

## INTERFAZ BASADA EN VISION PARA COMANDAR UN ROBOT MOVIL

Elisa Perez, Carlos Soria, Oscar Nasisi, Vicente Mut

*Instituto de Automática*  
*Universidad Nacional de San Juan*  
*Av. Libertador San Martín 1109 (o), 5400 San Juan*  
[eperez@gateme.unsj.edu.ar](mailto:eperez@gateme.unsj.edu.ar); {csoria;onasisi}@inaut.unsj.edu.ar

**Resumen:** En este trabajo se presenta el desarrollo de una interfaz basada en visión para realizar la navegación de un robot móvil por medio de un pasillo. A través de la captura de la imagen de una mano, se obtendrán las características de imágenes necesarias para obtener la referencia angular que será utilizada por el robot móvil. Los resultados experimentales se mostrarán usando primero un robot móvil en un entorno virtual y luego se utilizará un robot móvil real para hacerlo navegar en un pasillo real.

**Palabras Claves:** visión por computador, procesamiento de imágenes, robot móvil.

### 1. INTRODUCCION

Las interfaces basadas en visión (VBI, visual-based interfaces, en inglés) son sistemas perceptivos que utilizan la visión como canal de comunicación. Estos tipos de interfaces hacen que la interacción del usuario con las computadoras se realice de manera más intuitiva.

Las VBI buscan adquirir información acerca del usuario automáticamente, para interpretar los comandos explícitos o implícitos relacionados a los movimientos o gestos, de una parte o de todo el cuerpo de las personas. Entre los usos que pueden tener las VBI en aplicaciones con computadoras se pueden citar los siguientes: a) seguimiento de la cabeza, b) reconocimiento de expresiones faciales, c) seguimiento de los ojos, d) reconocimiento de los gestos de la mano/brazo, entre otras aplicaciones.

La VBI que realizan reconocimiento de gestos de la mano es la interfaz que representa la manera evolutiva más natural de las interfaces usuario-gráficas (GUI, en inglés), con las cuales nosotros interactuamos por medio de las manos a través del mouse, teclado y otros dispositivos. La idea es que la interfaz de reconocimiento de gestos pueda ayudar o reemplazar a los dispositivos mencionados para desempeñar de una manera más familiar tareas con las computadoras. En la Interacción Humano-Computadora (HCI, siglas en inglés) el reconocimiento de gestos puede ayudar a realizar

varias clases de tareas, dentro de las cuales podemos mencionar: a) Dibujo y posicionamiento, reemplazando el mouse; b) Navegación en ambientes virtuales y manipulación de objetos virtuales; c) Control de diferentes dispositivos electrónicos y robots; d) Comunicación a través de señas. En (Iannizzotto *et al.*, 2001) han desarrollado una interfaz en la cual reemplazan el mouse por la mano, con los dedos pulgar e índice realizan los clicks del mouse y deslizan el curso en la pantalla. El uso de la mano como herramienta de dibujo también a sido experimentado en un Panel de Interfaz Visual por (Zhang *et al.*, 2001), el sistema permite al usuario controlar el programa MsWindows Paint moviendo simplemente el dedo en un panel de forma cuadrada, y lo que el dedo realiza en el panel es reproducido en la pantalla de la PC, entonces se puede escribir con los dedos, dibujar etc. En (Moore *et al.*, 1998) usaron un sistema con una cama en donde el usuario desde allí puede controlar un conjunto de TV y lámparas. Estas interfaces son muy útiles para mejorar la calidad de vida en personas que posean alguna discapacidad, ya que les facilitan la manera de interactuar con el medio (Cook y Hussey, 2002). En este trabajo se desarrolla una interfaz basada en visión que reconoce los gestos y posición de la mano para comandar un robot móvil en la tarea de navegación a través de pasillos.

Este trabajo está organizado de la siguiente manera: en la sección 2 se explican las técnicas de procesamiento de imágenes usadas en esta interfaz.

La sección 3 modelado matemático del robot utilizado. El diseño del controlador se muestra en la sección 4. La descripción del entorno virtual utilizado se realiza en la sección 5. En la sección 6 se presentan los resultados experimentales y en la sección 7 se dan las conclusiones obtenidas.

## 2. PROCESAMIENTO DE IMAGENES

La interfaz desarrollada usa la visión artificial como canal de comunicación desde el usuario a la computadora u otro dispositivo.

Las cámaras de visión son herramientas de entrada no invasivas y actualmente sus costos son relativamente económicos. Por lo tanto la introducción de una VBIs en aplicaciones basadas en computadoras es un objetivo factible y no muy caro de implementar.

Para poder enviar consignas de ángulo de referencia a un robot móvil por medio de la interpretación de la posición de la mano y poder realizar una tarea de navegación, se necesita desarrollar un algoritmo de procesamiento de imágenes que detecte la presencia de la mano e identifique su orientación.

El primer paso que se desarrolló es la segmentación de la imagen y la ubicación de la mano en la imagen capturada por la cámara, esto se realizó trabajando en el espacio HSI (Hue-Saturation-Intensity, en inglés), ya que se sabe que el color de la piel se agrupa en regiones precisas en el espacio del color, (Kjeldsen, y Kender, 1996), (Zarit *et al.*, 1999).

Luego se buscan las características que representan correctamente la orientación de la mano, para comandar el robot. Las características de imagen elegidas son las coordenadas del centro de masa, el ángulo de orientación de la mano y el eje mayor y el eje menor de la mano.

Para calcular el centroide y los ejes se utilizaron los momentos de la imagen (Gonzalez y Wood, 2002), los cuales dan información global de la imagen. Los mismos involucran la suma de todos los píxeles de la imagen. Si  $I(x, y)$  es la intensidad de la imagen en la posición  $x, y$ , entonces los momentos de la imagen son:

$$\begin{aligned} M_{00} &= \sum_x \sum_y I(x, y) & M_{11} &= \sum_x \sum_y xyI(x, y) \\ M_{10} &= \sum_x \sum_y xI(x, y) & M_{01} &= \sum_x \sum_y yI(x, y) \\ M_{20} &= \sum_x \sum_y x^2 I(x, y) & M_{02} &= \sum_x \sum_y y^2 I(x, y) \end{aligned}$$

Las coordenadas del centro de masa se pueden obtener como:

$$x_c = \frac{M_{10}}{M_{00}} \quad y_c = \frac{M_{01}}{M_{00}} \quad (1)$$

Los ejes mayor y menor se obtuvieron aplicando las siguientes ecuaciones (Horn, 1986), para esto es necesario definir las siguientes variables intermedias:

$$a = \frac{M_{20}}{M_{00}} - x_c^2; \quad b = 2\left(\frac{M_{11}}{M_{00}} - x_c y_c\right); \quad c = \frac{M_{02}}{M_{00}} - y_c^2$$

Con las cuales se obtiene:

$$L_1 = \sqrt{6(a+c + \sqrt{b^2 + (a-c)^2})} \quad (2)$$

$$L_2 = \sqrt{6(a+c - \sqrt{b^2 + (a-c)^2})} \quad (3)$$

Para obtener la orientación de la mano se utiliza la herramienta Análisis de Componentes Principales, (PCA en inglés) (Gonzalez y Wood, 2002), la cual es una transformación lineal basada en técnicas estadísticas. A partir de la detección utilizando HSI, se obtiene la imagen binaria de la mano, la cual va a ser tratada como una población bidimensional, en otras palabras cada pixel de la mano es tratado como un vector bidimensional  $X = (x, y)^T$ , donde  $x$  e  $y$  son los valores de las coordenadas de ese punto con respecto a los ejes de referencia de la imagen. Se calcula el vector medio de estos vectores de la siguiente manera:

$$M_x = \frac{1}{k} \sum_{k=1}^k X_k \quad (4)$$

Donde  $k$  es el número de vectores que se obtiene de la imagen binarizada. Luego se procede a calcular la matriz de covarianza de la siguiente manera:

$$C_x = \frac{1}{k} \sum_{k=1}^k X_k X_k^T - M_x M_x^T \quad (5)$$

Luego de obtener la matriz  $C_x$  la cual es una matriz de dimensión 2x2, se calculan los autovectores de la misma, ya que con este método se establece un nuevo sistema de coordenadas cuyo origen queda en el centroide de la población, y cuyos ejes están en las direcciones de los autovectores de  $C_x$ , por lo tanto se puede establecer la rotación de la mano respecto a los ejes de la imagen. El ángulo obtenido es la referencia del ángulo de rotación del robot móvil. En la Fig. 2a se observa la imagen de la mano capturada por la cámara, en la Fig. 2b se muestra el resultado del procesamiento obteniéndose la detección de la mano con sus características, centroide, orientación y longitudes del eje mayor y menor,  $L_1$  y  $L_2$  respectivamente, graficados sobre la mano. En la Fig. 3 se muestran los ángulos medidos comparados con los ángulos reales. Para esta posición de la mano se tiene un error menor que  $|5.5|^\circ$ . En la Fig. 4b se muestra el resultado del procesamiento para la posición de la mano de costado, con sus características de imagen. En la Fig. 5 se muestran los ángulos medidos y reales. Para esta posición, mano de costado, se tiene un error menor que  $|3.6|^\circ$ . Los ángulos sensados para las dos posiciones de la mano tienen un rango de  $\pm 80^\circ$ .

La referencia de ángulos elegida se observa en la siguiente figura:

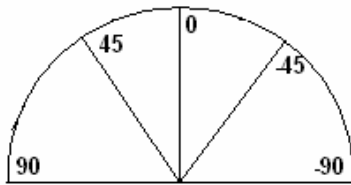


Fig 1. Referencia de Ángulos.

Para comandar el avance o parada del robot, se utilizará el valor del eje menor  $L_2$  (ec. 3). Si  $L_2 > L_{lim}$ , el robot debe parar, la mano tiene la posición que se muestra en la Fig. 1. Si  $L_2 < L_{lim}$ , el robot debe avanzar, esto significaba que la mano está de costado (Fig. 4).

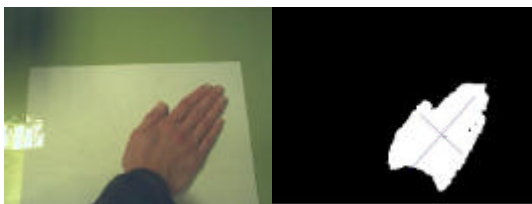


Fig.2. a) Imagen original. b) Imagen procesada, orientación y ejes mayor y menor detectados.

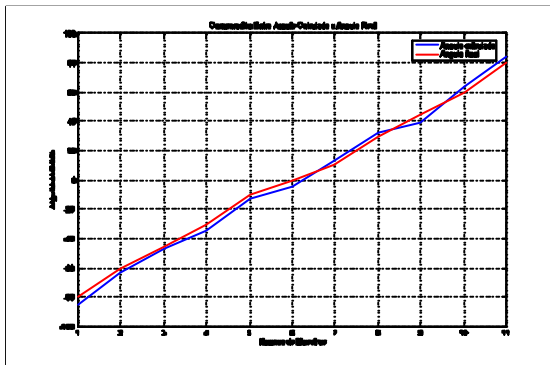


Fig. 3. Ángulos medidos para diferentes posiciones de la mano correspondientes a la Fig. 2.

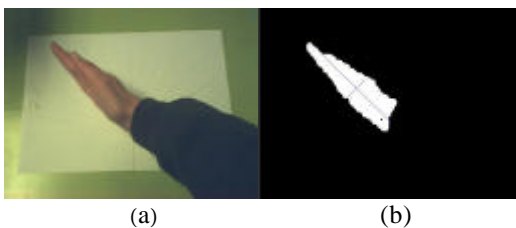


Fig. 4. a) Imagen original, mano de costado. b) Imagen procesada, orientación y ejes mayor y menor detectados.

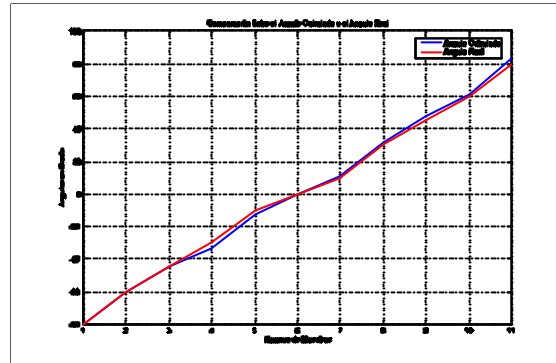


Fig 5. Ángulos medidos para diferentes posiciones de la mano correspondientes a la Fig. 4.

### 3. MODELO DEL ROBOT MOVIL

El robot móvil tipo uniciclo es un vehículo con dos ruedas motoras independientes en un solo eje y una rueda libre que permite mantener el equilibrio, como se muestra en la Fig. 6. El robot móvil se puede describir mediante el siguiente modelo de movimiento restringido en el plano

$$\begin{aligned} \dot{x} &= v \cos q \\ \dot{y} &= v \sin q \\ \dot{q} &= w \end{aligned} \quad (6)$$

donde  $x, y$  son las coordenadas cartesianas,  $q$  es el ángulo de orientación;  $v$  y  $w$  son las velocidades lineal y angular respectivamente. La restricción no-holonómica para el modelo (6) es

$$\dot{y} \cos q - \dot{x} \sin q = 0 \quad (7)$$

la cual especifica la trayectoria tangente a lo largo de cualquier trayectoria factible del robot. La suposición se hace en base a que el punto de referencia permanece en el centro de las dos ruedas motoras.  $v_1$  y  $v_2$  denotan las velocidades de las ruedas izquierda y derecha respectivamente. Las velocidades lineal  $v$  y angular  $w$  del robot móvil se pueden describir como  $v = (v_1 + v_2)/2$  y  $w = (v_1 - v_2)/L$ , donde  $L$  representa la distancia entre las dos ruedas motoras, (Fig. 7).



Fig. 6. El robot móvil Pioneer 2DX.

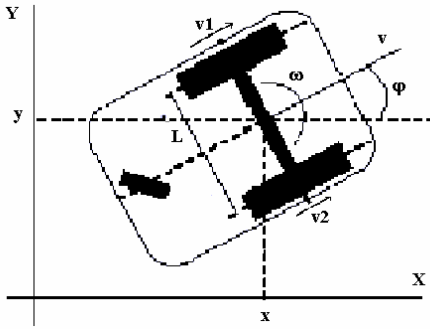


Fig.7. Descripción geométrica del robot móvil.

En el plano, el sistema unicycle tiene 3 grados de libertad, pero solo se puede controlar por medio de 2 acciones de control de entrada bajo la restricción no-holonómica. Cuando se considera el vector de estado  $(x, y, \theta)$ , se puede demostrar que el sistema (6) es controlable tanto en posición y en orientación (Canudas de Wit, 1997).

#### 4. DISEÑO DEL CONTROLADOR

La ley de control no lineal (comando de velocidad angular) aquí propuesta para cambiar la orientación del robot móvil para poder realizar la tarea de navegación está dada por:

$$\mathbf{w} = -k \tanh(\tilde{\mathbf{q}}) \quad (8)$$

donde  $k$  es una constante positiva,  $\mathbf{q}$  es la orientación del robot (cero grados para navegar, considerando el eje de movimiento del robot),  $\tilde{\mathbf{q}} = \mathbf{q}_d - \mathbf{q}$  es el error en la orientación del robot, y  $\mathbf{w} = \dot{\mathbf{q}}$  es el modelo cinemática del robot. La  $\tanh(\cdot)$  se utiliza para prevenir la saturación de los comandos ya que existe una cota en la velocidad angular del robot móvil.

##### 4.1. Análisis de Estabilidad

Reemplazando la ley de control expresada en (8) en la ecuación del sistema  $\dot{\mathbf{q}} = \mathbf{w}$ , se consigue la ley de control de lazo cerrado dada por:

$$\dot{\mathbf{q}} = -k \tanh(\tilde{\mathbf{q}}) \quad (9)$$

la cual es equivalente a

$$\dot{\tilde{\mathbf{q}}} = -k \tanh(\tilde{\mathbf{q}}) \quad (10)$$

Sea  $V = \frac{\tilde{\mathbf{q}}^2}{2}$  una función candidata de Lyapunov. Su derivada en el tiempo es.

$$\dot{V} = -\tilde{\mathbf{q}} k \tanh(\tilde{\mathbf{q}}) \quad (11)$$

la cual es una función definida negativa, por lo tanto se puede concluir que  $\mathbf{q}(t) \rightarrow \mathbf{q}_d(t)$  cuando  $t \rightarrow \infty$ .

Entonces el sistema de control de lazo cerrado es asintóticamente estable, por lo tanto está garantizado efectivamente que la orientación del robot cambiará acordeamente.

##### 4.2. Ley de Control de Velocidad Lineal

También se desea que el robot cambie su velocidad lineal  $v$  de acuerdo a si el robot está maniobrando para orientarse en la dirección correcta ( $\tilde{\mathbf{q}}$  no es nulo). De otra manera, se puede mover rápido, alcanzando la máxima velocidad lineal posible  $V_{max}$ . Entonces, la ley de control propuesta es

$$v = V_{max} \cos(\tilde{\mathbf{q}}) \quad (12)$$

$V_{max}$  debe ser definido teniendo en cuenta los límites físicos del robot.

#### 5. ENTORNO VIRTUAL

En la actualidad, los gráficos generados por computador se han convertido en grandes aliados de diferentes tipos de profesionales del diseño, debido a su gran utilidad en la manera de expresar claramente una idea. Dichos gráficos van desde simples trazos bidimensionales hasta complicadas estructuras o modelos tridimensionales desarrollados dentro de un ambiente virtual y así poder ser sometidos a diferentes procesos tales como: animación en 3D, imágenes interactivas en 360°, imágenes 3D estéreo e incluso VRML (Lenguaje de Modelado de Realidad Virtual).

Se puede definir la Realidad Virtual como un conjunto de herramientas de gráficos 3D, simulación e interacción que permiten al usuario operar directamente con un entorno generado por computador, interactivamente y en tiempo real, sin percepción de la existencia de un interfaz hombre-máquina. La realidad virtual puede ser de dos tipos: inmersiva y no inmersiva. Los métodos inmersivos de realidad virtual con frecuencia se ligan a un ambiente tridimensional creado por computadora el cual se manipula a través de cascos, guantes u otros dispositivos que capturan la posición y rotación de diferentes partes del cuerpo humano. La realidad virtual no inmersiva utiliza medios como el que actualmente nos ofrece Internet en el cual podemos interactuar a tiempo real con diferentes personas en espacios y ambientes que en realidad no existen sin la necesidad de dispositivos adicionales a la computadora. La realidad virtual no inmersiva ofrece un nuevo mundo a través de una ventana de escritorio. Este enfoque no inmersivo tiene varias ventajas sobre el enfoque inmersivo, como: bajo costo y fácil y rápida aceptación de los usuarios. Los dispositivos inmersivos son de alto costo y generalmente el usuario prefiere manipular el ambiente virtual por medio de dispositivos familiares como lo son el teclado y el ratón que por medio de cascos pesados o guantes.

El tipo de diseño gráfico realizado en el simulador del robot Pioneer 2DX (Fig. 8) es el de realidad virtual no inmersiva, es decir no hay dispositivos externos como cascos o guantes digitales para lograr capturar otros sentidos diferentes al oído y a la vista.

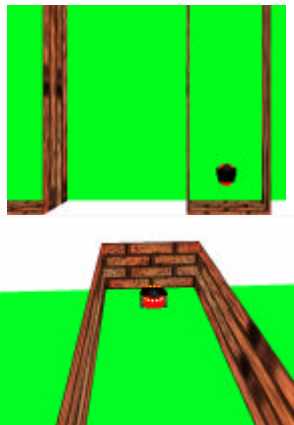


Fig. 8. Entorno virtual del robot móvil Pioneer 2DX.

## 6. RESULTADOS EXPERIMENTALES

Para validar el sensado y el control propuesto, se utilizará un entorno virtual con el modelo del robot móvil Pioneer 2DX y un entorno real, con el robot móvil real, el entorno tanto virtual como real es el pasillo del Instituto de Automática (INAUT) de la UNSJ. El valor de  $k$  es de 3.5, el de  $V_{max}$  es de 70mm/s y  $L_{min}=70$  [píxeles]. La cámara que se usó para obtener las imágenes es una webcam Creative. La computadora en la que se realizó los algoritmos de procesamiento de imágenes posee un microprocesador Pentium IV 1.2 Ghz, y el algoritmo de procesamiento de imágenes fue implementado en Matlab®, en su versión 6.5.

### 6.1. Navegación en un entorno virtual

En la Fig. 9 se muestra la referencia angular dada por el movimiento de la mano para guiar al robot virtual y el ángulo girado efectivamente por el robot para cumplir con la referencia. En la Fig. 10 se muestra la trayectoria realizada por el robot virtual sobre el pasillo virtual. En la Fig. 11 se muestran las evoluciones de las velocidades lineales y angulares del robot móvil virtual durante la navegación a través de los pasillos.

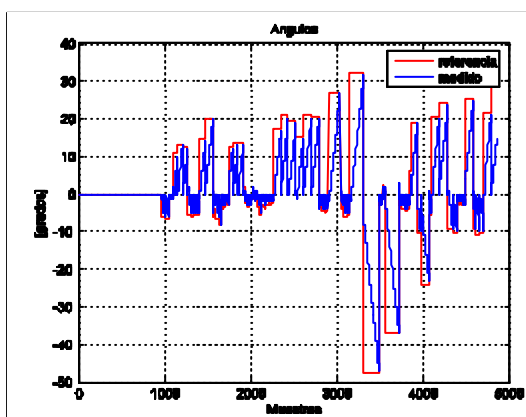


Fig. 9. Evolución de los ángulos de referencia y medido

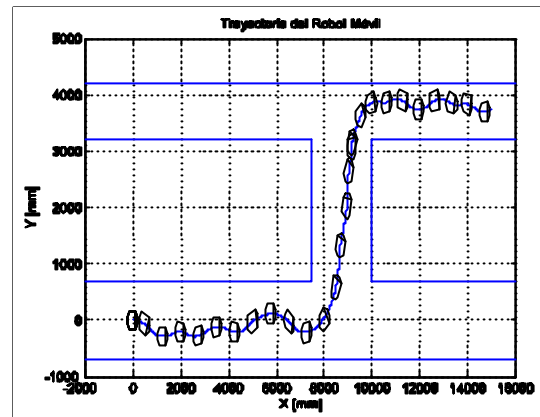


Fig. 10. Trayectoria del robot móvil en el entorno virtual.

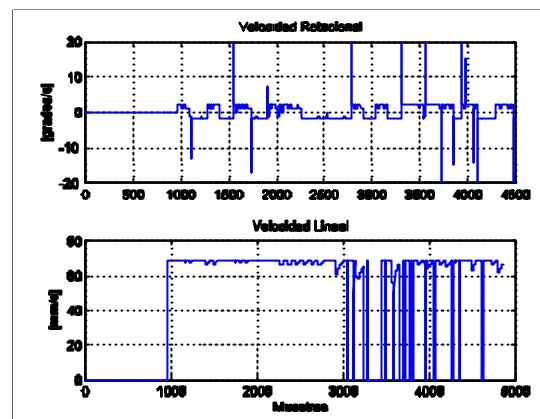


Fig. 11. Evolución de las velocidades lineales y angulares

### 6.2 Navegación en un entorno real

Para la navegación en el entorno real, se utilizarán los mismos valores de las constantes de los controladores. En la Fig. 12 se muestra la referencia angular dada por el movimiento de la mano para guiar al robot real y el ángulo girado efectivamente por el robot para cumplir con la referencia. En la Fig. 13 se muestra la trayectoria realizada por el robot móvil sobre el pasillo real. En la Fig. 14 se muestran las evoluciones de las velocidades lineales y angulares del robot móvil real durante la navegación a través de los pasillos.

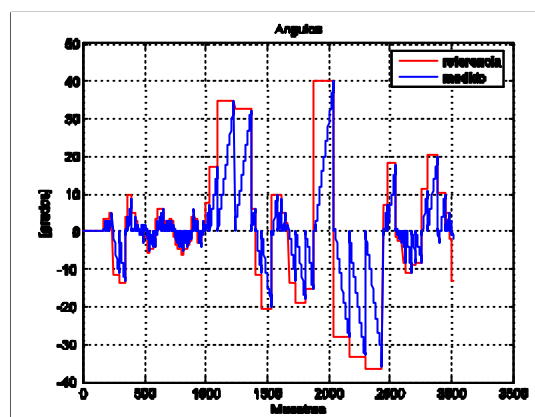


Fig. 12. Evolución de los ángulos de referencia y medido.

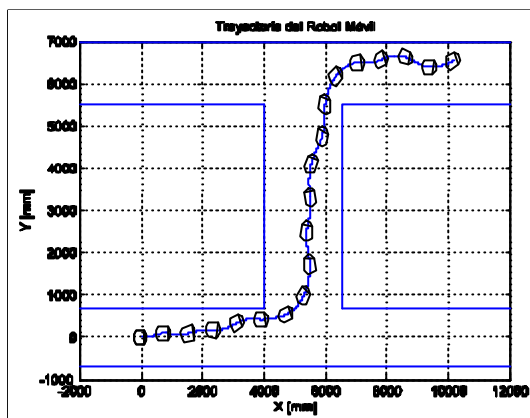


Fig. 13. Trayectoria del robot móvil en el entorno real.

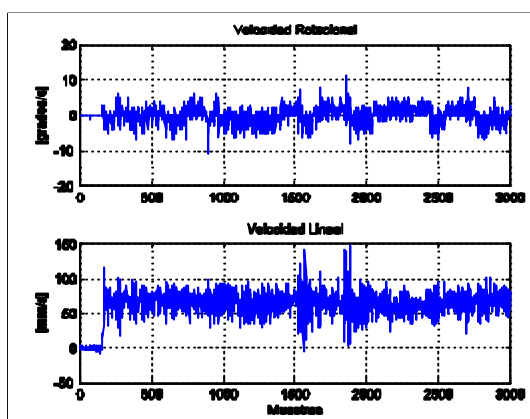


Fig. 14. Evolución de las velocidades lineal y angular del robot móvil.

## 7. CONCLUSIONES

Un sistema de interfaz basado en visión ha sido implementado para realizar la navegación de un robot móvil tanto en un entorno real como en un entorno virtual. El controlador no lineal propuesto, garantiza que los cambios necesarios en la orientación del robot móvil se cumplan, de tal manera que se pueda realizar la navegación a través de los pasillos. La utilización de un entorno virtual permite un aprendizaje y entrenamiento efectivo previo a trabajar con el robot real

Como futuro trabajo, se plantea realizar la navegación del robot móvil utilizando la identificación de los movimientos de la cabeza utilizando visión artificial.

## REFERENCIAS

- Canudas de Wit, B. Siciliano and G. Bastin (Eds.), (1997). *Theory of Robot Control*, Springer.
- Cook A. and Hussey S., (2002). *Assistive Technologies Principles and Practice*. Mosby
- Gonzalez R., R. Woods, (2002). *Tratamiento Digital de Imágenes*, Prentice Hall.
- Horn B. K P., (1986). *Robot Vision*. MIT Press, McGraw-Hill.
- Iannizzotto, G., Villari, M. & Vita, L. (2001). *Hand tracking for human-computer interaction with Graylevel VisualGlove: turning back to the*

*simple way*. Proceedings of the Workshop on Perceptive User Interfaces (PUI 2001), Orlando, FL, USA, November 15–16.

Kjeldsen, R. and Kender, J. (1996). Finding skin in color images. Proceedings of the 2<sup>nd</sup> International Conference on Automatic Face and Gesture Recognition (FG '96), Killington, Vermont, USA.

Moore D. J., Essa, I. A. and Hayes III, M. H. (1998). Objectspaces: context management for human activity recognition. Technical Report GIT-GVU-98-26, Georgia Institute of Technology, Graphics, Visualization and Usability Center, August.

Zarit B. D., Super, B. J. and Quek, F. K. H. (1999). Comparison of .ve color models in skin pixel classification. Proceedings of ICCV '99 (International Workshop on Recognition, Analysis and tracking of Faces and Gestures in Real-Time Systems), pp. 58–63, Corfu, Greece.

Zhang Z., Wu, Y., Shan, Y. & Shafer, S. (2001). Visual panel: virtual mouse, keyboard and 3D controller with an ordinary piece of paper. Proceedings of the Workshop on Perceptive User Interfaces, Orlando, FL, USA, November 15–16.