

Programación Visual de Robots para Rutinas de Posicionamiento

Mata Hurtado Omar Fernando

Instituto Tecnológico Superior de Uruapan
Carretera Uruapan-Carapan No. 5555 Col. La Basilia
Uruapan, 60015
mata1786@hotmail.com

Resumen

El presente trabajo exhibe el razonamiento formal y su metodología para la programación de robots implementando algoritmos de visión artificial en rutinas de posicionamiento. Así, el motivo principal de ello, es que un robot sea capaz de aprender una tarea únicamente con visualizar la acción a realizar. Además de formularse un estado de conciencia al robot valiéndose del sentido de la visión.

Palabras clave: Artificial, Robot, Aprendizaje, Inteligencia Artificial, Redes Neuronales.

1. Introducción

Un estado ideal de la Ingeniería del Conocimiento e Inteligencia Artificial es el dotar de conciencia y perspicacia a sistemas artificiales, con el fin de crear máquinas autónomas capaces de realizar tareas con la misma eficacia y raciocinio que un ser humano. Con la formulación del actual trabajo se presenta una metodología alterna, expresada de manera formal, para la creación de una técnica de aprendizaje en un robot, mediante visión artificial.

El proceso de aprendizaje se sustenta en aquel presentado en el ser humano, donde cada uno de los sentidos, ya sea el olfato, tacto, vista, etc., nos permiten la creación de un estado de conciencia (EC) de nuestro entorno. Así, un objeto que produce cierto sonido, es captado por nuestros sentidos creando un estado de conciencia específico. Esto es almacenado en nuestra memoria de tal manera que al escuchar nuevamente ese sonido, es posible hacer referencia

conscientemente al objeto o alguna propiedad del mismo (forma, color, etc.).

Un comportamiento semejante se presenta en el aprendizaje de tareas secuenciales, como puede ser, el posicionamiento de un objeto. Donde primeramente se observan variables como orden, trayectoria, número de repeticiones, objetos involucrados, etc., para posteriormente memorizarlas y llevarlas a cabo. De ahí es posible verificar una implicación crítica del sentido de la vista en el proceso de aprendizaje.

La problemática en un robot, segado del medio que lo rodea, es la obvia incapacidad de un procesamiento autónomo, generando así sistemas torpes y rutinarios. Más aun, un sistema de visión artificial comúnmente es empleado en los sistemas robotizados como un sensor de múltiples variables. Sin embargo, si para este sensor se formulan un conjunto de algoritmos de procesamiento que constituyan una conciencia de su medio y a su vez funcionen como herramientas de aprendizaje, es posible asemejar el comportamiento humano en tareas de asociación mental. Con base en tales asociaciones, el objetivo central del actual documento es dotar a un robot de un sistema de aprendizaje basado en visión artificial, para que pueda memorizar secuencias de posicionamiento de objetos. Esto lo logrará observando solamente dicha secuencia, ya sea que ésta la realice un segundo robot o inclusive un humano.

2. La Percepción Visual

El sistema de captación visual está conformado por un conjunto de algoritmos de visión artificial los cuales están aplicados para proporcionar la mínima información de las

variables necesarias para el proceso de aprendizaje, y posteriormente, para el estado de conciencia.

Aunque existen múltiples variables que pueden ser capturadas en un sistema de visión, éste trabajo se centra principalmente en tres de ellas: forma, color y trayectoria de los objetos. La forma es la variable con mayor información para el reconocimiento de los objetos. El color es la segunda de ellas, debido a que proporciona una identidad adicional inclusive para cuerpos con la misma forma. Finalmente, la trayectoria de un objeto no solamente nos indica que existe un movimiento en éste, sino también el tipo del mismo, velocidad, entre otras características.

Para entender la formulación de los distintos algoritmos de visión, se establece lo siguiente: A) El sistema de visión está situado por encima del área de trabajo, visualizando el entorno de frente. B) Al existir desplazamiento en algún objeto se supone que la manipulación es de manera que no se obstruya significativamente el cuerpo del objeto. C) Se omite la calibración de cámaras, así como la corrección de distorsiones en las imágenes adquiridas debidas a la lente de la cámara. D) La cámara tiene una resolución aceptable y una velocidad de captación considerable como para no permitir pérdidas de información desmedidas.

2.1 Detección del Entorno

El procesamiento comienza con la adquisición de una secuencia visual (video) la cual estará siendo captada por una cámara con la suficiente resolución como para no causar pérdidas significativas de información. Puesto que el video es una sucesión continua de imágenes, es necesario discretizarlo de la forma $I_1, I_2, I_3, \dots, I_n$, donde n es el número total de particiones de la imagen continua. El tiempo asociado con una imagen I_i es aquel expresado como el producto del tiempo de muestreo z y el valor ordinal de dicha imagen dentro de una secuencia de n imágenes. Donde:

$$z = \frac{\text{tiempo total de secuencia visual}}{\text{numero de imagenes}} \quad (1)$$

Así el tiempo t asociado para la imagen I_i son $i \cdot z$ segundos. Este parámetro refleja que una imagen expresada normalmente como $F(x, y)$ ahora será referida como $F(x, y, t)$. Por lo tanto la secuencia de video original queda transformada en una secuencia discretizada expresándola como:

$$\text{video discretizado} = \sum_{t=1}^{z=\text{temporal}} I(x, y, t_z) \quad (2)$$

Nótese que si el valor de $z \rightarrow 0$, se tendría el video original, ya que dicha secuencia de video estaría conformada de un número infinito de imágenes. Aunque esto, no puede ser conseguido por la presente metodología, z debe ser lo más pequeño posible para evitar perder información crucial en el movimiento de la escena.

La segmentación en cada imagen esta expresada como una función discriminante D en términos de los valores RGB. De la forma:

$$D(r, g, b) = \begin{cases} 0, & r \leq \delta \leq Th_r \\ & g \leq \delta \leq Th_g \\ & b \leq \delta \leq Th_b \\ \alpha, & r \leq \delta \leq Th_r \\ & g \leq \delta \leq Th_g \\ & b \leq \delta \leq Th_b \end{cases} \quad (3)$$

El valor de α es un identificador para cada color donde para dos funciones discriminantes D_i y D_j , $\alpha_i \neq \alpha_j$. Según el grado de efectividad deseado, el tipo de color y la iluminación del entorno, Th adquirirá distintos valores. Así como el número de funciones discriminantes dependerá de los colores a considerar. Para todo el grupo de imágenes a procesar es el mismo grupo de funciones D .

El producto de esta función sobre cada $F_i(x, y, t_z)$ da origen a una nueva imagen segmentada S_i :

$$S_i(x, y, t_z) = \sum_{r=1}^{r=C} D_r(r, g, b) \quad (4)$$

Donde C es el número total de colores a detectar. Nótese que S_i , está constituida de hasta un máximo de C valores α , y valores cero. A éstos últimos los llamaremos valores nulos. Dichos valores nulos juegan un papel importante para una función complementaria en esta etapa, la cual es la contabilización de los objetos de la escena. Y a su vez, sirve como principio fundamental para la detección de forma y movimiento.

La metodología de contabilización de objetos, es suponer que cada objeto se distingue por un color característico. Así, cada vez que ese

color se presente implica la existencia de tal objeto en el tiempo t_i .

Entonces, el objeto esta manifestado en forma de una $D(r, b, g)$ en una $S(x, y, t_i)$. De tal manera que haciendo un algoritmo de conectividad, considerando que los limites de tal objeto son los valores nulos o una función $D(r, b, g)$ distinta, es posible contabilizar los objetos que existen en la escena.

La identificación de la forma comienza con la obtención del centroide $\Phi(\bar{x}, \bar{y}, t)$ de cada objeto contabilizado previamente [1].

$$\bar{x} = \frac{\sum x_p}{A} \quad (5)$$

$$\bar{y} = \frac{\sum y_p}{A} \quad (6)$$

Las ecuaciones (5) y (6) en conjunto, hacen referencia en un sistema cartesiano a un punto central de un cuerpo B . En general, como resultado de estas operaciones, se tendrán n centroides correspondientes a n cuerpos identificados en la contabilización. Por ello, es crítico realizar una discriminación al terminar de contar el número de objetos en la escena, con el fin de eliminar falsos elementos.

Con base al centroide, el proceso de captación de formas de los distintos cuerpos encuentra su sustento en los descriptores de Fourier, para ello, se considera cada par coordenado del contorno que describe la figura como un número complejo $s(k) = x(k) + jy(k)$ para $k = 0, 1, 2, 3, \dots, N - 1$. [2]. Con ello se aplica la transformada discreta de Fourier de $s(k)$:

$$a(u) = \frac{1}{N} \sum_{k=0}^{N-1} s(k) \exp[-j2\pi uk/N] \quad (7)$$

Para $u = 0, 1, 2, \dots, N - 1$.

Con objeto de hacer a los descriptores de Fourier invariantes ante rotaciones, traslaciones y escalados de los objetos, se llevan a cabo las siguientes acciones:

- Se toma solamente el modulo ($|a^u| = 1$).
- Se elimina la componente $F(0)$.
- Se dividen las componentes resultantes por la componente $F(1)$.

Un análisis espacial en la secuencia de imágenes rastreando el centroide con respecto al tiempo, es un método factible para determinar la trayectoria punto a punto de cada objeto. Ya que para un objeto B solo existirá un punto en el tiempo t , que representará su existencia.

A continuación se muestran algunas imágenes que denotan distintos tipos de trayectorias para cualquier cuerpo. Estas imágenes se encuentran expresadas en un espacio tridimensional desde un tiempo inicial hasta t_{n+2} .

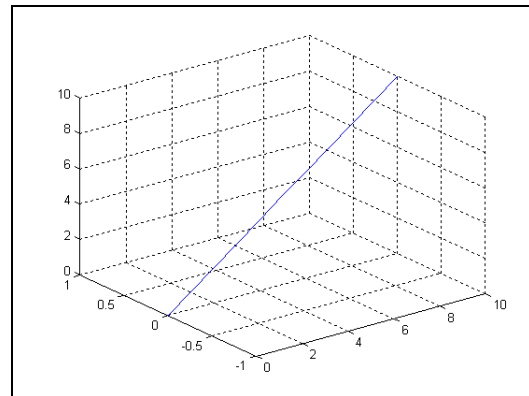


Figura 1. Movimiento de un cuerpo sobre el eje x con respecto al tiempo.

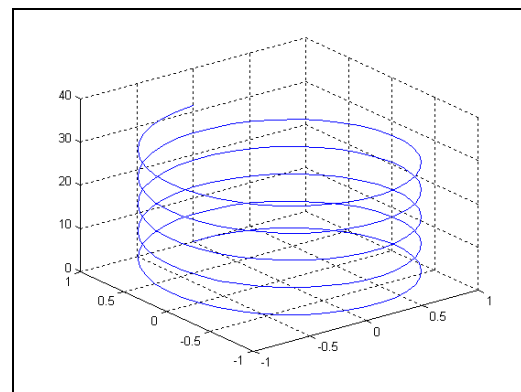


Figura 2. Trayectoria circular de un cuerpo con respecto al tiempo.

3. Ordenamiento de Parámetros y Programación Autónoma

Una tarea de posicionamiento implica la manipulación de un objeto para llevarlo de una posición p_0 a una posición p_1 , en las cuales usualmente $p_0 \neq p_1$.

En la sección 2.1 se describieron las metodologías para la obtención de los distintos parámetros que tienen influencia en nuestro sistema. Sin embargo, el aprendizaje se centra en la identificación de los cambios de posición, los cuales son una función del color y forma de cada objeto en un incremento de tiempo.

El proceso de aprendizaje está conformado por la combinación de todas las funciones correspondientes a todos los parámetros, desde un t_1 hasta t_{n+2} . Donde para cada B en el tiempo t_i existe un $S(x, y, t_i)$, $\Phi(x, y, t_i)$ y un conjunto de $a_B(u)$. La unión de estas tres funciones para un tiempo específico constituirá el Estado de Conciencia para dicho instante.

Existe una información adicional de alta relevancia en la conformación del proceso de aprendizaje, la cual viene a complementar la unión de los estados de conciencia para cada acción a realizar. Nótese anteriormente, que cada vez que se hacía alusión a una sucesión de eventos referidos al tiempo, estos daban origen en un tiempo t_1 . Esto se refería al momento en el que comenzaba una secuencia de posicionamiento.

Sin embargo, para completar el proceso de aprendizaje y ayudar a la creación de un estado subconsciente (del cual se hablará más adelante) es necesario conocer lo que existe en el tiempo t_0 . Dentro del aprendizaje, éste será el primer estado de conciencia que reciba la máquina. Tal estado de conciencia contiene todos los parámetros visuales antes de que se comience la secuencia de posicionamiento.

3.1 Árbol de Conocimiento

Todas las secuencias de posicionamiento aprendidas, en conjunto con los parámetros que dan origen al aprendizaje de éstas, están condensadas en un Árbol de Conocimiento.

Tal Árbol de Conocimiento posee una estructura característica como lo muestra la siguiente figura:

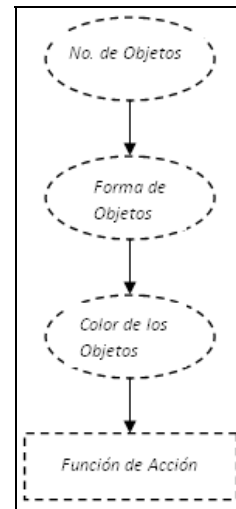


Figura 3. Estructura del Árbol de Conocimiento del Robot Aprendiz.

La figura anterior muestra el orden del Árbol de Conocimiento del robot aprendiz, y la jerarquía de cada parámetro frente a los demás. Cada nodo contiene la información obtenida en el proceso de aprendizaje correspondiente a un valor específico de una secuencia específica. Es decir, si para una secuencia p le corresponde (entre otros parámetros) un número de objetos q , y, para una secuencia r le corresponden un número de objetos s , donde $q \neq s$, entonces el árbol de conocimiento estará formado de dos nodos en el primer nivel. Por otro lado, si existen dos secuencias para las cuales $q = s$ entonces solamente se tendrá un nodo en el nivel de número de objetos con 2 hijos en el nivel inmediato inferior. Esto es así para cada una de las variables, donde tal árbol manifiesta una característica aditiva conforme a la teoría de grafos.

La información que conforma el árbol de conocimiento se obtiene completamente del estado de conciencia originado en t_0 . De ahí la importancia de tal estado. El Árbol de Conocimiento encuentra una manifestación física mediante las funciones de acción. Estas contienen las coordenadas de la trayectoria aprendida previamente.

3.2. El Estado Subconsciente, la Captación de la Realidad y su Conexión con las Secuencias de Posicionamiento

Una vez completado el proceso de aprendizaje todos los estados conscientes pasan a formar parte de las memorias del robot. Para acceder a ellos debe de existir una conexión entre

la realidad actual y el estado subconsciente. Donde, el estado subconsciente está formado por toda la información de los parámetros en el tiempo t_0 .

Considerando que el robot, continuará teniendo una sistema visual una vez terminado su aprendizaje, éste estará en un proceso de captación de la realidad constantemente, de tal manera que al estar percibiendo la realidad actual, el robot evoque todo aquello que tiene en su subconsciente. Lo anterior implica una comparación constante de su entorno actual con su subconsciente, donde este último es expresado como su Árbol de Conocimiento.

Si en un entorno visual existe un objeto con forma ξ y color α_i al cual se le asoció una trayectoria η en el aprendizaje, y tal forma, color y número existen en el Árbol de Conocimiento *solo entonces* tal robot producirá un movimiento del tipo η .

4. Las Funciones de Acción

Como previamente se comentó, las funciones de acción almacenan todos aquellos puntos de la trayectoria a realizar.

Para que el robot aprendiz, una vez conformado su proceso de aprendizaje pueda realizar la trayectoria de posicionamiento en el mundo real, es necesario que cada punto significativo de la trayectoria se convierta de las coordenadas imagen a las coordenadas del mundo real. Ya que éstas últimas son las que servirán de parámetros para que el robot realice la trayectoria de posicionamiento.

4.1 Modelo de la Cámara

Si los puntos de una imagen están dados por las coordenadas [3]:

$$\begin{bmatrix} u \\ v \end{bmatrix} \quad (8)$$

Para los cuales expresados en coordenadas homogéneas son:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \quad (9)$$

Considerando que el modelo de cámara es linear, la transformación que relaciona un punto

del espacio real tridimensional con su respectivo punto en la imagen es:

$$C \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \\ t \end{bmatrix} \quad (10)$$

Con base en (10) y siguiendo un conjunto de operaciones algebraicas es posible llegar al siguiente par de ecuaciones, las cuales están en función de parámetros de la cámara implementada.

$$a_1x + b_1y + c_1z + d_1 = 0 \quad (11)$$

$$a_2x + b_2y + c_2z + d_2 = 0 \quad (12)$$

Así para un determinado valor de Z y un conjunto de coordenadas x e y, es posible obtener el punto de desplazamiento del robot:

$$X = \frac{Z(b_1c_2 - b_2c_1) + (b_1d_2 - b_2d_1)}{(a_1b_2 - a_2b_1)} \quad (13)$$

$$Y = \frac{Z(a_2c_1 - a_1c_2) + (a_2d_1 - a_1d_2)}{(a_1b_2 - a_2b_1)} \quad (14)$$

Si se da seguimiento a todos los puntos conforme al tiempo se tiene completa la trayectoria η asociada con la rutina de posicionamiento desde t_0 hasta $t_{\eta+z}$.

5. Resultados

Con el propósito de llevar a cabo la metodología desarrollada en los apartados anteriores, se desarrolló un software que funge como medio para el aprendizaje de las distintas trayectorias de posicionamiento realizadas por un robot.

Con el fin de *cerrar* la respuesta del sistema se experimentó incorporando una red neuronal al sistema de reconocimiento del software. Esto da por hecho que el sistema solamente recibe una especie de patrones (formas) y da una respuesta conforme a un entrenamiento previo. Así, trata de *aproximar* todos los elementos visuales a ciertas respuestas específicas. Para este caso ésta ha sido entrenada orientada a elementos geométricos.

El software comentado ha sido creado en la plataforma de Visual Basic 6. La interfaz gráfica (GUI) se muestra a continuación.

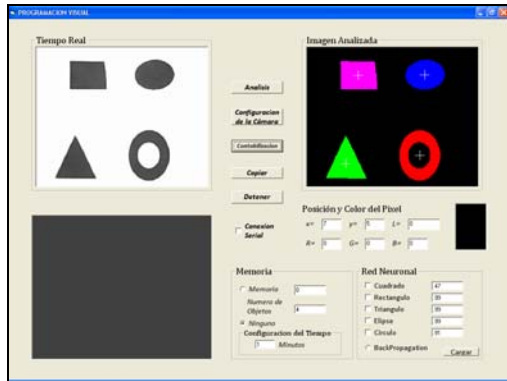


Figura 4. Interfaz gráfica creada para el aprendizaje de rutinas de posicionamiento.

Como se observa en la figura 4, la GUI consta de varios elementos de entrada y salida. Los principales elementos de salida son: Ventana de Análisis y Ventana de Trayectoria. Así como una Ventana de Video en Tiempo Real y algunos parámetros de entrenamiento de la red neuronal son los elementos más significativos de entrada.

La figura 5 muestra la trayectoria de posicionamiento aprendida por el robot en la Ventana de Trayectoria (esquina inferior izquierda). En ella es posible visualizar 3 nodos. Esto es debido a que el periodo de muestreo se estableció para discretizar la secuencia en tres imágenes. Esto se realizó para un objeto circular, de color negro.

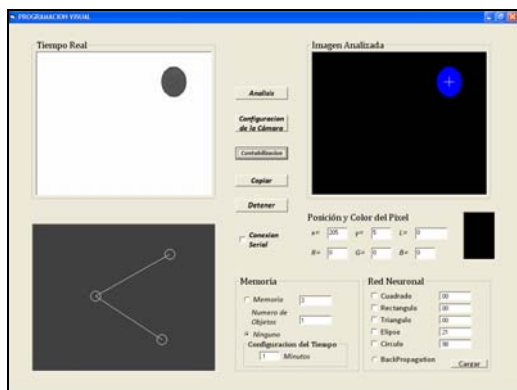


Figura 5. La Ventana de Trayectoria muestra el camino recorrido por un objeto circular.

6. Conclusiones

Lo mostrado en el presente artículo es un proceso de reconocimiento básico sin embargo eficiente, en el cual se deben asumir condiciones ideales en algunos aspectos. Más aun, este solamente involucra tareas de posicionamiento únicas o repetitivas de un robot, centrándose en el *objeto*. El propósito de este proceso de aprendizaje es la programación de un robot mediante un algoritmo dinámico y no estático. Lo que le da facilidad al robot aprendiz de memorizar una secuencia. Donde, el robot aprendiz observa la secuencia a copiar realizada por otro robot o inclusive por un humano.

Actualmente se continua con el estudio y aplicación de procesamientos más robustos para generar un conocimiento más completo y confiable. Para ello se continúa observando la relación entre las variables visuales que infieren en un proceso de aprendizaje del entorno.

Referencias

- [1] Pajares G., Santos M. *“Inteligencia Artificial e Ingeniería del Conocimiento”*, Ed. RA-MA. , España, 1ª Edición, 2005.
- [2] Gonzalez R., Woods R. *“Digital Image Processing”*, Ed. Prentice Hall, USA., 3ª Edición, 2008.
- [3] Vernon D., *“Machine Vision”*, Ed. Prentice Hall, U.K, 1ª Edición, 1991.
- [4] Bishop C., *“Neural Networks for Pattern Recognition”*, Ed. Oxford, U.K., 1ª Edición, 1995.