

SEGUIMIENTO DE PERSONAS POR ROBOT MÓVIL BASADO EN SEGMENTACIÓN DINÁMICA Y CONTROL DIFUSO

Leonardo Saldivar Piñón, Mario Ignacio Chacón Murguía

Instituto Tecnológico de Chihuahua
 Laboratorio de Sistemas de Percepción Visual con Aplicaciones en Robótica
 Ave. Tecnológico # 2909 Colonia 10 de Mayo
 C.P. 31310 Chihuahua, Chih.
 mchacon@ieee.org

1. INTRODUCCIÓN

La interacción entre personas y robots es cada día más frecuente en diversas áreas del quehacer humano [1]. Una parte importante en esta interacción es la capacidad del robot para percibir diferentes estímulos generados por una persona [2]. Por tal motivo es necesario contar con sistemas que sean capaces de percibir dichos estímulos, interpretarlos y alimentarlos al robot para que este a su vez responda a dicho estímulo [3],[4].

El presente trabajo describe y muestra las pruebas de un sistema basado en visión para proporcionar la capacidad a un robot móvil de seguir a una persona y mantenerse a una distancia definida. El sistema se basa en segmentación dinámica y un control difuso. Se utiliza un robot LEGO el cual tiene montado una cámara inalámbrica a color Draganfly Eyecam que funciona a una frecuencia de 2.4 GHz.

El algoritmo de seguimiento utiliza imágenes de una resolución de 176x120 en el plano de color RGB. En primera instancia se realiza una segmentación de color [5] utilizando la distancia Mahalanobis para la detección del torso del individuo. De la región segmentada se obtienen las características necesarias para realizar el control del robot utilizando Lógica Difusa [6],[7]. El algoritmo se compone de 4 secciones: Inicialización, pre-Procesamiento, adquisición de parámetros y control difuso. A continuación se describen cada una de estas secciones.

2. SEGMENTACIÓN DINÁMICA

2.1 INICIALIZACIÓN

En esta etapa el usuario realiza el entrenamiento del algoritmo para que este pueda realizar el seguimiento de la persona en los siguientes cuadros de video. El entrenamiento es mediante la

selección de una región de color. De esta región se obtiene la máscara $M(x,y)$ con la cual se obtendrá la información necesaria acerca de los pixeles seleccionados dentro de la región de interés ROI que se utilizarán más adelante.

Mediante la distancia Mahalanobis, $D_{mh}(x,y)$ entre los pixeles de entrenamiento y los demás pixeles del escenario se define las principales partes de interés del escenario $B(x,y)$.

$$B(x,y) = \begin{cases} 1 & D_{mh}(x,y) \leq \tau \\ 0 & D_{mh}(x,y) > \tau \end{cases} \quad (1)$$

$$D_{mh}(x,y) = \sqrt{(\beta - \partial)^T C^{-1} (\beta - \partial)} \quad (2)$$

donde

$$\partial = [\partial_H(x,y) \quad \partial_S(x,y) \quad \partial_V(x,y)] \quad (3)$$

$$\beta = [H(x,y) \quad S(x,y) \quad V(x,y)] \quad (4)$$

y

$$\partial_H(x,y) = H(x,y) - \mu_{MH} \quad (5)$$

$$\partial_S(x,y) = S(x,y) - \mu_{MS}$$

$$\partial_V(x,y) = V(x,y) - \mu_{MV}$$

τ es el umbral de binarización y μ_{MH} , μ_{MS} y μ_{MV} son las medias de los componentes del espacios de color HSV.

La Figura 1 muestra un ejemplo de obtención de regiones de interés preliminares para el seguimiento.



Figura 1. a) Cuadro de video, b) Regiones preliminares para seguimiento.

2.2 PRE PROCESAMIENTO

El objetivo de esta etapa del algoritmo es mejorar las regiones $B(x,y)$ de la etapa anterior para poder obtener mejores resultados en el algoritmo. Para esto, se aplican diferentes operaciones morfológicas. La última operación realizada es la eliminación de regiones no deseadas de la imagen binaria, en caso de que existan, en donde en base a un punto de referencia se obtienen las distancias de los centroides de las regiones dentro de la imagen binaria con respecto al punto de referencia. En éste algoritmo en el proceso de eliminación de regiones no deseadas el punto de referencia en el primer cuadro de video es el centro de la imagen, pero para los siguientes cuadros de video el punto de referencia es el centroide de la región deseada, el cual se va actualizando en cada cuadro de video.

La Figura 2 ilustra un caso de eliminación de regiones no deseadas en donde se muestran dos objetos, uno de los cuales es la región no deseada. Primero se observa un objeto en el primer cuadro de video, Figura 2a, en donde el punto de referencia (punto color rojo) es el centro de la imagen. En el siguiente cuadro, Figura 2b, se observa que se movió el objeto y el punto de referencia se actualizó, ahora es el centroide de la región del cuadro anterior. En la Figura 2c aparece un nuevo objeto que genera una región no deseada en la imagen binaria por lo que se procede a eliminarlo de la imagen obteniendo las distancias de los centroides al punto de referencia y eliminando la región que tenga la mayor distancia con respecto al punto de referencia.

2.3 ADQUISICIÓN DE PARÁMETROS.

De la imagen binaria resultante del proceso anterior, en caso de que sea el primer cuadro de video, se obtiene el parámetro de altura A_s de la región de interés de la imagen binaria que será utilizado como punto de ajuste para determinar a qué distancia del usuario debe permanecer el robot móvil al momento de hacer el seguimiento. La

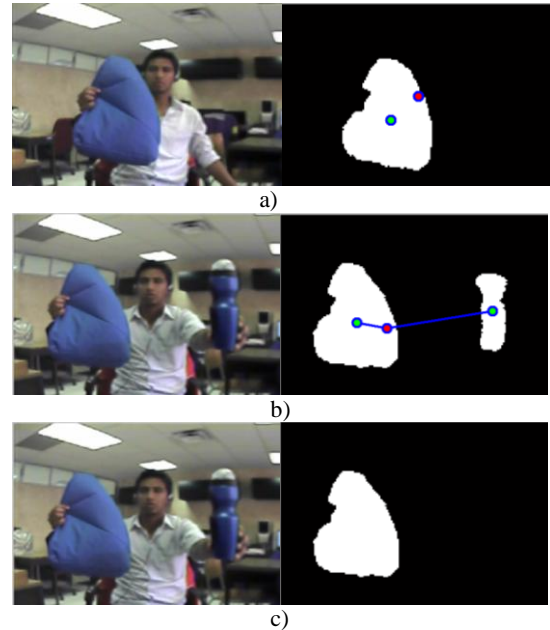


Figura 2. Proceso de eliminación de regiones no deseadas.

altura de la región de interés se obtiene con las siguientes ecuaciones:

$$S = \begin{bmatrix} t(1,1) + t(2,1) + \dots + t(c,1) \\ t(1,2) + t(2,2) + \dots + t(c,2) \\ \vdots \\ t(1,r) + t(2,r) + \dots + t(c,r) \end{bmatrix} \quad (6)$$

$$li = \begin{cases} d & \text{Si } S(d) > 0 \text{ y } S(d-1) = 0 \\ \text{nulo} & \text{De otra manera} \end{cases} \quad (7)$$

$$ls = \begin{cases} d & \text{Si } S(d) = 0 \text{ y } S(d-1) > 0 \\ \text{nulo} & \text{De otra manera} \end{cases} \quad (8)$$

$$As = ls - li \quad (9)$$

Donde S es un vector que contiene la suma de cada una de las filas de la imagen binaria; t representa la imagen binaria; r y c son el número de filas y columnas de la imagen binaria respectivamente; li es la fila en la que comienza la región binaria; ls es la fila en la que termina la región binaria.

De esta manera el robot debe mantener al usuario a determinada distancia utilizando el parámetro A_s y a la mitad del eje x de la imagen que en este caso es en la columna de pixeles numero 86 ya que se

están trabajando con imágenes de 120 filas por 172 columnas.

Para los siguientes cuadros de video se obtienen la altura A_i y el centroide (S_x, S_y) de la región de interés de la imagen binaria obtenida del pre procesamiento. Los datos A_i y (S_x, S_y) son utilizados para obtener los parámetros distancia v y posición κ los cuales son valores normalizados de -1 a 1 y son calculados de la siguiente manera:

$$\kappa = (S_x - y_x) / y_x \quad (10)$$

Donde y_x es el valor del centro de la imagen en el eje x , en este caso es 86 y κ es el parámetro posición con valores entre -1 y 1.

$$v = (A_i - A_s) / A_s \quad (11)$$

Los parámetros de posición y distancia contienen la información acerca de la ubicación del usuario con respecto al robot y son las entradas utilizadas por el control difuso que fue implementado para el control del robot y que es explicado detalladamente a continuación.

3. CONTROL DIFUSO

En este trabajo se implementa el control difuso para controlar el robot con el fin de seguir a una persona a una determinada distancia y colocarla en el centro de la imagen. El control difuso es un sistema de inferencia Mamdani [7] con un método de desfuzificación de centroide para el cual se definieron dos variables de entrada y dos variables de salida. Las variables de salida son llamadas *Motor Der* y *Motor Izq* que corresponden a los motores derecho e izquierdo del robot móvil respectivamente. Estas salidas cuentan con 2 funciones de membresía, FM, triangulares y 1 trapezoidal. Las variables de entrada definidas son: Distancia Δ y Posición ρ . La variable Δ es la entrada que nos indica que tan lejos o cercas está la persona que se está siguiendo con respecto al robot móvil. Esta variable de entrada cuenta con 3 FM trapezoidales. La variable ρ es la variable de entrada que nos indica si la persona se encuentra hacia el lado derecho o izquierdo del robot o en su defecto en el centro de la imagen el cual es el objetivo. Para esta variable de entrada también se definieron 3 FM trapezoidales.

Las funciones de membresía de las entradas y salidas fueron definidas de acuerdo a la realización del análisis del comportamiento del robot al seguir

a una persona así como de pruebas realizadas con el mismo. Se definieron los criterios necesarios para que el seguimiento fuera el deseado, como lo es la velocidad a la que el robot se movía dentro de un escenario y las vibraciones que el movimiento generaba y su efecto en la adquisición de imágenes. También se analizó la velocidad con la que el robot gira con respecto al tiempo en que la persona sale de la imagen al momento de que el robot se encuentra girando. Los valores lingüísticos que se utilizaron se muestran en la Tabla I.

Tabla I Valores lingüísticos de las variables.

Variable lingüística	Tipo	Valores lingüísticos
<i>Distancia</i>	Entrada	<i>Lejos</i> <i>Normal</i> <i>Cercano</i>
<i>Posición</i>	Entrada	<i>Izquierda</i> <i>Centro</i> <i>Derecha</i>
<i>Motor Der</i>	Salida	<i>Atrás</i> <i>Parado</i> <i>Adelante</i>
<i>Motor Izq</i>	Salida	<i>Atrás</i> <i>Parado</i> <i>Adelante</i>

El sistema diseñado consta de 9 reglas las cuáles se representan tabularmente en las siguientes matrices asociativas difusas.

Tabla II reglas para el motor derecho.

MOTOR DER		POSICIÓN		
		IZQUIERDA	CENTRO	DERECHA
DISTANCIA	LEJOS	Adelante	Adelante	Parado
	NORMAL	Adelante	Parado	Parado
	CERCANO	Parado	Atrás	Atrás

Tabla III reglas para el motor izquierdo.

MOTOR IZQ		POSICIÓN		
		IZQUIERDA	CENTRO	DERECHA
DISTANCIA	LEJOS	Parado	Adelante	Adelante
	NORMAL	Parado	Parado	Adelante
	CERCANO	Atrás	Atrás	Parado

Las gráficas de mapeo entrada – salida de cada sistema se muestran en la Figura 3.

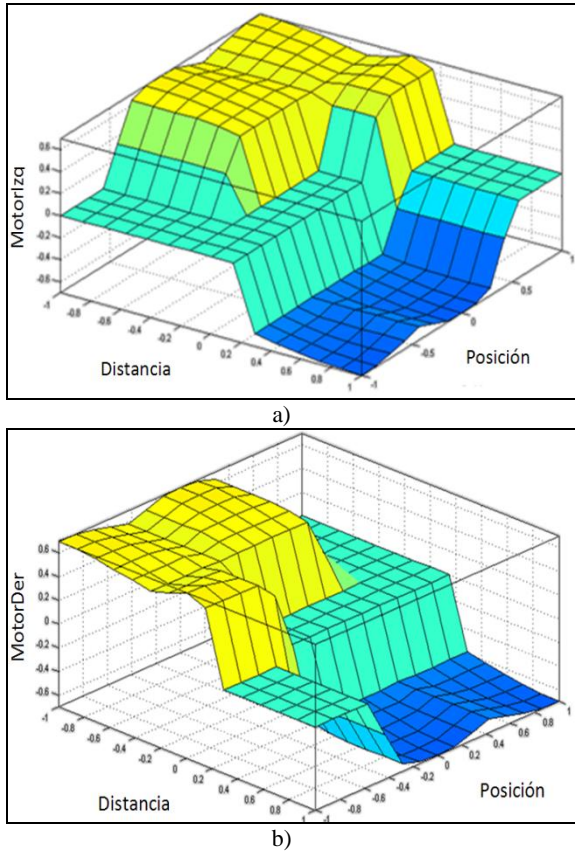


Figura 3. Gráficas de mapeo entrada – salida: a) Motor Izq, b) Motor Der.

4. RESULTADOS Y CONCLUSIONES

Resultados de experimentación donde se evalúan las salidas de los motores y el tiempo de procesamiento nos indica que el algoritmo de seguimiento de personas propuesto en este trabajo tiene un buen desempeño. En cuanto al seguimiento los experimentos muestran que en más del 90% del tiempo la posición y distancia se encuentran dentro de los rangos deseados. Respecto al tiempo de procesamiento se obtuvo mediante experimentación con 456 cuadros un tiempo promedio de ejecución de 0.1094 segundos por cuadro de video teniendo un tiempo máximo de 0.1423 seg y uno mínimo de 0.0919 seg. En otras palabras el algoritmo tiene la capacidad de procesar 9.15 cuadros por segundo lo que permite que éste algoritmo sea implementado en aplicaciones en tiempo real.

Los principios expuestos en este trabajo puede ser utilizado en aplicaciones de tiempo real para manipular robots que realicen alguna actividad específica en ambientes industriales o bien en sustituir inspecciones en ambientes de alto riesgo para humanos.

AGRADECIMIENTOS

Al Fondo Mixto de Fomento a la Investigación Científica y Tecnológica CONACYT-Gobierno del Estado de Chihuahua por el apoyo a este trabajo bajo el proyecto Clave CHIH-2009-C02-125358.

5. REFERENCIAS

- [1] T. Fong, I. Nourbakhsh, y K. Dautenhahn, "A Survey of Socially Interactive Robots: Concepts, Design, and Applications," The Robotics Institute Carnegie Mellon University, Reporte técnico CMU-RI-TR-02-29, 2002
- [2] L. Mündermann, S. Corazza, and T. Andriacchi, "The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications," *Journal of NeuroEngineering and Rehabilitation*, vol. 3, no. 6, pp. 2-11, March 2006
- [3] H. Kawamoto and Y. Sankai, "Power Assist Method Based on Phase Sequence Driven by Interaction between Human and Robot Suit," *Proceedings of the 2004 IEEE International Workshop on Robot and Human Interactive Communication* Kurashiki, Okayama Japan, pp 491-496, September, 2004
- [4] I. Rasool y H. Miyamoto, "Face and Arm-Posture Recognition for Secure Human-Machine Interaction," in *Systems, Man and Cybernetics*, 2008, 2008, pp. 411 - 417.
- [5] Clement Menier, Emond Boyer, and Bruno Raffin, "3D Skeleton-Based Body Pose Recovery," in *Proceedings of the Third International Symposium on 3D Data Processing, Visualization and Transmission*, North Carolina, USA, 2006, pp. 389-396.
- [6] I. Hernández and Ochoa I., "Control difuso y construcción de un mecanismo capaz de golpear con distintos efectos una bola de billar," Departamento de Ingeniería Electrónica, Escuela de Ingeniería, Universidad de las Américas Puebla, Puebla, México., Tesis 2004.
- [7] H. Tsoukalas and Uring R. E, *Fuzzy and Neural Approaches in Engineering*., Wiley, 1997.