La Mecatrónica en México, Vol. 3, No. 3, páginas 84 - 91, Septiembre 2014
Disponible en línea en www.mecamex.net/revistas/LMEM
ISSN en trámite, 2014 Derechos de autor y derechos conexos, Asociación Mexicana de Mecatrónica A.C

Un Esquema 3D para la Descripción Visual de Gestos Dinámicos

Medrano-Aguilar José Jesús[1], Avilés-Arriaga Héctor Hugo[2], Gómez-Jáuregui David Antonio[3], Herrera-Rivas Hiram[4] y Nuño-Maganda Marco Aurelio[5]

Universidad Politécnica de Victoria, Parque Científico y Tecnológico de Tamaulipas
Carretera Victoria – Soto la Marina Km 5.5, Cd. Victoria Tamaulipas, México C.P. 87138
[1]1229012@upv.edu.mx, [2]havilesa@upv.edu.mx, [4]hherrerar@upv.edu.mx,[5]mnunom@upv.edu.mx
[3]Université Paris-Sud / LIMSI-CNRS, Orsay, France
[3]Gomez-jau@limsi.fr

Resumen

La interacción humano-robot usando gestos es una forma de comunicación natural para instruir robots de servicio. Sin embargo, la naturalidad y efectividad de este tipo de comunicación depende de una adecuada descripción de los gestos. En la literatura reciente es cada vez más común el uso de atributos tridimensionales debido a la actual disponibilidad de sistemas de bajo costo RGB-D. Por lo tanto, en este documento se describe el desarrollo de un sistema de reconocimiento visual de gestos que usa 7 atributos 3D de postura y movimiento capturados con el Microsoft Kinect. Para la localización y seguimiento del usuario se incluye la calibración del Kinect, eliminación del fondo de la imagen, detección del rostro, medidas antropométricas y seguimiento de la mano derecha por color de piel. Se proponen 5 gestos simples que son reconocidos mediante modelos ocultos de Markov. Para los experimentos se creó una base de datos con 2165 gestos obtenidos por una persona. Las pruebas preliminares muestran un 98.3% de clasificación correcta de estos gestos. Los resultados obtenidos sugieren la viabilidad de implementar el sistema visual de reconocimiento de gestos con atributos 3D para instruir a un robot de servicio.

***Palabras Clave*— Reconocimiento de Gestos, Robots de Servicio, Interacción Humano-Robot.**

I. INTRODUCCIÓN

El reconocimiento visual de gestos es una línea de investigación que ha despertado gran interés en los últimos años. En la literatura es bien sabido que el reconocimiento depende de diferentes elementos como son: i) la elección de los atributos para describir los ademanes y ii) el esquema de reconocimiento. Por un lado, gracias a la aparición de videocámaras RGB-D de bajo costo en la literatura se pueden encontrar cada vez más propuestas que utilizan información tridimensional para describir los gestos [1,2]. Por otro lado, los modelos ocultos de Markov son los modelos probabilistas más usados para realizar la clasificación de los gestos.

Por tanto, en este documento se presenta el diseño y desarrollo de un sistema visual de reconocimiento de gestos que utiliza atributos 3D de un MS Kinect. Esta propuesta es una extensión al trabajo previo descrito en [3] el cual se basa en un sistema monocular RGB y atributos 2D. El reconocimiento de gestos se realiza por medio de modelos ocultos de Markov [4]. Se consideran 5 gestos: a) apuntar a la derecha, b) apuntar a la izquierda, c) saludar, d) acercar y e) detenerse como se puede observar en la Fig. 1.

El sistema incorpora en total 7 atributos de los cuales 3 son de movimiento (desplazamiento) y 4 atributos de postura (posición) de la mano derecha del usuario con respecto a la cabeza y torso. La localización del usuario se realiza por eliminación del fondo de la imagen, detección del rostro y medidas antropométricas. El seguimiento de la mano derecha del usuario se ejecuta por color de piel. Para la fusión de información de profundidad, de color y reconstrucción de la escena se implementó un sistema de calibración para la cámara RGB-D que en este trabajo corresponde al MS Kinect.

Los gestos descritos anteriormente se eligieron para la instrucción de nuestro robot de servicio SerBot II [5]. Para realizar los experimentos se construyó una base de datos con 2165 gestos ejecutados por una persona. Nuestras pruebas iniciales obtuvieron un 98.3% de clasificación correcta de los gestos. Estos resultados muestran la efectividad del sistema visual de reconocimiento y la viabilidad del sistema para instruir a un robot de servicio.

La Mecatrónica en México, Vol. 3, No. 3, páginas 84 - 91, Septiembre 2014
Disponible en línea en www.mecamex.net/revistas/LMEM
ISSN en trámite, 2014 Derechos de autor y derechos conexos, Asociación Mexicana de Mecatrónica A.C

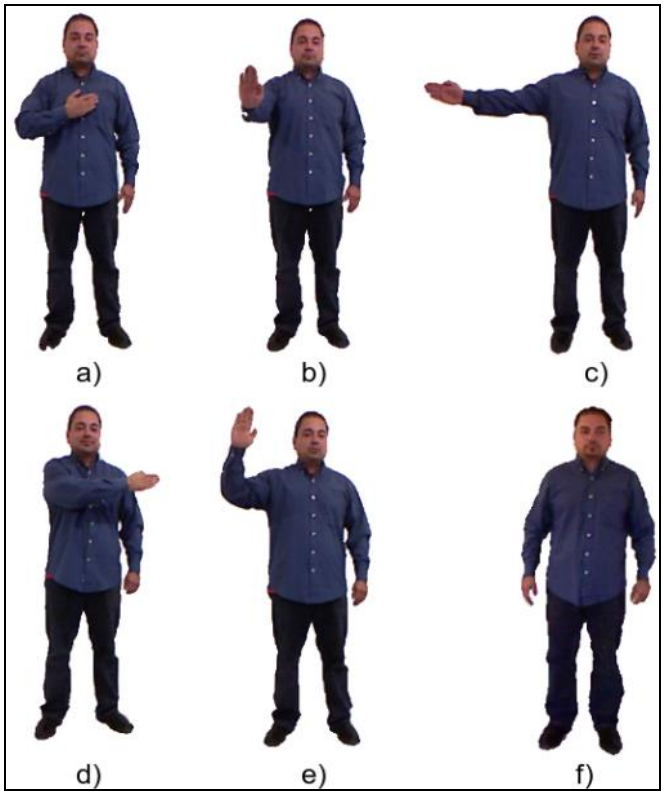

Fig. 1 Gestos propuestos en este trabajo: a) acercar, b) detener, c) derecha, d) izquierda, e) saludar y f) posición de descanso inicial y final de cada gesto.

II. TRABAJOS RELACIONADOS

El uso de dispositivos RGB-D ha aumentado recientemente para el desarrollo de sistemas funcionales de reconocimiento de gestos. En [6] se propone un sistema de reconocimiento de gestos de la mano que con información 3D y adicionalmente lecturas de un acelerómetro que proporciona un seguimiento y pose angular bajo situaciones de oclusión. La nube de puntos de la mano se compara con plantillas para reconocimiento. Palacios et al. [7] proponen reconocer posturas utilizando un cerco convexo para describir la mano. Similar a nuestra propuesta, en [8] se combina información de movimiento y de postura 3D del usuario. Tao et al. [9] describen una combinación de modelos ocultos de Markov multicapa y redes neuronales *backpropagation* para el reconocimiento de gestos continuos usando velocidad angular y aceleración. Estos trabajos utilizan al MS Kinect como dispositivo de entrada. Recientemente, R. Mangera [10] propone el dispositivo Asus Xtion Pro Live para la esqueletización del cuerpo y reconocimiento de 10 posturas tomando como atributos los ángulos de las articulaciones y su posición relativa a la cabeza. El clasificador en este trabajo es *k-means*. Aunque los atributos y métodos de clasificación suelen ser extensiones directas de algoritmos presentes en interfaces de gestos 2D, las propuestas anteriores muestran en general una mejora de la descripción del usuario y un desempeño competitivo en comparación a sus antecesores.

III. METODOLOGÍA

La metodología se divide en 4 partes principales: i) el desarrollo del sistema visual para la captura de gestos, ii) la construcción de la base de datos de gestos, iii) la selección de los atributos para describir los gestos y iv) la definición de los modelos ocultos de Markov para reconocimiento. Cada una de estas etapas se describe a continuación.

La Mecatrónica en México, Vol. 3, No. 3, páginas 84 - 91, Septiembre 2014
Disponible en línea en www.mecamex.net/revistas/LMEM
ISSN en trámite, 2014 Derechos de autor y derechos conexos, Asociación Mexicana de Mecatrónica A.C

3.1. Sistema de análisis visual

Nuestro sistema visual utiliza el MS Kinect para la captura de los gestos. El proceso de análisis visual se divide en tres partes: a) acceso a los datos b) calibración de parámetros intrínsecos y extrínsecos de las cámaras RGB e IR y c) el análisis de las imágenes en la secuencia de video.

a) Acceso a datos. El MS Kinect consiste en una cámara de video RGB para luz visible y una cámara IR que detecta patrones de luz infrarroja (emitidos por el mismo dispositivo). Estos sensores producen imágenes bidimensionales en RGB $I_{RGB}(x,y)$ y mapas de profundidad "cruda" $I_{IR}(x,y)$ del ambiente, respectivamente (x, y, indican las coordenadas de cada pixel). Existen diversas librerías como Kinect SDK de Microsoft, OpenNI+SensorKinect o Libfreenect para el acceso a los datos. En este trabajo se utiliza Libfreenect debido a que es una librería de código abierto con ejemplos comprensibles en lenguaje C, con un buen desempeño y no incorpora algoritmos de procesamiento de imágenes, lo cual nos obliga a comprender las técnicas necesarias para implementar funcionalidades disponibles en otras librerías.

b) Calibración del MS Kinect. Una vez obtenido el acceso, es necesario calibrar los parámetros intrínsecos de las cámaras y estimar sus coeficientes de distorsión radial y tangencial. Con este propósito se implementó una aplicación basada en OpenCV [11] que requiere múltiples vistas de un tablero de ajedrez y supone un modelo geométrico tipo "pinhole" para la cámara. El tablero es de 4 por 7 esquinas y cuadros de 5cm. La calibración se realiza individualmente para cada cámara y se usan 14 imágenes en cada caso. En particular, en la calibración de la cámara IR se bloqueó el emisor infrarrojo y se usó luz natural para iluminar el tablero [12]. Sin embargo, la imagen del sensor IR suele ser oscura y la detección de algunas esquinas del tablero no es siempre precisa. Nuestra aplicación permite la corrección manual de estos errores al momento de la detección usando un ratón.

La Tabla I muestra un ejemplo de parámetros intrínsecos $<f_x, f_y, c_x, c_y>$ obtenidos y redondeados a 3 decimales. Los coeficientes de distorsión $<k_1, k_2, p_1, p_2>$ de la Tabla II se emplean para remover efectos de distorsión en todas las imágenes.

TABLA I. EJEMPLO DE PARÁMETROS INTRÍNSECOS DEL MS KINECT.

	f_x	f_y	c_x	c_y
RGB	5.155e+02	5.161e+02	3.351e+02	2.253e+02
IR	5.968e+02	5.972e+02	3.352e+02	2.267e+02

TABLA II. EJEMPLO DE COEFICIENTES DE DISTORSIÓN DEL MS KINECT.

	k_1	k_2	p_1	p_2
RGB	-1.349e-02	1.310e-01	2.242e-03	-1.837e-05
IR	-1.784e-01	6.239e-01	-5.802e-03	2.001e-03

A partir de los parámetros intrínsecos de la cámara IR y de $I_{IR}(x,y)$ se puede representar la escena como una nube de puntos $P_{IR}(x,y,z)=(X_{xy}, Y_{xy}, Z_{xy})$ en coordenadas "del mundo" y con origen en la cámara IR para la conversión de profundidad cruda a metros. Cada $P_{IR}(x,y,z)$ se obtiene de acuerdo a las siguientes fórmulas:

$$Z_{xyz} = 0.1236 * \tan\left(d_{xy} / 2842.5 + 1.1863\right) \quad (1)$$

$$X_{xyz} = \left(\left(x - c^{IR}{}_x\right) / f^{IR}{}_x\right) * Z_{xyz} \quad (2)$$

$$Y_{xyz} = \left(\left(y - c^{IR}{}_y\right) / f^{IR}{}_y\right) * Z_{xyz} \quad (3)$$

Donde $f^{IR}{}_x$, $f^{IR}{}_y$, $c^{IR}{}_x$, $c^{IR}{}_y$ son los parámetros intrínsecos del sensor IR. La ecuación 1 fue propuesta en [13] para la conversión de profundidad cruda a metros. En nuestra experiencia los valores válidos de profundidad van de 436 a 1020 (.5m a 5m, aproximadamente).

La Mecatrónica en México, Vol. 3, No. 3, páginas 84 - 91, Septiembre 2014
Disponible en línea en www.mecamex.net/revistas/LMEM
ISSN en trámite, 2014 Derechos de autor y derechos conexos, Asociación Mexicana de Mecatrónica A.C

La nube de puntos es muy útil pero se requiere color para enriquecer la reconstrucción 3D. Libfreenect accede a parámetros de fábrica para el *registro* de imagen (el alineamiento uno-a-uno de los píxeles de profundidad y RGB para combinar 3D y color). No obstante, también reduce el campo visual del dispositivo. De esta forma, se optó por una alineación *vía* software que calcula la rotación R y traslación $\boldsymbol{t}$ entre ambos sensores a partir de imágenes del tablero tomadas por ambas cámaras usando OpenCV [14]. Ejemplos de R y $\boldsymbol{t}$ son:

$$R = \begin{pmatrix} 9.997\text{e-}01 & -9.175\text{e-}03 & -2.191\text{e-}02 \\ 9.078\text{e-}03 & 9.999\text{e-}01 & -4.495\text{e-}03 \\ 2.195\text{e-}02 & 4.295\text{e-}03 & 9.997\text{e-}01 \end{pmatrix} \quad (4)$$

$$t = \begin{pmatrix} 3.553\text{e-}02 \\ 6.546\text{e-}03 \\ -1.252\text{e-}01 \end{pmatrix} \quad (5)$$

Estas matrices representan los parámetros extrínsecos de las cámaras con los cuales se transforma cualquier punto $P_{IR}(x,y,z)$ *hacia* $P_{RGB}(x,y,z)$, es decir, como si $P_{IR}(x,y,z)$ fuese visto por la cámara RGB en sus propio sistema coordenado de referencia. La siguiente ecuación realiza esta transformación [15]:

$$P_{xyz}^{RGB} = R^T\left(P_{xyz}^{IR} - t\right) \quad (6)$$

Una vez calculados los puntos $P_{RGB}(x,y,z)$, éstos son proyectados en el plano de $I_{RGB}(x,y)$ por medio de los parámetros intrínsecos de esta cámara. El resultado son nuevas coordenadas $x_{RGB}y_{RGB}$ donde puede consultarse en $I_{RGB}(x,y)$ el color correspondiente a $P_{IR}(x,y,z)$. Un ejemplo en OpenGL [16] del resultado de este procedimiento se presenta en la Fig. 2. Este procedimiento y el sistema visual han sido probados satisfactoriamente en particular los modelos del MS Kinect (1414 y 1473).

Fig. 2 Ejemplo de la reconstrucción 3d del ambiente con registro de imagen.

c) Análisis de imágenes. Para analizar la escena se realizó la reconstrucción de una tercera imagen bidimensional que contiene el registro de cada pixel de color con información de profundidad $I_{IR}(x,y)$ que a través de las coordenadas xy hace accesible a la información 3D de $P_{IR}(x,y)$. Así, esta nueva imagen permite obtener la forma y color que se usó para la localización y seguimiento del usuario. Posteriormente se procede a la eliminación del fondo utilizando el algoritmo de modelos mixtos gaussianos [17]. Para la eliminación del fondo se utiliza información 3D en lugar de 2D. El Sistema requiere de 15 segundos para el aprendizaje. Una vez eliminado el fondo el usuario puede entrar a la escena dando paso a la detección del rostro por medio de algoritmos de clasificación Haar en cascada [18]. En seguida se estiman las subregiones pertenecientes al torso y a la mano derecha del usuario en base a la proporción del rostro. Para la localización de la mano se utiliza un algoritmo que trabaja de la siguiente manera: a) por clasificación probabilístico de pixeles de piel y b) por un algoritmo de crecimiento de regiones para la segmentación de

La Mecatrónica en México, Vol. 3, No. 3, páginas 84 - 91, Septiembre 2014
Disponible en línea en www.mecamex.net/revistas/LMEM
ISSN en trámite, 2014 Derechos de autor y derechos conexos, Asociación Mexicana de Mecatrónica A.C

la piel. El algoritmo trabaja sobre la subregión de la posición inicial de la mano. De manera similar el seguimiento se hace por medio de una ventana que se actualiza en base al centroide de la mano.

Para la etapa de reconocimiento se utiliza una secuencia de observaciones para cada modelo oculto de Markov calculada por medio del algoritmo hacia-adelante. El cálculo de los parámetros del modelo oculto de Markov se encuentra dado por secuencias de entrenamiento en base al algoritmo Baum-Welch. La consulta de los datos de prueba y entrenamiento se lleva a cabo por medio de una base de datos que almacena los gestos en texto plano. Finalmente se muestra como salida en pantalla el tipo de gesto reconocido. Las partes del sistema propuesto se basan en el diagrama de flujo propuesto por Shrivastava [19]. La Fig. 3 muestra la ejecución del sistema de reconocimiento en el que el usuario ejecuta un gesto. Se indica también la detección del rostro y estimación de subregiones del torso y mano derecha. El rectángulo superior de color rojo indica la detección del rostro del usuario. El recuadro azul claro indica la posición del torso. El recuadro inferior rojo indica la posición inicial de la mano. El recuadro azul marino cercano a la esquina inferior izquierda del cuadro del torso marca la zona inicial de búsqueda de la mano. El recuadro amarillo es la región de búsqueda de la mano en cada imagen. Finalmente, el rectángulo verde es el resultado de la segmentación de la mano. Este último rectángulo (rectángulo verde) sirve para la extracción de la información "cruda" de los gestos que se describe en la siguiente sección.

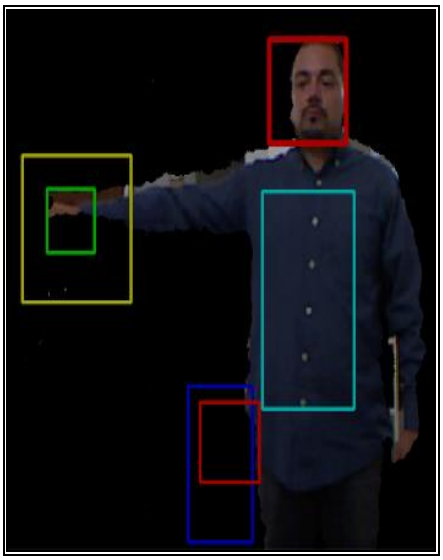

Fig. 3 Ejemplo del sistema visual para el seguimiento de la mano derecha del usuario

3.2. Base de datos

Para este trabajo se generó una base de datos con 5 gestos naturales y un total de 3165 muestras, La ejecución de los gestos fueron realizadas por uno de los autores. Cada muestra está compuesta por el número T de observaciones que componen al gesto (donde $4 < T < 15$) y de las observaciones en secuencia. Cada observación consiste en: a) las coordenadas (x_m,y_m,z_m) de las esquinas superior e inferior del rectángulo que segmenta la mano derecha, b) las coordenadas (x_t,y_t,z_t) de las esquinas superior e inferior del rectángulo que segmenta al torso y c) las coordenadas (x_r,y_r,z_r) del rostro.

Los datos se grabaron en archivos de texto plano. Los gestos comienzan y terminan siempre con el brazo del usuario en una posición de descanso. Nuestro participante ejecutó todos los gestos de frente a la cámara a una distancia de 2 metros aproximadamente. La altura del MS Kinect fue de 1.50 metros en base a la posición cercana a la que se encuentra el MS Kinect en SerBot II. Las observaciones se registraron cada 4 imágenes a una velocidad aproximada de 20 imágenes por segundo.

La Mecatrónica en México, Vol. 3, No. 3, páginas 84 - 91, Septiembre 2014
Disponible en línea en www.mecamex.net/revistas/LMEM
ISSN en trámite, 2014 Derechos de autor y derechos conexos, Asociación Mexicana de Mecatrónica A.C

3.3. Descripción de los atributos

De la información "cruda" descrita en la sección anterior se obtienen 3 atributos de movimiento y 4 de postura para describir a nuestros gestos. Los atributos de movimiento son Δx, Δy y Δz que describen cambios en la posición *XYZ* de la mano y nos permiten describir el desplazamiento de la mano en el espacio 3D reconstruido de la escena. Cada una de estas variables pueden tomar uno de 3 valores {(+),(-),(0)} que indican incremento, decremento ó no cambio dependiendo de la posición de la mano en la imagen inmediata anterior en la secuencia, respectivamente. Por ejemplo, si el usuario mueve su mano hacia arriba entonces Δy = (+); si la mano se mueve hacia abajo, Δy = (-) y si no hay movimiento vertical, Δy = (0).
Los atributos de postura son *forma*, *arriba*, *derecha* y *torso*. Éstos representan la apariencia de la mano y relaciones espaciales entre la mano y otras partes del cuerpo como el rostro y el pecho. La apariencia de la mano es representada por el atributo *forma* y puede tomar 3 valores: (+) si la mano muestra la palma, (-) si la mano está en posición horizontal ó (0) si la mano está inclinada a la derecha o a la izquierda. Esta información es fácilmente deducida por comparación de la longitud de los lados del rectángulo que segmenta a la mano. El atributo *derecha* indica si la mano está a la derecha de la cabeza; *arriba* indica si la mano está arriba de la cabeza y *torso* describe si la mano está sobre el torso del usuario. Las variables de postura son binarias *{V, F}* y toman su valor dependiendo de si la condición correspondiente se satisface o no. El número posible de observaciones diferentes es 648. Así, un gesto es una secuencia de vectores de estos 7 atributos en el siguiente orden: $<\Delta x, \Delta y, \Delta z,$ *forma, arriba, derecha, torso*$>$.

3.4. Modelos Ocultos de Markov para reconocimiento

Para el reconocimiento se utilizan los modelos ocultos de Markov (MOM). Un MOM λ describe la evolución de un proceso a través de un conjunto de estados identificados en cada tiempo *t* por una variable S_t y un conjunto de transiciones posibles entre ellos. Sin embargo, el estado S_t no es directamente identificable, es decir, está "oculto" y sólo es accesible a través de las observaciones O_t que se obtienen del sistema (en nuestro caso, los valores de atributos descritos en la sección anterior).

Un modelo oculto de Markov sigue el siguiente modelo de probabilidad:

$$P(O|\lambda)=\sum_{\forall S} P(S_1|\lambda)\prod_{t=1}^{T-1} P(S_{t+1}|S_t,\lambda)\prod_{t=1}^{T-1} P(O_t|S_t,\lambda) \tag{7}$$

donde $P(S_1|\lambda)$ es la distribución de probabilidad inicial de los estados, $P(O_t|S_t, \lambda)$ es la función de probabilidad de las observaciones dados los estados y $P(S_{t+1}|S_t, \lambda)$ es la distribución de probabilidad de las transiciones entre los estados a través del tiempo. La topología "hacia adelante" de 3 estados del modelo utilizado para describir cada gesto se muestra en la Fig. 4. La construcción de un clasificador de gestos requiere construir un MOM λ_i para cada gesto. Así, para el entrenamiento de cada uno de los modelos ocultos de Markov se utiliza una versión logarítmica del algoritmo Baum-Welch. En este trabajo se modificó la implementación de modelos ocultos de Markov de T. Kanugo [20] para múltiples observaciones. Para reconocimiento, un criterio de máxima-verosimilitud se utiliza para seleccionar el modelo λ_i que maximice $P(O \mid \lambda_i)$.

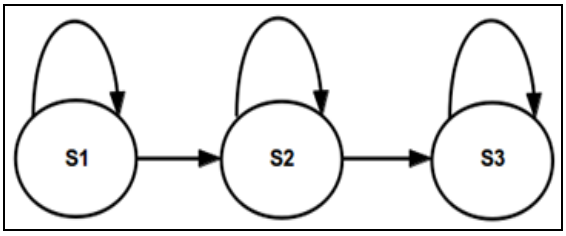

Fig. 4 Topología lineal de 3 estados o clases utilizada en los modelos ocultos de Markov

IV. EXPERIMENTOS Y RESULTADOS

4.1. Condiciones experimentales

Para la experimentación se seleccionaron aleatoriamente 200 muestras de cada tipo de gesto para entrenamiento de los modelos ocultos de Markov como se puede observar en la Tabla III. El resto de las muestras fueron utilizadas para reconocimiento. Este proceso fue realizado una única vez. Sin embargo, a diferencia de otras propuestas similares en las cuales son comunes pruebas de validación cruzada de K repeticiones con 90% de los datos disponibles para entrenamiento, en nuestra experimentación se utilizan más muestras de prueba. Los experimentos fueron ejecutados usando una laptop con un procesador Intel Core 7 y 8Gb de memoria RAM con SO Ubuntu 12.04.

La Mecatrónica en México, Vol. 3, No. 3, páginas 84 - 91, Septiembre 2014
Disponible en línea en www.mecamex.net/revistas/LMEM
ISSN en trámite, 2014 Derechos de autor y derechos conexos, Asociación Mexicana de Mecatrónica A.C

TABLA III. CLASIFICACIÓN DE LOS GESTOS, TOTAL DE MUESTRAS DE CADA GESTO Y CANTIDAD DE MUESTRAS SELECCIONADAS PARA ENTRENAMIENTO Y PARA PRUEBA.

	Número de muestras		
Gesto	**Datos de Entrenamiento**	**Datos de prueba**	**Total de Muestras**
Acercar	200	382	582
Parar	200	441	641
Derecha	200	454	654
Izquierda	200	405	605
Saludar	200	483	683

4.2. Resultados y discusión

Los resultados de reconocimiento se describen en la Tabla IV. El número total de muestras por gesto se muestra en la Tabla III. Los datos de prueba son resultado de la resta de las 200 muestras del total de cada tipo de gesto. Se muestra el total de resultados correctos y de error, así como el porcentaje correcto de reconocimiento obtenido de cada gesto en particular.

TABLA IV. RESULTADOS DE CLASIFICACIÓN DE GESTOS.

Gesto	**Datos de prueba**	**Correctos**	**Error**	**Reconocimiento (%)**
Acercar	382	377	5	98.69
Parar	441	423	18	95.91
Derecha	454	448	6	98.67
Izquierda	405	405	0	100.00
Saludar	483	475	8	98.34

El promedio ponderado de reconocimiento obtenido es de un 98.3%. Este resultado es muy similar al obtenido en el trabajo de Zhong que obtiene un porcentaje de 96.67% [21]. Los resultados de reconocimiento se pueden observar de forma más detallada en la matriz de confusión de la Tabla V. Como se puede observar el gesto de izquierda no tuvo ningún error de reconocimiento lo cual puede resultar peculiar debido a que en la ejecución del gesto la mano pasa por el recuadro del torso, lo cual se podría pensar que se confunda con el gesto de acercar que no es el caso. Por otra parte el gesto de parar tuvo más errores de reconocimiento con respecto a todos los demás, esto pudiera deberse a los movimientos de trayectorias similares a las de otros gestos. De esta manera, observando estos resultados, en general se puede decir que la clasificación de gestos no presenta una deficiencia considerable. Sin embargo esta investigación nos conduce a la necesidad de encontrar métodos alternativos de relajar el sistema para el reconocimiento de gestos para diferentes usuarios.

TABLA V. MATRIZ DE CONFUSIÓN.

Gesto	**Acercar**	**Parar**	**Derecha**	**Izquierda**	**Saludar**	**Total**
Acercar	377	2	0	3	0	382
Parar	3	423	4	2	9	441
Derecha	0	3	448	1	2	454
Izquierda	0	0	0	405	0	405
Saludar	0	3	5	0	475	483

V. CONCLUSIONES Y TRABAJO FUTURO

En este documento se detalló el procedimiento del desarrollo de un sistema visual de reconocimiento de gestos de un usuario utilizando diversas técnicas de visión por computadora con el apoyo del MS Kinect. Este sistema nos proporciona una base sólida hacia desarrollos e implementaciones más complejas de diversos algoritmos de reconocimiento visual por computadora en ambientes tridimensionales. Los experimentos arrojan resultados con porcentajes de efectividad total de reconocimiento de gestos de un 98.3%, lo cual se considera bueno respecto a otros sistemas similares que se encuentran en la literatura los cuales tienen una efectividad total de reconocimiento de gestos de un 96.67%.

Como trabajo futuro se propone también la realización de un modelo geométrico en 3D de la cabeza, brazo y torso del usuario basado en paralelepípedos, para mejorar la localización. También se propone el uso de filtros de partículas para el seguimiento de la mano, así como la integración de más gestos de reconocimiento y realización de experimentos del sistema con distintos usuarios y la implementación del sistema en un robot de servicio.

Referencias

[1] Rimkus, K., Bukis, A., Lipnickas, A., & Sinkevicius, S, *"3D human hand motion recognition system*" In *Human System. Interaction (HSI), 2013 The 6th International Conference on*, Págs 180-183, 2013.
[2] Sigal, L., Fleet, D. J., Troje, N. F., & Livne, M., *"Human attributes from 3d pose tracking*" *In Computer Vision–ECCV 2010*, Págs 243-257, 2010.
[3] Avilés–Arriaga H.H., Sucar–Succar L.E., Mendoza–Durán C.E., Pineda–Cortés L.A. ., *"A Comparison of Dynamic Naive Bayesian Classifiers and Hidden Markov Models for Gesture Recognition"*, J. Appl. Res. Technol. Vol. 9 No.1, México, Abr, 2011.
[4] Elmezain, M., Al-Hamadi, A., & Michaelis, B, *"Hand trajectory-based gesture spotting and recognition using HMM"*, In *Image Processing (ICIP), 2009 16th IEEE International Conference on, P*ágs 3577-3580, 2009.
[5] Camberos A., Tovar C., Medrano J., Arriaga L, Avilés H., *"Avances en el Desarrollo y Construcción del Robot de Servicio SerBot II"*, *Proceedings of the 12vo. Congreso Nacional de Mecatrónica. Por aparecer.*
[6] P. Trindade, J. Lobo, J. P. Barreto, Hand gesture recognition using color and depth images enhanced with hand angular pose data, Proc. of the 2012 IEEE International Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI), Págs. 71-76. 2012.
[7] Palacios J. M., Sagues C., Montijano E. and Llorente S., *"Human-Computer Interaction Based on Hand Gestures Using RGB-D Sensors*", Sensors, Volumen 13, Número 9 Págs. 11842-11860. 2013.
[8] Liang B, Zheng L., Shah S.A.A., Bennamoun M., Boussaid F., El-Sallam A.A., Urschler M., Bornik A., Donoser M., Babahajian P., *"Three dimensional motion trail model for gesture recognition"*, IEEE International Conference on Computer Vision (ICCV), Págs. 684–69. 2013.
[9] Tao C., Liu G., *"Human Robot Interaction using gesture recognition based on RGB-D Sensor"*, Journal of Convergence Information Technology (JCIT), Volumen 8, Número 11, Págs. 448-454, 2013.
[10] Mangera R., *"Static gesture recognition using features extracted from skeletal data"*, Twenty-Fourth Annual Symposium of the Pattern Recognition Association of South Africa, 2013.
[11] Bradski G., Kaehler A., Learning OpenCV: Computer Vision with the OpenCV Library, O'Reilly Media, 2008.
[12] ROS.org, Intrinsic calibration of the Kinect cameras. Disponible en:http://wiki.ros.org/openni_launch/Tutorials/IntrinsicCalibration. Última visita: 27 de julio del 2014.
[13] Magnenat S. Disponible en: https://groups.google.com/forum/#!topic/openkinect/AxNRhG_TPHg. Última visita: 14 de julio del 2014.
[14] N. Burrus, Kinect Calibration. Disponible en: http://nicolas.burrus.name/index.php/Main/HomePag. Últimavisita: 14 de julio del 2014.
[15] Trindade P, Lobo J, Barreto J. P., "Hand gesture recognition using color and depth images enhanced with hand angular pose data", Proc. of the 2012 IEEE International Conf. on Multisensor Fusion and Integration for Intelligent Systems (MFI), Págs. 71-76. 2012.
[16] OpenGL. Disponible en: http://www.opengl.org/. Última visita: 14 de julio del 2014.
[17] Yihsin H., Takao N., Toru Y., Eri, s-S, Norio T., *"A Hand Gesture Recognition System Based on GMM Method for Human-Robot Interface"*, *Robot, Vision and Signal Processing (RVSP), 2013 Second International Conference on* , Págs 291 - 294 . 2013.
[18] Nishimura, J. ; Kuroda, T. *"Versatile Recognition Using Haar-Like Feature and Cascaded Classifier"*, *Sensors Journal, IEEE*, Págs 942 - 951. 2010.
[19] Shrivastava R. *"A hidden Markov model based dynamic hand gesture recognition system using OpenCV"*, *Proceedings of the 2013 3rd IEEE International Advance Computing Conference*. IACC 2013, Págs. 947 – 950. 2013.
[20] Kanugo T., *"UMDHMM: Hidden Markov Model Toolkit, Extended Finite State Models of Language"*, A. Kornai (editor), Cambridge University Press. 1999. Disponible en: http://www.kanungo.com/software/software.html. Última visita: 27 de julio del 2014.
[21] Zhong Y., Yi L., Weidong C., Yang Z., *"Dynamic hand gesture recognition using hidden Markov models"*, *Proceedings of 2012 7th International Conference on Computer Science and Education*. ICCSE 2012, Págs 360 – 365. 2012.

Agradecimientos

Este trabajo está financiado parcialmente por el proyecto PROMEP /103.5/12/3620