precisión comparable de FEMD métrica basada en detección de dedos con cerca-convexo en descomposición.

*A. Cálculo aritmético*

Cálculo aritmético es una interesante aplicación de HCI. En lugar de interactuar con el ordenador mediante el teclado o el ratón, introducimos comandos aritméticos al ordenador a través de los gestos de la mano. Como se muestra en la Fig. 19, 14 los gestos de la mano están definidos para representar 14 comandos, a saber número 0-9 y el operador  ,

, , , Respectivamente.

Reconociendo cada gesto de entrada como un comando, la computadora puede realizar cálculos aritméticos instruido por el

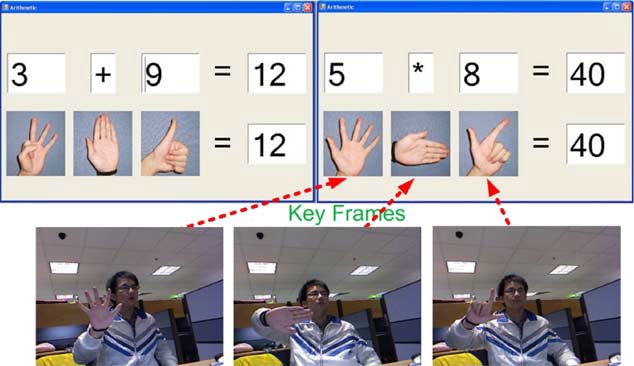


Fig. 18. Cálculo aritmético.

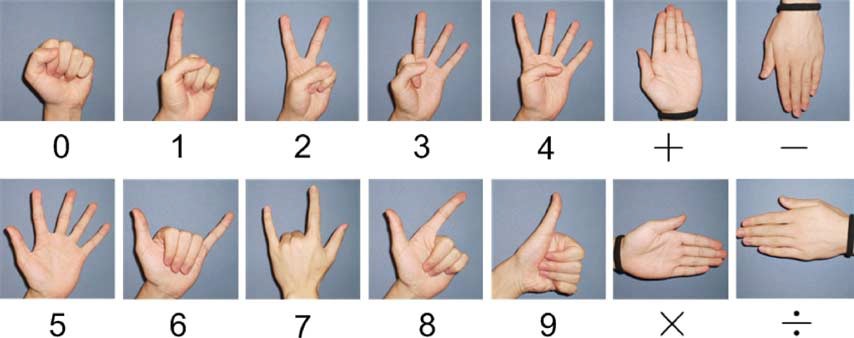


Fig. 19. Los 14 comandos gestuales en nuestro sistema de cálculo aritmético.

El usuario. Dos ejemplos se muestran en la Fig. 18. Los fotogramas clave se muestran tan bien.

*B. Rock-Paper-Scissors Juego*

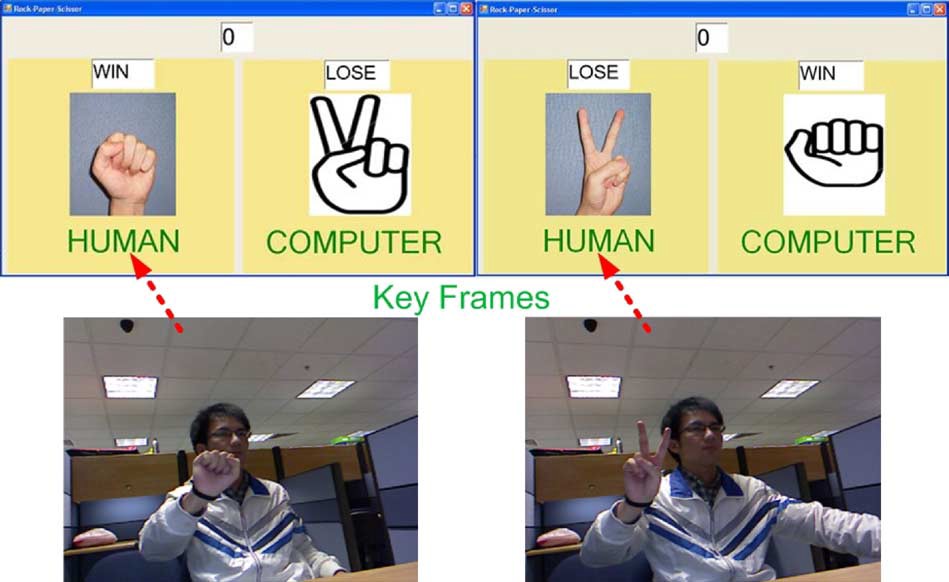
Piedra-papel-tijeras es un juego tradicional. La regla es la piedra rompe las Tijeras Tijeras corta Papel;; y papel envuelve el rock. En esta demostración, construimos una Piedra-papel-tijeras sistema de juego jugado entre un ser humano y una computadora. El ordenador elige aleatoriamente un arma, y el gesto del usuario es reconocido por nuestro sistema. Según la regla de juego, nuestro sistema puede decidir quién es el ganador. Fig. 20 muestra dos ejemplos.

Estas dos demos ha sido demostrada en la ACM Multi-

Media 2011, etc. se ejecuta con precisión en tiempo real. Es factible construir más interesantes demostraciones de Kinect en la parte superior de la mano del sistema de reconocimiento de gestos. El gesto de mano dataset hemos recogido con Kinect Sensor y el vídeo de demostración de la técnica [40] que muestran estas dos aplicaciones están disponibles HCI en <http://eeeweba.ntu.edu.sg/>computervision/personas/home/renzhou.

VI. CY FUTURE ONCLUSION WORK

Reconocimiento de gestos de mano para aplicaciones en la vida real es muy difícil debido a sus requisitos en cuanto a la robustez, precisión y eficacia. En este trabajo, presentamos un robusto basado en parte a mano usando el sistema de reconocimiento de gestos Kinect Sensor. Una novela métricas de distancia, distancia del transportador Finger-Earth (FEMD), se utiliza para la medida de disimilitud, que representa la forma de la mano como una firma con cada dedo parte como un clúster y sanciona el dedo-agujeros vacíos. Extensos experimentos en un desafiante 10-gesto dataset validar que nuestra parte de manos a base de sistema de reconocimiento de gestos es precisa y eficiente. Más específicamente, nuestra FEMD mano basado en el sistema de reconocimiento de gestos alcanza el 93,2% de exactitud media y se ejecuta en 0.0750 s por fotograma cuando se utiliza el método de detección de umbral de dedo de descomposición. Y logra una mejor precisión de 93,9% cuando se utiliza

Fig. 20. Piedra-papel-tijeras juego.

Un método de detección de dedos más precisa, sin embargo, en detrimento de la eficiencia. Teniendo tanto la exactitud y eficiencia en conside- eración, utilizamos umbralizaciï descomposición para detección de dedos en nuestro sistema de demostración en tiempo real.

Una importante contribución de nuestro trabajo es la distancia métrica

Sobre la base de representación basada en parte. Distancia tradicional mea- sures como contextos de forma distancia y ruta similitud no es robusto a las distorsiones y variaciones de forma, ya que sus representantes- sentations, es decir, la forma y el esqueleto, los contextos no son consistentes en el caso de variaciones de mano o graves distorsiones en el local. La distancia FEMD propuesto se basa en una métrica basada en parte repre- sentation que representa la forma de una mano como una firma con cada dedo parte como un clúster. Esta representación permite visualizar la com- putation sobre las características globales, lo que es resistente a las distor- ciones. Y es robusto a la articulación, orientación, cambios de escala, como se indica en la sección IV-B2.

Otra contribución de este trabajo es el de la vida real de apli- caciones HCI hemos construido en la cima de nuestra mano el sistema de reconocimiento de gestos. Demuestra que la técnica de reconocimiento de gestos con la mano podemos imitar las comunicaciones entre humanos, e involucrar a mano gesto como una manera intuitiva y natural de interactuar con las máquinas. Por lo tanto nos puede beneficiar a nuestra vida cotidiana en muchos aspectos, tales como el suministro de ayudas para las personas con deficiencias auditivas, y mantener ab- soluto esterilidad en los entornos de atención de salud mediante en- terfaces sin contacto a través de los gestos.

Nuestra investigación futura se centrará en explorar más eficiente la representación basada en parte, para manejar el problema se muestra en la Fig. 9 y la eficiencia en la devolución de cerca-convexo decomposi- ción basado en el método de detección de dedos. Y vamos a seguir desarrollando interesantes aplicaciones HCI de la mano del sistema de reconocimiento de gestos.

References

[1] Z. Ren, J. Yuan, y Z. Zhang, "reconocimiento de gestos de mano robustos basados en tierra con el dedo del transportador de distancia con una cámara de profundidad de productos básicos", en *Proc. ACM Int. Conf. Multimedia*, 2011, pp. 1093-1096.

[2] X. Bai, L. J. Latecki y W.-Y. Liu, "esqueleto de poda por contornos Particionamiento con la curva discreta evolución", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 29, págs. 449-462, 2007.

[3] X. Bai y L. J. Latecki, "Ruta esqueleto similitud gráfica, coincidentes"

*IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 30, págs. 1-11, 2008.

[4] J. P. Wachs, M. Kolsch, H. Stern, y Y. Edan, "Visión basada en gestos de mano aplicaciones", *Commun. ACM*, vol. 54, págs. 60-71, 2011.

[5] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, y X. Twombly, "Vi-

Sion de manos a base de estimación de pose: una revisión", *Comput. Imagen de visión*

*Entender.*, vol. 108, págs. 52-73, 2007.

[6] S. Mitra y T. Acharya, "reconocimiento de gestos: una encuesta", *IEEE Trans.*

*Syst., hombre, Cybern. C, Appl. Rev.*, vol. 37, pp. 311-324, 2007.

[7] G. R. S. Murthy y R. S. Jadon, "un examen de la visión basada en el reconocimiento de gestos de mano", *Int. J. Inf. Technol. Los conocimientos sobre. Administrar.*, vol. 2, págs.

405-410, 2009.

[8] G. Dewaele, F. y R. Horaud Devernay, "El movimiento de la mano de 3D punto de trayectorias y una superficie lisa, modelo" en *Proc. Eur. Conf. Visión por Computador*, Praga, República Checa, 2004, págs. 495-507.

[9] J. Shotton, A. Fitzgibbon, M. Cook, y A. Blake, "en tiempo real derechos suponen el reconocimiento de piezas de profundidad solo imágenes", en *Proc. IEEE Conf. Reconocimiento de Patrones y visión por computador*, Colorado Springs, CO, USA, 2011.

[10] P. Cai, D. Gallup, C. Zhang, y Z. Zhang, "Seguimiento de la cara deformable en 3D con una cámara de profundidad de productos básicos", en *Proc. Eur. Conf. Visión de redes informáticas*, Creta, Grecia, 2010, pp. 229-242.

[11] KinectHacks, Microsoft Kinect demostraciones de aplicaciones. [Online]. Aprovecha-

Capaz de: [http://kinecthacks.net/.](http://kinecthacks.net/)

[12] J. Wang, Z. Liu, Y. Wu, y J. Yuan, "Minería actionlet ensemble para reconocimiento de acción con cámaras de profundidad", en *Proc. IEEE Conf. Reconocimiento de Patrones y visión por computador*, Providence, RI, USA, 2012.

[13] S. Belongie, J. Malik, y J. Puzicha, "Shape Matching y reconocimiento de objetos mediante la forma contextos", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 24, págs. 509-522, 2002.

[14] H. Ling y D. W. Jacobs, "Forma clasificación utilizando el interno-dis-

Cia", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 29, págs. 286-299,

2007.

[15] K. Siddiqi, S. Bouix, A. R. Tannenbaum, y S. W. Zucker, "Hamilton- Jacobi esqueletos", *Int. J. Comput. Visión*, vol. 48, pp. 215-231, 2002.

[16] A. Wilson y A. Bobick, "paramétricas modelos ocultos de Markov para ges- ture reconocimiento", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 21, págs.

884-900, 1999.

[17] H. Lee y J. Kim, "Hmm un umbral basado en un enfoque modelo para ges- ture reconocimiento", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 21, págs.

961-973, 1999.

[18] M.-C. Su, "Un enfoque basado en reglas fuzzy a mano espacio-temporales ges- ture reconocimiento", *IEEE Trans. Syst., hombre, Cybern. C, Appl. Rev.*, vol.

30, págs. 276-281, 2000.

[19] C. Kwok, D. Fox, y M. Meila, "filtros de partículas en tiempo real," *Proc.*

*IEEE*, pp. 469-484, 2004.

[20] A. Doucet, N. de Freitas, y N. Gordon*, Secuencial en Monte Carlo.*

*La práctica*.   Nueva York, NY, EE.UU.: Springer-Verlag, 2001.

[21] C. Chua, H. Guan, y Y. Ho, "Modelo 3D basados en postura de la mano de esti- mación desde una sola imagen 2D," la *imagen Visión Comput.*, vol. 20, págs.

191-202, 2002.

[22] N. Shimada, Y. Shirai, Y. Kuno, y J. Miura, "gesto de mano estima- ción y perfeccionamiento del modelo utilizando la cámara monocular ambigüedad limita- ción de la desigualdad limitaciones," en *Proc. IEEE Int. Conf. Cara y ges- ture reconocimiento*, Nara, Japón, 1998, pp. 268-273.

[23] E. Holden, "el reconocimiento visual de movimiento de la mano", tesis de Doctorado, Departamento Comput. Sci., Univ. de Australia Occidental, Crawley, Australia, 1997.

[24] Y. Fang, K. Wang, J. y H. Cheng Lu, "una mano en tiempo real Método de reconocimiento de gestos", en *Proc. IEEE Int. Conf. Multimedia y Expo*,

2007, pp. 995-998.

[25] D. Lowe, "apenas parametrizado de modelos 3D para imágenes", *IEEE Trans.*

*Pauta Anal. Mach. Intell.*, vol. 13, pp. 441-450, 1991.

[26] T. Starner, J. Weaver, y A. Pentland, "en tiempo real- lan señas americano meden reconocimiento con escritorio y ordenador portátil basado video", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 20, pp. 1371-1375, 1998.

[27] M. H. Yang, N. Ahuja, y M. Tabb, "Extracción de 2-d motion tra- jectories y su aplicación a mano reconocimiento de gestos", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 29, pp. 1062-1074, 2002.

[28] M. Bray, E. Koller-Meier y L. V. Gool, "filtrado de partículas inteligentes

 Seguimiento de mano 3D", en *Proc. IEEE Int. Conf. El rostro y el gesto recon- nición*, Los Alamitos, CA, EE.UU., 2004, pp. 675-680.

[29] J. Lin, Y. Wu, y T. Huang, "3D basado en el modelo de seguimiento de mano utilizando el método de búsqueda directa estocástico", en *Proc. IEEE Int. Conf. El rostro y el gesto de reconocimiento*, Seúl, Corea, 2004, págs. 693-698.

[30] J. Segen y S. Kumar, "Gesto vr: Visión 3d basados en la interfaz de mano

Para interacción espacial", en *Proc. ACM Int. Conf. Multimedia*, 1998, págs.

455-464.

[31] M. Reale, S. Canavan, L. Yin, K. Hu, y T. Hung, "un sistema de interacción multi-gestos usando un disco iris 3D modelo de estimación de la mirada y una apariencia activa el modelo 3D de mano apuntando", *IEEE Trans. Mul- timedia*, vol. 13, pp. 474-486, 2011.

[32] E. Keogh, L. Wei, X. Xi, S. Lee, y M. Vlachos, "LB\_keogh justifica la indexación de formas exactas bajo invariancia de rotación con representaciones arbitrarias y medidas de distancia", en *Proc. Int. Conf. Bases de datos muy grandes*, 2006, pp. 882-893.

[33] Y. Rubner, C. Tomasi, y L. J. Guibas, "La tierra del transportador de distancia como una métrica para recuperación de imágenes", *Int. J. Comput. Visión*, vol. 40, págs.

99-121, 2000.

[34] R. Datta, D. Joshi, J. J. Z. Li y Wang, "recuperación de imágenes: Ideas, En-

Fluencias, y las tendencias de la new age", *ACM Comput. Surv.*, vol. 40, págs.

1-60, 2008.

[35] J. Zhang, S. Lazebnik y C. Schmid, "características locales y kernels para clasificación de textura y categorías de objetos: un exhaustivo estudio", *Int. J. Comput. Visión*, vol. 73, pp. 213-238, 2007.

[36] K. Grauman y T. Darrell, "Fast contorno aproximado coincidencia usando

Distancia del transportador de la tierra", en *Proc. IEEE Conf. Visión por computador y Pat- tern reconocimiento*, Washington, DC, USA, 2004, págs. 220-227.

[37] H. Ling y L. Okada, "una distancia del transportador de tierra eficiente algoritmo robusto para comparación de histograma", *IEEE Trans. Pauta Anal. Mach. Intell.*, vol. 29, págs. 840-853, 2007.

[38] Z. Ren, J. Yuan, C. W. Li y Liu, "mínimo cerca-convexo decompo-

Sición de forma sólida representación", en *Proc. IEEE Int. Conf. Visión de redes informáticas*, 2011, pp. 303-310.

[39] Z. Ren, J. Yuan, y W. Liu, "cerca de la Mínima forma convexa decompo- sición", *IEEE Trans. Pauta Anal. Mach. Intell.*, que será publicado.

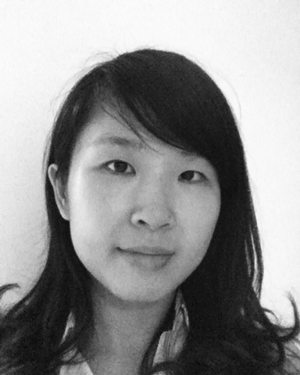
[40] Z. Ren, J. Meng, J. Yuan, y Z. Zhang, "gesto de mano robustos recogni- ción con Kinect Sensor", en *Proc. ACM Int. Conf. MultiMedia*, Scotts- dale, AZ, USA, 2011, pp. 759-760.

**Ren Zhou**es actualmente un estudiante de doctorado en el Departamento de Ciencias de la Computación en la Universidad de California, Los Ángeles. Él recibió el M.Eng. grado con premio de la Escuela de Ingeniería Eléctrica y Electrónica en la Universidad Tecnológica de Nanyang, Singapur, en 2012. Antes de eso, él recibió el B.Eng.  Con el más alto honor del Departamento de Electrónica e Ingeniería de la información en la Universidad de Huazhong de Ciencia y Tecnología, China, en 2010.

De 2010 a 2012, fue jefe de proyecto en el Laboratorio de Tecnología de medios en la Universidad Tecnológica de Nanyang. Su investigación en- terests incluyen Computer Vision, la interacción hombre-máquina, estadística y ma- chine aprendizaje.

**Junsong Yuan**(M'08) es un profesor asistente en la escuela Nanyang de EEE, Nanyang Technology- Universidad de ical (NTU). En la actualidad, sirve como Director del Programa de Análisis de vídeo en Infocomm Centro de Excelencia (INFINITUS), Escuela de la EEE, la NTU. Recibió su doctorado de la Universidad de Northwestern, en Illinois, EE.UU. y M.Ing. de la Universidad Nacional de Singapur. Fue seleccionado para la clase especial para los jóvenes talentosos de la Universidad de Huazhong de Ciencia y Tecnología y recibió un B.Ing. de comunicación Ingeniería en 2002.

La Dra. Sus intereses de investigación incluyen la del yuan de visión por computador, análisis de video, búsqueda visual y minería, la interacción con el ordenador, etc. Ha publicado más de 90 artículos en las principales revistas y conferencias de visión por computador, patrón recon- nición, minería de datos y multimedia. Él sirve como editor, co-presidenta del PC y miembro revisor de numerosas revistas internacionales/conferencias/talleres/reuniones especiales. Recibió el premio de tesis de doctorado CEA pendientes desde el Norte- Western University y el Doctorado Premio Spotlight de IEEE conferencia de Visión por computador y reconocimiento de patrones (Conferencia CVPR'09). Él da tutoriales análisis de la acción humana y análisis de vídeo en algunas conferencias como ICIP'13, FG'13, ICME VRCAI SIGGRAPH'12,'12,'12 y PCM. Él ha presentado tres patentes de EE.UU. y dos patentes estadounidenses provisionales.

**Jingjing Meng**(M'09) recibió la B.E. grado en electrónica e ingeniería de la información de la Universidad de Huazhong de Ciencia y Tecnología, China, en 2003, y la maestría en ciencias de la computación de la Universidad Vanderbilt, Nashville, TN, EE.UU. en 2006.

Desde 2007 a 2010, fue un ingeniero del personal senior de investigación con el Centro de Investigación Aplicada de Motorola, Schaumburg, IL, USA. Actualmente es investigador asociado a tiempo parcial y estudiante de doctorado en la Escuela de Ingeniería Eléctrica y Electrónica.

Ella ha presentado dos patentes de EE.UU. y nosotros dos solicitudes de patente provisional. Sus intereses de investigación actuales incluyen la visión informática, interacción de humanos y ordenadores, y análisis de imágenes y vídeo.

**Zhengyou Zhang**(F'05) obtuvo la licenciatura en ingeniería electrónica de la Universidad de Zhejiang, Hangzhou, China, en 1985, la maestría en ciencias de redes informáticas de la Universidad de Nancy, Nancy, Francia, en 1987, y el doctorado en ciencias de la computación y el Doctorado en Ciencias (Habilitación à diriger des recherches) de la Universidad de París XI, París, Francia, en 1990 y 1994, respectivamente.

Él es un investigador principal con Microsoft re- búsqueda, Redmond, WA, EE.UU., y la investigación- Man ager de la Multimedia, interacción y comu-

Catión (MIC) Grupo. Antes de unirse a Microsoft Research en marzo de 1998, fue con el INRIA (Instituto Nacional Francés para la Investigación en Informática y Control), Francia, durante 11 años y fue un Científico Investigador Senior de

1991. En 1996-1997, pasó un año sabático como investigador invitado en el Instituto Internacional de Investigación de Telecomunicaciones Avanzadas (ATR), Kyoto, Japón. Ha publicado más de 200 documentos en jour- nals arbitradas internacionales y conferencias, y es coautor de los siguientes libros: 3-D escena dinámica análisis: un enfoque basado en estéreo (Springer-Verlag, 1992); Geometría Epipolar en estéreo, el movimiento y el reconocimiento de objetos (Kluwer, 1996); Computer Vision (Academia de Ciencias de China, 1998, 2003, en chino); detección de la cara y la adaptación (Morgan y El Claypool, 2010), y se enfrentan a la geometría y aparecen- ance Modeling (Cambridge University Press, 2011). Le ha dado un número de ponencias en conferencias internacionales.

La Dra. Zhang es el editor en jefe fundador de la IEEE Transactions on au- tonomous desarrollo mental, un editor asociado de la Revista Internacional de Visión por computador, y un editor asociado de visión de máquina y solici- tudes. Se desempeñó como Editor Asociado de la IEEE Transactions on Pattern- ysis anal e Inteligencia Artificial desde 2000 hasta 2004, un editor asociado de la IEEE Transactions on Multimedia desde 2004 hasta 2009, entre otros. Él ha estado en los comités de programa en numerosas conferencias internacionales en las esferas de desarrollo mental autónomo, visión por computador, procesamiento de señales, multimedia y la interacción humano-ordenador. El Programa sirvió como copresidente de la Conferencia Internacional sobre Multimedia y Expo (ICME), en julio de 2010, un programa copresidente de la Conferencia internacional ACM sobre Multimedia (ACM MM), en octubre de 2010, y un programa copresidente de la ACM de la Confe- rencia Internacional sobre interfaces multimodales (ICMI), Noviembre de 2010. Él está cumpliendo una Asamblea copresidente del taller internacional del IEEE sobre Multimedia Signal Processing (MMSP), octubre de 2011.