

Reconstrucción 3D-2D de gestos usando información de vídeo monocular aplicada a un brazo robótico

3D-2D Gesture reconstruction using monocular video information applied to a robotic arm

*Sandra Nope**, *Humberto Loaiza*, *Eduardo Caicedo*

Grupo en Percepción y Sistemas Inteligentes, Universidad del Valle. Calle 13 N.º. 100-00, Santiago de Cali – Colombia

(Recibido el 16 de febrero de 2009. Aceptado el 15 de febrero de 2010)

Resumen

Se presenta un modelo que permite la reconstrucción de gestos realizados por un brazo humano, mediante la información extraída por un sistema de visión monocular, constituido por una cámara web de bajas prestaciones técnicas sin calibrar. El modelo se probó en un grupo de gestos realizados por diferentes personas y a diferentes distancias. El error cuadrático medio de las trayectorias de las articulaciones del brazo robótico medidas sobre la proyección 2D de la reconstrucción 3D, para los gestos escogidos, fue inferior a 0,32 píxeles.

----- *Palabras clave:* Visión Artificial, brazo robótico, cinemática directa, imitación, reconstrucción de gestos

Abstract

A model for gesture reconstruction performed by a human arm is presented. The model uses the information from a monocular vision system constituted by an uncalibrated basic webcam. A set of gestures were used to test the model, those gestures were performed by different people and distances from the camera. The mean square error of the robotic arm joints' trajectories was less than 0.317 pixels. This error was computed on the 2D projection of the 3D reconstruction for the selected gestures.

----- *Keywords:* Artificial vision, robotic arm, direct cinematic, imitation, gesture reconstruction

* Autor de correspondencia: teléfono: +57 + 2 + 339 19 80 ext 116, fax: + 57 + 2 + 339 23 61 ext 112, correo electrónico: sandrano@univalle.edu.co (S. Nope)

Introducción

El diseño de controladores para tareas robóticas es una actividad realizada usualmente por personas especializadas en la programación de robots; y, aún para ellos, a menudo se trata de un proceso complicado que requiere esencialmente ajustar a mano un controlador nuevo y diferente para cada tarea. En contraposición a este procedimiento, podría pensarse en que los robots interactúen de forma efectiva con los humanos; y de esta manera, incluso usuarios con pocos conocimientos podrían programar robots mediante esta interacción. Bajo esa premisa, surgió el aprendizaje por demostración o aprendizaje por imitación.

El aprendizaje por imitación, es una técnica poderosa en la adquisición de conocimiento usada por personas y animales que, aplicada a la robótica, permitiría la programación de robots con la habilidad de aprender comportamientos complejos e interactuar inteligentemente con el ambiente. Estas características de programación han sido un gran reto para los investigadores ya que los métodos convencionales de programación (textual, guiado y por reglas de aprendizaje) no son completamente adecuados y su investigación aún sigue abierta.

La meta a futuro es lograr que un robot aprenda por imitación un conjunto de gestos a través de demostraciones realizadas por “maestros” humanos, que, entre otras cosas, no comparten las mismas características antropomórficas. La imitación en este caso, debe entenderse como aquella que se presenta “cuando un comportamiento es observado, entendido y reproducido” [1] o imitación activa. Para ello, la primera fase consiste en obtener buenas aproximaciones a las ejecuciones realizadas por diferentes maestros humanos. Estas aproximaciones, pueden considerarse como una forma de imitación denominada imitación pasiva o mímica [2], es decir, aquella que consiste en replicar los gestos o movimientos del maestro o demostrador sin buscar entender esos gestos o la meta de sus acciones. Una discusión entre imitación activa y pasiva se puede encontrar en [3].

Así, el propósito de este trabajo es presentar un modelo que permita obtener los valores de las variables articulares de un brazo robótico con seis grados de libertad, de tal forma que imite la ejecución realizada por un demostrador humano de un grupo de cuatro gestos que son capturados a través de una cámara Web sin calibrar. Para alcanzar este objetivo, se requieren al menos dos etapas: la reconstrucción monocular y la estimación de los valores apropiados de las variables articulares del robot.

El desafío de la primera etapa es que debe hacer frente a la ambigüedad en cuanto a la profundidad, ya que pueden existir varias configuraciones tridimensionales que permitan que un punto se ubique en una determinada posición dentro de la imagen, y es un proceso que ha sido abordado por diferentes autores como [4-12]. Así mismo, involucra a su vez dos fases: el seguimiento de características en las imágenes (en este caso las articulaciones) y la reconstrucción tridimensional.

En cuanto a la estimación de los valores de las variables articulares que dan lugar a la imitación, se presentan las expresiones analíticas para un brazo robótico constituido por tres articulaciones, cada una con dos grados rotacionales de libertad. Sin embargo, el procedimiento presentado puede ajustarse para obtener las expresiones de un brazo robótico con una configuración diferente.

En la figura 1 se presenta el diagrama de bloques del sistema propuesto:

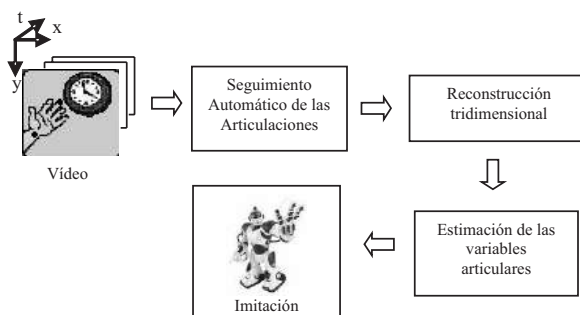


Figura 1 Diagrama de bloques del sistema

Lo interesante del método propuesto respecto a los trabajos previos, es que no sólo se logra hacer

seguimiento automático de los movimientos humanos sino que adicionalmente, el sistema hace una adaptación a sus características antropomórficas propias, de acuerdo con sus grados de libertad. En los trabajos citados, se emplean algoritmos de mayor complejidad algorítmica que seguramente redundan en mayores tiempos de cómputo; mientras que aquí se obtiene una relación directa de coordenadas de imagen a variables articulares del robot, cuyos resultados muestran la versatilidad del método, siendo capaz de seguir casi cualquier movimiento, y cuya limitación está relacionada con el uso de una cámara monocular y la ambigüedad que resulta en cuanto a la profundidad de las articulaciones del brazo. Lo anterior se resolvió haciendo algunas asunciones que limitan el espacio de trabajo del robot y los gestos que pueden ser imitados (ej. Que la profundidad del dedo medio es menor que la de la muñeca; la de la muñeca es menor que la del codo; la del codo es menor que la del hombro).

En este trabajo se explica el método empleado para el seguimiento automático de las articulaciones del brazo del demostrador, se presentan las transformaciones que deben realizarse para convertir las coordenadas sobre las imágenes de las articulaciones a coordenadas tridimensionales, las cuales son posteriormente usadas para estimar los valores de las variables articulares del brazo robótico. Para probar las expresiones analíticas encontradas se analizan y discuten los resultados de la reconstrucción 3D de cinco gestos y al final se presentan las conclusiones y trabajo futuro.

Metodología

Seguimiento automático de las articulaciones

Para hacer el seguimiento de las articulaciones de forma automática, se recurrió a marcadores negros pegados a cada articulación del demostrador (hombro, codo, muñeca y punta del dedo medio). Es de aclarar que la ubicación de cada marcador no requiere alta precisión por dos razones princi-

pales: primero porque como se indicó anteriormente, el sistema debe ajustarse al hecho de no poseer las mismas características antropomórficas que el demostrador; y segundo, por tratarse de articulaciones rígidas.

El sistema requiere que, al inicio, se indique manualmente la ubicación de cada marcador sobre la primera trama del vídeo. Esta información se usa para definir la posición de cada articulación sobre la imagen, posición que corresponde al centroide de la región concatenada de píxeles de color negro, alrededor de la ubicación inicial dada por el usuario.

Para segmentar los marcadores de color negro, se utilizaron dos umbrales en el plano V, en el espacio de color HSV, que fueron heurísticamente determinados; así, se consideran puntos de color negro aquellos cuyo valor en el plano V está en el rango de 0 a 0,25.

Para el seguimiento automático de los marcadores en el vídeo, se realiza una exploración dentro de una ventana cuadrada de 10×10 píxeles², centrada en la posición de la juntura de la trama precedente, de tal forma que optimice su localización en tramas subsecuentes, disminuyendo así el espacio de búsqueda.

A los píxeles dentro de las ventanas de búsqueda, que fueron identificados como puntos de color negro, se les aplica una operación morfológica de mayoría, de tal manera que se elija el grupo concatenado de mayor cantidad de píxeles (en caso de que haya más de un grupo), y, al igual que en el caso anterior, se asigna el centroide de dicho grupo como posición de la juntura sobre la imagen.

Para obtener buenas estimaciones de las posiciones de las articulaciones sobre las imágenes, debe garantizarse que los marcadores no se traslapen con objetos del fondo que compartan el mismo color. También, es fundamental que los marcadores permanezcan visibles durante toda la ejecución del movimiento y que no se traslapen o pasen demasiado cerca entre ellos. Si esto ocurre, el algoritmo será incapaz de seguirlas apropiadamente.

Con este algoritmo de seguimiento es posible obtener las coordenadas de cada una de las articulaciones en el plano de la imagen durante la ejecución de un gesto. Sin embargo, y teniendo en cuenta que en este caso se quiere recuperar las variables articulares que debe tener un brazo robótico para aproximar la pose del brazo del demostrador, es necesario rotar cada una de las imágenes de los vídeos de la ejecución del demostrador (*allo-imagen*), para que la imagen que vea el robot corresponda a un gesto realizado por él (*ego-imagen*).

El sistema de coordenadas en una imagen, usualmente corresponde al que aparece sobre la imagen de la persona en el lado derecho de la figura 2, el cual, por facilidad, se traslada al hombro del brazo del imitador (robot) como se muestra en el lado izquierdo de la figura 2. Este cambio en el sistema de referencia se expresa matemáticamente a través de las ecuaciones en (1):

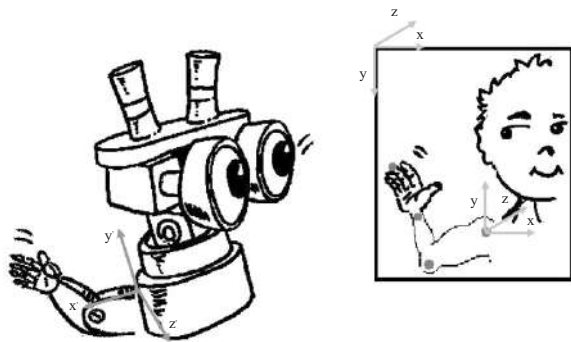


Figura 2 Sistema de coordenadas para la recuperación de las coordenadas tridimensionales desde las imágenes de vídeo

$$\begin{aligned}
 u_d' &= u_d - u_h & v_d' &= a - v_d - v_h \\
 u_m' &= u_m - u_h & v_m' &= a - v_m - v_h \\
 u_c' &= u_c - u_h & v_c' &= a - v_c - v_h \\
 u_h' &= 0 & v_h' &= 0
 \end{aligned} \tag{1}$$

donde (u_d', v_d') son las coordenadas de imagen del dedo en el nuevo sistema de coordenadas, y (u_d, v_d) sobre el antiguo. De la misma forma, (u_m', v_m') (u_m, v_m) son las coordenadas de la muñeca,

(u_c', v_c') (u_c, v_c) son las coordenadas del codo, y (u_h', v_h') (u_h, v_h) son las coordenadas del hombro en los sistemas coordenados correspondientes. La constante a corresponde a la altura de la imagen y depende de la resolución de la cámara.

Hasta aquí, el sistema aún no ha logrado obtener la ego-imagen, puesto que si el nuevo sistema de referencia se ubica sobre el brazo robótico, tendría las direcciones que se muestran a la izquierda de la figura 2 sobre el robot, en donde se puede apreciar que se requiere invertir los ejes x y z para obtener la ego-imagen. El cambio correspondiente se realiza durante el cálculo de las coordenadas tridimensionales como se explicará posteriormente.

Con esta información y, basados en el trabajo de Taylor [11], es posible recuperar las coordenadas tridimensionales de cada marcador, tal como se explica en el siguiente apartado.

Reconstrucción tridimensional

En su trabajo, Taylor [11] presenta una forma de recuperar las coordenadas tridimensionales, desde cualquier tipo de imagen, realizando las siguientes asunciones:

Se provee la correspondencia entre las articulaciones en el modelo del brazo y los marcadores en la imagen. En este caso, y como se explicó en el apartado anterior, esta relación se indica a través de la primera imagen y el algoritmo sigue las marcas automáticamente.

La relación entre los marcadores en la escena y sus correspondientes en la imagen, pueden modelarse como una proyección ortográfica escalada. Bajo estas condiciones, las coordenadas de un punto en la escena (X, Y, Z) pueden relacionarse con las coordenadas de su proyección en la imagen, (u, v) a través de la ecuación (2), en donde el parámetro s corresponde a un factor de escala desconocido.

$$\begin{pmatrix} u \\ v \end{pmatrix} = s \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{2}$$

Se conocen *a priori* las longitudes relativas de los segmentos en el modelo del brazo. Así, si se conoce por ejemplo la longitud l de un segmento de línea en una imagen bajo proyección ortográfica, la proyección de los dos puntos finales en el espacio (X_1, Y_1, Z_1) y (X_2, Y_2, Z_2) sobre las de las imágenes (u_1, v_1) y (u_2, v_2) , respectivamente, dan como resultado las ecuaciones en (3):

$$\begin{aligned} l^2 &= (X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2 \\ u_1 - u_2 &= s(X_1 - X_2) \\ v_1 - v_2 &= s(Y_1 - Y_2) \end{aligned} \quad (3)$$

$$dZ = (Z_1 - Z_2) = \sqrt{\frac{l^2 - ((u_1 - u_2)^2 + (v_1 - v_2)^2)}{s^2}}$$

En otras palabras, este análisis permite calcular la configuración de los puntos en la escena como función del parámetro de escala s . Para un valor determinado de escala s , aún son posibles dos soluciones correspondientes al signo de dZ , lo que refleja el hecho de que cualquiera de los puntos puede ser escogido como el de menor profundidad. Todo esto lleva a que deba conocerse, además, cuál es el punto real con menor profundidad, o sea, el más cercano a la cámara.

Esta ambigüedad puede resolverse limitando el espacio de trabajo del robot; por ejemplo, en los gestos ejecutados en los videos se puede verificar que el punto más cercano a la cámara corresponde al del dedo medio.

Así mismo, para definir un valor apropiado de s puede usarse la inecuación (4), que surge del hecho de que dZ no puede ser un número complejo:

$$s \geq \frac{\sqrt{((u_1 - u_2)^2 + (v_1 - v_2)^2)}}{l} \quad (4)$$

Otra opción para estimar el valor de s , es colocar un segmento perpendicular al plano de la imagen y de longitud conocida (L), y medir su longitud en píxeles (l); de esta manera, el valor de s quedaría determinado por (5):

$$s = \frac{l}{L} \quad (5)$$

Al aplicar las ideas de Taylor para el caso del brazo robótico, se obtiene como resultado las ecuaciones en (6), en donde además, se ha invertido el signo de las coordenadas X y Z , que corresponde a la última transformación necesaria para convertir la *allo-image* en la *ego-image*, y cuyo efecto es reflejar geoméricamente los puntos usando como eje de simetría al eje y .

$$\begin{aligned} X_c &= -\frac{u_c'}{s} \\ Y_c &= \frac{v_c'}{s} \\ Z_c &= \sqrt{L_{brazo}^2 - (X_c^2 + Z_c^2)} \\ X_m &= -\frac{(u_m' - u_c')}{s} + X_c \\ Y_m &= \frac{(v_m' - v_c')}{s} + Y_c \\ Z_m &= \sqrt{L_{antebrazo}^2 - \left(\frac{(u_m' - u_c')^2 + (v_m' - v_c')^2}{s^2} \right)} + Z_c \\ X_d &= -\frac{(u_d' - u_m')}{s} + X_m \\ Y_d &= \frac{(v_d' - v_m')}{s} + Y_m \\ Z_d &= \sqrt{L_{mano}^2 - \left(\frac{(u_d' - u_m')^2 + (v_d' - v_m')^2}{s^2} \right)} + Z_m \end{aligned} \quad (6)$$

En donde, (X_c, Y_c, Z_c) , (X_m, Y_m, Z_m) y (X_d, Y_d, Z_d) son las coordenadas tridimensionales estimadas del codo, la muñeca y dedo medio, respectivamente. L_{brazo} , $L_{antebrazo}$ y L_{mano} son las longitudes de los enlaces del brazo robótico que se consideran conocidas.

Una vez que se cuenta con las coordenadas tridimensionales de las articulaciones, se pueden determinar los valores de las variables articulares del brazo robótico, que le permitiría reproducir el movimiento del brazo del demostrador.

Estimación de las variables articulares

Para determinar la cinemática directa del robot, se presentan, a través de la figura 3, los sistemas de referencia en cada una de las articulaciones.

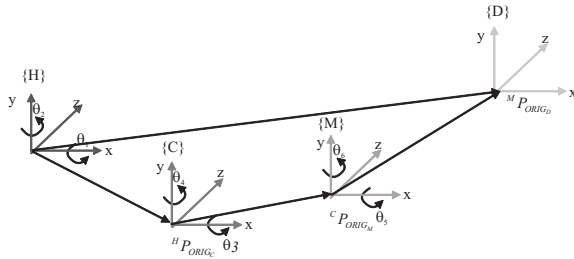


Figura 3 Sistemas de referencia para el brazo robótico

En la figura 3, el sistema de referencia $\{H\}$ corresponde al del hombro (que es tomado también como sistema de referencia global), $\{C\}$, $\{M\}$ y $\{D\}$ son los sistemas de referencia del codo, muñeca y dedo medio, respectivamente. ${}^H P_{ORIG_H}$ y ${}^H P_{ORIG_C}$ corresponden a la ubicación respecto al sistema de referencia $\{H\}$ del hombro y del codo, respectivamente; ${}^C P_{ORIG_M}$ es la ubicación de la muñeca respecto al sistema de referencia $\{C\}$, y ${}^M P_{ORIG_D}$ es la ubicación del dedo medio respecto a la muñeca.

Por otro lado, el brazo robótico cuenta con dos grados de libertad rotacionales en cada articulación, uno sobre el eje x' y otro sobre el eje y' , que son los que permiten el movimiento de flexo-extensión y de aducción-abducción, respectivamente. De acuerdo con esto, puede verificarse que el modelo directo del robot está definido por las ecuaciones (7) y (8).

$${}^C P_{ORIG_M} = Rxy^{-1}(\theta_2, \theta_1)({}^H P_{ORIG_M} - {}^H P_{ORIG_C}) \quad (7)$$

$${}^M P_{ORIG_D} = Rxy^{-1}(\theta_4, \theta_3)Rxy^{-1}(\theta_2, \theta_1)({}^H P_{ORIG_D} - {}^H P_{ORIG_M}) \quad (8)$$

En donde, $Rxy^{-1}(q_p, q_2)$ es la matriz de rotación inversa, y (q_3, q_4) son las rotaciones sufridas por el codo del demostrador y (q_5, q_6) las de la muñeca.

Por lo tanto, los valores de las variables articulares del brazo robótico quedan determinados si se conocen los valores de estos ángulos, los cuales se calculan consecutivamente tal como se explica a continuación.

Puede verificarse que un movimiento de aducción-abducción en el hombro del demostrador (θ_1) y de un movimiento de flexo-extensión (θ_2), puede estimarse de las coordenadas tridimensionales de acuerdo con las ecuaciones en (9), y que estos valores afectan los sistemas los demás sistemas de referencia como se observa en las ecuaciones (7) y (8).

$$\theta_2 = \sin^{-1}\left(\frac{Y_c}{L_{brazo}}\right) \quad (9)$$

$$\theta_1 = \tan^{-1}\left(\frac{X_c}{Z_c}\right)$$

Una vez estimados los valores articulares del hombro (θ_1, θ_2), se puede calcular ${}^C P_{ORIG_M}$ haciendo uso de la ecuación (7). Estas coordenadas corresponden a $[X'_m, Y'_m, Z'_m]$, y son empleadas en las ecuaciones en (10) para estimar los valores de los ángulos del codo (θ_3, θ_4).

$$\theta_4 = \sin^{-1}\left(\frac{Y'_m}{L_{antebrazo}}\right) \quad (10)$$

$$\theta_3 = \tan^{-1}\left(\frac{X'_m}{Z'_m}\right)$$

Del mismo modo, una vez estimados los valores de las variables articulares del hombro (θ_1, θ_2) y del codo (θ_3, θ_4), se halla ${}^M P_{ORIG_D}$ de acuerdo con la ecuación (8). El resultado es las coordenadas tridimensionales $[X'_d, Y'_d, Z'_d]$, que son empleadas en las ecuaciones en (11) para estimar los valores de los ángulos de la muñeca (θ_5, θ_6).

$$\theta_6 = \sin^{-1}\left(\frac{Y'_d}{L_{mano}}\right) \quad (11)$$

$$\theta_5 = \tan^{-1}\left(\frac{X'_d}{Z'_d}\right)$$

Así, se obtienen todos los valores articulares necesarios para que un brazo robótico con seis

grados de libertad, replique la pose del brazo de un demostrador que ha sido grabado a través de una cámara web sin calibrar. Así mismo, que el procedimiento descrito, al ser secuencial, implica que si a través de los grados de libertad de una articulación no se puede alcanzar una posición determinada, en las siguientes estimaciones se pueda “compensar”, de tal forma que se alcance la posición deseada por el efector final.

Experimentación

El procedimiento descrito anteriormente fue empleado para reconstruir cuatro gestos realizados por diferentes personas, en diferentes posiciones y distancias respecto a la cámara web monocular sin calibrar. El gesto 1 corresponde a un movimiento de aducción-abducción de muñeca (saludar); el gesto 2 corresponde un movimiento de flexo-extensión de muñeca (abanicar); el gesto 3 corresponde a un movimiento de rotación de muñeca en sentido inverso a las manecillas del reloj; y, el gesto 4 corresponde a acercar y alejar la mano abierta respecto a la cámara.

Durante la grabación de los vídeos las condiciones de iluminación y de velocidad de los gestos fueron semi-controladas. En el primer caso con el objeto de garantizar el seguimiento automático de las marcas, y en el segundo, evitar que dada la velocidad de captura de la cámara, los objetos en movimiento aparezcan borrosos en las imágenes. Sin embargo, condiciones como la distancia del demostrador a la cámara y su posición frente a la misma, no fueron controladas.

El semi-control en las condiciones de iluminación se logro mediante el uso de luz artificial dentro del espacio de oficina usado para la captura de vídeo, mientras que en el caso de la luz solar sólo se tuvo en cuenta no realizar grabaciones en la noche en donde las condiciones de iluminación son pobres. En el caso de la velocidad de los gestos, a los demostradores se les presentó un ejemplo de ejecución apropiada de los diferentes gestos. Por otro lado, solo cuando la velocidad era excedida, se les solicitaba una nueva ejecución “un poco más lenta”.

Note que el hecho de utilizar vídeos implica la repetición de la estimación de las variables articulares de cada articulación en cada una de las imágenes que lo componen, así mismo, es más desafiante que usar imágenes independientes pues la ejecución del gesto debe mantener suavidad en la ejecución de tal forma que las simulaciones por parte del robot se vean naturales.

Para verificar la exactitud del método se calcula el error cuadrático medio de las trayectorias sobre cada una de las imágenes de las articulaciones del demostrador y del brazo robótico, en los diferentes vídeos. Para ello, se proyectan los valores de las variables articulares del brazo al plano de la imagen, de tal forma que sean comparables con las posiciones de las articulaciones obtenidas durante el seguimiento. No se realiza una estimación del error en coordenadas tridimensionales, debido a que no se cuenta con la información real de éstas durante las ejecuciones realizadas por los demostradores.

En las pruebas se empleó un valor de $s=10$ (valor que cumple con la restricción establecida en la inequación (4) para cada articulación, en cada instante de tiempo, y en cada una de las secuencias de video en las que se realizaban los gestos.

Los valores usados para L_{upper_arm} , $L_{forearm}$ y L_{hand} fueron de 24,5, 24 y 11,5 cm., respectivamente, y corresponden a las longitudes de los enlaces del brazo robótico.

Resultados

Las figuras 4 a 7 presentan gráficamente algunos resultados obtenidos, uno por gesto. En la parte superior de cada una de estas figuras aparece resaltada la trayectoria estimada en píxeles de imagen, mientras que en la inferior aparece la proyección en el plano de la imagen de las coordenadas tridimensionales.

La figura 8 presenta un ejemplo adicional que corresponde a un movimiento aleatorio, el cual permite visualizar la versatilidad del método.



Figura 4 Ejemplo de reconstrucción del gesto 1



Figura 5 Ejemplo de reconstrucción del gesto 2

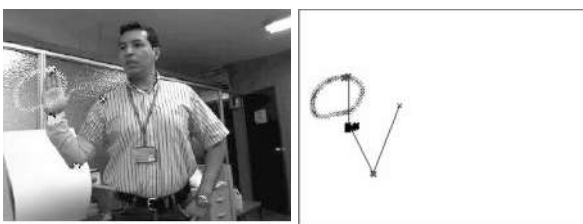


Figura 6 Ejemplo de reconstrucción del gesto 3



Figura 7 Ejemplo de reconstrucción del gesto 4



Figura 8 Ejemplo de reconstrucción de un gesto aleatorio

La tabla 1 presenta el promedio del error cuadrático medio entre las coordenadas de las trayectorias de las articulaciones (codo, muñeca y dedo medio) del demostrador en el plano de la imagen, y la producida por los valores angulares estimados con los proyectados en el plano de la imagen.

Tabla 1 Promedio del error cuadrático medio entre las coordenadas de las trayectorias para los gestos presentados

	<i>Error Cuadrático Medio</i>	
	<i>x</i>	<i>y</i>
Gesto 1	0,0364	0,0234
Gesto 2	0,1615	0,1082
Gesto 3	0,0192	0,0215
Gesto 4	0,1465	0,1149
Aleatorio	0,2746	0,3170

En la tabla 1 puede observarse que hay una relación directa entre el error en las trayectorias de las articulaciones y la traslación del hombro observada durante la ejecución de los gestos. Por ello, el menor error ocurre durante la ejecución del gesto 1 y 3. Mientras que los mayores errores ocurren en el gesto 4 y el gesto con un movimiento aleatorio. Sin embargo, en todos los casos el error se considera pequeño para muchas aplicaciones prácticas, más aun cuando son estimadas a partir de una imagen monocular tomada desde una cámara web sin calibrar.

Discusión

Es importante resaltar que, aunque en todos los casos no se cumple en forma estricta la segunda asunción de Taylor, relacionada con la proyección ortográfica entre las marcas de la escena y sus correspondencias en las imágenes, el robot obtiene buenas aproximaciones a las trayectorias realizadas por el demostrador, en especial para la mano.

Aunque los grados de libertad definidos para el robot no permiten que su hombro realice traslaciones, el seguimiento de esta articulación indica que sí presenta movimientos de traslación. Estos movimientos se compensan durante las estimaciones de las variables articulares del codo, muñeca y dedo medio, tal como puede apreciarse en las imágenes inferiores de las figuras 4 a 8 en donde el hombro aparece sin movimiento.

Algo similar ocurre durante la ejecución por parte del demostrador del gesto 3 (figura 6), en donde al final del mismo, usualmente ocurre una rotación sobre el eje x (al rededor del eje longitudinal del antebrazo), que corresponde a un grado de libertad adicional con el que no cuenta el robot. Este grado de libertad también es compensado por las articulaciones del robot para alcanzar las posiciones en la imagen alcanzadas por el demostrador.

Conclusiones y trabajo futuro

Se presentó un modelo que permite la reconstrucción tridimensional de gestos realizados por un brazo, mediante la información extraída por una sola cámara web monocular sin calibrar y en condiciones de iluminación semi-controladas. El procedimiento descrito es simple y de bajo costo computacional.

El error cuadrático medio de las trayectorias de las articulaciones para los gestos escogidos durante las pruebas, osciló entre 0,0215 y 0,317 píxeles, lo que indica que puede ser usado en múltiples aplicaciones de seguimiento (*tracking*) y aprendizaje por demostración, entre otras. Se verificó que el procedimiento es sensible a los movimientos de traslación del hombro del demostrador, debido a que el brazo del robot no cuenta con este grado de libertad. Sin embargo, el algoritmo es robusto y realiza ajustes de tal forma que se alcance la posición deseada en el efector final, ajustándose a sus grados de libertad. Esta limitación podría eliminarse dotando al robot con los grados de libertad necesarios para seguir cualquier clase de movimiento.

Los trabajos futuros se enfocarán a relacionar la información de movimiento visual extraída de las secuencias de vídeo [13], con los valores de las variables articulares estimadas mediante el método presentado aquí y, de esta manera, lograr el aprendizaje por demostración de un grupo de gestos por parte de un brazo robótico real.

Agradecimientos

Agradecemos al Programa de apoyo a doctorados de Colciencias, a la Universidad del Valle y al Instituto Técnico Superior (IST) – de Portugal, por el soporte a este trabajo. Un reconocimiento especial al profesor José Santos-Victor del IST – Portugal por su orientación, consejo y apoyo a este proyecto. Agradecemos también a la Institución Universitaria Tecnológica de Comfacaucá – ITC sede Popayán, por brindar el tiempo y el espacio para que S. N. pudiera perfeccionar el presente artículo

Referencias

1. E. L. Thorndike. *Animal Intelligence*. Ed. Macmillan. New York. 1911. Citado en: B. G. Galef *Imitation in animals: History, definition and interpretation of data from psychological laboratory*. T. Zentall & B. G. Galef (editores). Ed. Hillsdale. New Jersey. 1988. pp. 3-28.
2. H. Barlow. "Possible principles underlying the transformation of sensory messages". Rosenblith (editor). *Sensory Communication*. Ed. MIT Press. Cambridge. MA (USA). 1961. pp. 217-234.
3. Y. Derimis, G. Hayes. "Imitations as a dual-route process featuring predictive and learning components: a biologically plausible computational model". *Imitation in animals and artifacts*. Ed. MIT Press. Cambridge. MA (USA). 2002. pp. 327-361.
4. H. Sidenbladh, M. J. Black, L. Sigal. "Implicit probabilistic models of human motion for synthesis and tracking". *European Conference on Computer Vision*. Copenhagen. Dinamarca. 2002. pp. 784-800.
5. N. Howe, M. E. Leventon, W. T. Freeman. "Bayesian reconstruction of 3d human motion from single-camera video". En S. Solla, T. Leen. K. R. Müller (editores.). *Advances in neural information processing systems*. Cambridge. MA: MIT Press. 1999. pp. 820-826.

6. R. Bowden. "Learning statistical models of human motion". *IEEE Workshop on Human Modeling, Analysis and Synthesis*. Hilton Head. Carolina del Sur. USA. 2000. pp. 10-17.
7. C. Sminchisescu, B. Triggs. "Kinematic jump processes for monocular 3D human tracking". *Proceedings of the International Conference on Computer Vision and Pattern Recognition*. Madison, (WI). 2003. Vol. 1. pp. 69-76.
8. M. J. Park, M. G. Choi, S. Y. Shin. "Human motion reconstruction from inter-frame feature correspondences of a single video stream using a motion library". *2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*. San Antonio. Texas (USA). 2002. pp. 113-120.
9. C. Bregler, J. Malik. "Tracking people with twists and exponential maps". *IEEE International Conference on Computer Vision and Pattern Recognition*. Anchorage. (Alaska). 1998. pp. 8-15.
10. E. D. Bernardo, L. Goncalves, P. Perona. "Monocular tracking of the human arm in 3d: realtime implementation and experiments". *IEEE International Conference on Pattern Recognition*. Viena. Austria. 1996. pp. 622-626.
11. C. J. Taylor. "Reconstruction of articulated objects from point correspondences in a single uncalibrated image". *Computer Vision and Image Understanding*. Hilton Head. Carolina del Sur. USA. Vol. 1. 2000. pp. 349-363.
12. M. Cabido, J. Santos Victor. "Visual Transformations in Gesture Imitation: what you see is what you do". *2003 IEEE International Conference on Robotics & Automation*. Taipei. (Taiwan). 2003. pp. 2375-2381.
13. S. Nope, E. Caicedo, H. Loaiza. "Aplicaciones del Movimiento y su Representación Biológica en el Reconocimiento de Gestos". *Ingeniería y Competitividad*. Vol. 8. 2006. pp. 55-63.