

Reconocimiento y localización visual 3D de objetos para robots manipuladores

Rodríguez G. J., Gómez J. L., Barranco A. I.,
Martínez S., Sandoval J.

Instituto Tecnológico de La Paz
Boulevard Forjadores de Baja California Sur No.4720, La Paz, B. C. S., México.
(joseluis.mx@live.com.mx, gjrj_1987@hotmail.com, abarrancog0900@ipn.mx
saulmd@itlp.edu.mx, jsandoval@itlp.edu.mx)

Resumen: Consideramos el problema de manipular objetos conocidos, localizados dentro del espacio de trabajo de un robot manipulador. La propuesta que presentamos incluye un sistema de reconocimiento de objetos, un sistema de cámaras convencionales calibradas para ubicar objetos en el espacio 3D y la programación necesaria para el control de un robot comercial que permite la manipulación de objetos. La metodología para calibrar el sistema de referencia de las cámaras con el sistema de referencia del robot es documentada a detalle. A fin de validar el esquema propuesto, se utilizó el robot RV-2AJ de Mitsubishi para manipular un objeto con la ayuda de dos cámaras convencionales.

Palabras claves: Robot Manipulador, Visión estereoscópica, Reconocimiento de Objetos

I. INTRODUCCIÓN

La generación actual de robots incluye un sistema de visión para la planeación de movimientos, identificación de objetos a manipular y supervisión de tareas ejecutadas. Esto motiva a mejorar e innovar sistemas de visión como elementos sensoriales en el control del robot y potenciar su programación. *Visual Servoing* es el término utilizado para referirse al control de robots con retroalimentación visual (Corke, 2011), que considera desde los materiales de los eslabones, los motores y los circuitos, hasta el sistema de visión en el modelo general del sistema. A diferencia de trabajos que presentan una metodología compacta (Colombo, 1995), (López, 2006) y (Ching, 2011); en este trabajo consideramos dos sistemas independientes: *El sistema de visión y el robot*. Ambos sistemas se relacionan entre sí mediante una transformación lineal que permite unificar el marco de referencia de las cámaras y del robot, logrando que el robot conozca la localización exacta del objeto a manipular desde las cámaras. En este trabajo se considera al robot manipulador como un sistema con entradas, salidas y restricciones. Las entradas son las coordenadas cartesianas del objeto a manipular, las salidas corresponden a la posición del actuador para tomar el objeto y las restricciones son el espacio de trabajo del robot y la velocidad máxima de cada uno de los motores del mismo.

Por otro lado, en esta propuesta, el sistema de visión trabaja sólo como un elemento sensorial que determina la ubicación exacta de la pieza a manipular y le envía la información al robot para que tome la

pieza y la lleve hasta el lugar especificado. El sistema de visión tiene como entrada el video del espacio de trabajo del robot-visión para reconocer la pieza a manipular, la salida es la posición en 3D de la pieza a manipular. Después de que el robot manipulador toma la pieza, el sistema de visión continua enviando la información de su posición, a pesar de que el robot ya no la necesita debido a que su pinza electroneumática (gripper) la está manipulando. El robot, las cámaras de visión y el objeto a manipular, como elementos principales de nuestro esquema, son mostrados en la Figura 1.

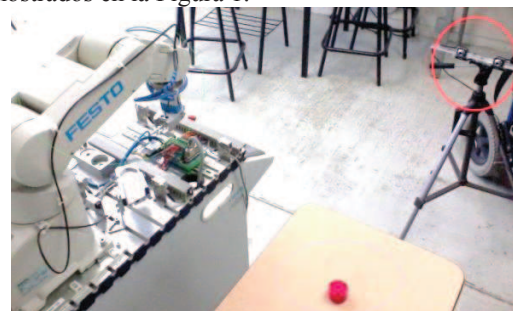


Figura 1: Sistema de visión estereoscópica con dos cámaras convencionales usadas para reconocer y localizar la pieza a manipular por el robot.

La principal contribución consiste en un esquema de robot-visión para manipular objetos conocidos en un entorno arbitrario, con la integración y descripción de las relaciones del reconocimiento de objetos 3D, su localización en el espacio de trabajo y el control del robot para manipular el objeto de interés. Cabe

resaltar que para el reconocimiento en 3D consideramos tres propiedades importantes de los objetos: Color y forma. Esas propiedades permiten discriminar con mejor certidumbre los objetos buscados de los que no son útiles en el proceso. La medición de dimensiones en el sistema nos proporciona un marco de trabajo unificado entre el robot y el sistema de visión.

El documento se encuentra organizado de la siguiente manera: La sección II define la estructura general del sistema, en la sección III muestra la metodología del sistema, los resultados experimentales son mostrados en la sección IV y finalmente en la sección V damos algunas conclusiones.

II. ESTRUCTURA GENERAL DEL SISTEMA

De manera general, el sistema de reconocimiento y localización 3D genera la posición del objeto a manipular y la envía como señal de referencia del sistema de control del robot manipulador. El flujo de información está representado por el diagrama jerárquico de módulos de la Figura 2, donde el sistema percibe del entorno imágenes conformadas por píxeles, que de acuerdo a la extracción de rasgos, interpreta la escena y produce un razonamiento a nivel de tarea; esto permite planear un movimiento del robot, generando una trayectoria de acuerdo con la dinámica de las articulaciones del mismo para reaccionar de acuerdo con la percepción visual.

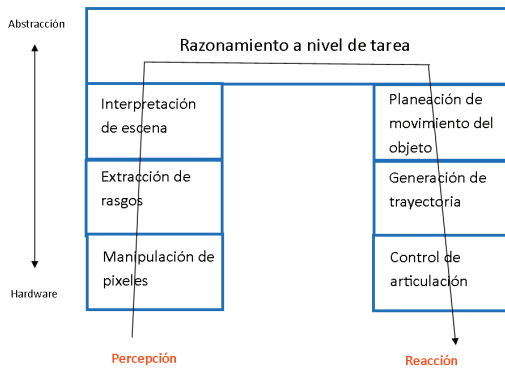


Figura 2: Estructura general del modelo jerárquico robot-visión.

4. METODOLOGÍA

La metodología propuesta se divide en dos etapas: *Offline* y *online*. Para la primera etapa se requiere:

- Calibración de cámaras.

- Calibración estereoscópica.
- Calibración entre sistemas de referencia cámara-robot.

Mientras que para la etapa online se consideran los siguientes pasos:

- Reconocimiento de objetos basado en su color y forma.
- Cálculo del centroide del objeto en las imágenes derecha e izquierda.
- Triangulación basada en los centroides.
- Cálculo de la posición en coordenadas del robot.
- Planeación de la trayectoria del robot.
- Ejecución de la trayectoria.

El sistema de visión propuesto en este documento considera dos técnicas: la primera es el reconocimiento de objetos por invariantes de Hu; y la segunda, la triangulación de puntos con visión estereoscópica. La finalidad de utilizar ambas técnicas se fundamenta en la necesidad de que el sistema reconozca un objeto deseado y lo localice respecto a las coordenadas del robot para su manipulación.

4.1 Calibración de cámaras

Para la localización automática se requiere la calibración de las cámaras. En Zhang, se propone una técnica de calibración basada en la observación de una plantilla estrictamente plana (tablero de juego de ajedrez) desde varias posiciones (Zhang, 2000). La ventaja de este método es que permite obtener los parámetros de la cámara fácilmente a partir de un sistema de referencia expresado en un plano cuadrículado y no es necesario conocer las posiciones 3D de los puntos, tampoco es necesario conocer las posiciones de la cámara desde donde se han tomado las imágenes. El plano se puede mover simplemente con la mano. Esto hace que sea una técnica muy flexible. En la Figura 3, se observa un patrón cuadrículado en diferentes orientaciones y posiciones.

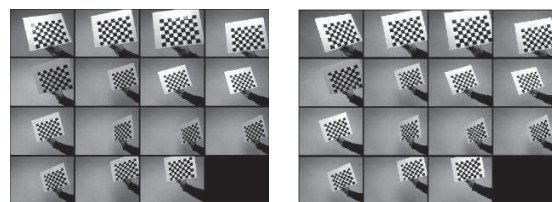


Figura 3: Patrón de calibración en ambas cámaras para usar el método de Zhang, con diferentes orientaciones y traslaciones para una mejor aproximación.

El modelo de la cámara que utiliza es el PinHole:

$$s\tilde{m} = A[R\ t]\tilde{M} \quad (1)$$

donde:

s: escalar que representa la escala de la imagen,

$\tilde{m} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$: es el mismo punto en 2D del objeto en coordenadas homogéneas,

A: es la matriz de parámetros intrínsecos de la cámara,

$[R\ t]$: es la matriz de parámetros extrínsecos de la cámara, y

$\tilde{M} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$: es un punto en 3D del objeto en coordenadas homogéneas.

Por otro lado, para calcular A utilizamos las siguientes relaciones entre A y B:

$$B = A^{-T}A^{-1} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} = \begin{bmatrix} \frac{1}{\alpha^2} & -\frac{c}{\alpha^2\beta} & \frac{cv_0-v_0\beta}{\alpha^2\beta} \\ -\frac{c}{\alpha^2\beta} & \frac{c^2}{\alpha^2\beta^2} + \frac{1}{\beta^2} & -\frac{c(cv_0-v_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} \\ \frac{cv_0-v_0\beta}{\alpha^2\beta} & -\frac{c(cv_0-v_0\beta)}{\alpha^2\beta^2} - \frac{v_0}{\beta^2} & \frac{(cv_0-v_0\beta)^2}{\alpha^2\beta^2} + \frac{v_0}{\beta^2} + 1 \end{bmatrix} \quad (2)$$

donde:

$$b = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^T \quad (3)$$

$$[h_1 \ h_2 \ h_3] = \lambda A[r_1 \ r_2 \ t] \quad (4)$$

$$h_1^T A^{-T} A^{-1} h_2 = 0 \quad (5)$$

$$h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-1} h_2 \quad (6)$$

la i-ésima columna de H es:

$$h_i = [h_{i1} \ h_{i2} \ h_{i3}]^T \quad (7)$$

y

$$v_{ij}^T = \begin{bmatrix} h_{i1}h_{j1} & h_{i1}h_{j2} + h_{i2}h_{j1} & h_{i2}h_{j2} & h_{i3}h_{j1} \\ & h_{i1}h_{j3} & h_{i3}h_{j2} + h_{i2}h_{j3} & h_{i3}h_{j3} \end{bmatrix} \quad (8)$$

se tiene que:

$$h_i^T B h_i = v_{ij}^T b \quad (9)$$

al considerar (5) y (6)

$$\begin{bmatrix} v_{12}^T \\ (v_{11} - v_{12}) \end{bmatrix} b = 0 \quad (10)$$

Cuando se ha estimado B se pueden obtener los parámetros intrínsecos de la cámara que forman la matriz A según:

$$v_0 = \frac{B_{12}B_{13} - B_{11}B_{23}}{B_{11}B_{22} - B_{12}^2} \quad (11)$$

$$\lambda = B_{33} - \frac{[B_{13}^2 + v_0(B_{12}B_{13} - B_{11}B_{23})]}{B_{11}} \quad (12)$$

$$\alpha = \sqrt{\frac{\lambda}{B_{11}}} \quad (13)$$

$$\beta = \sqrt{\frac{\lambda B_{11}}{B_{11}B_{22} - B_{12}^2}} \quad (14)$$

$$c = -\frac{B_{12}\alpha^2\beta}{\lambda} \quad (15)$$

$$u_0 = \frac{cv_0}{\alpha} - \frac{B_{13}\alpha^2}{\lambda} \quad (16)$$

Una vez que A es conocida, los parámetros extrínsecos para cada imagen son fácilmente calculados como:

$$r_1 = \lambda A^{-1} h_1 \quad (17)$$

$$r_2 = \lambda A^{-1} h_2 \quad (18)$$

$$r_3 = r_1 \times r_2 \quad (19)$$

$$t = \lambda A^{-1} h_3 \quad (20)$$

$$\lambda = \frac{1}{\|A^{-1}h_1\|} = \frac{1}{\|A^{-1}h_2\|} \quad (21)$$

4.2 Calibración estereoscópica

Debido a que necesitamos conocer la localización de objetos en un espacio 3D, calibramos el sistema estereoscópico simplemente tomando la misma escena del patrón cuadrículado con ambas cámaras para conocer el vector que une el punto principal de la cámara derecha e izquierda. En la Figura 4, se ilustra la calibración estereoscópica, donde los números en cada plano cuadrículado identifican a cada una de las fotografías en orden temporal de

captura. Las dimensiones de los ejes están dadas en milímetros.

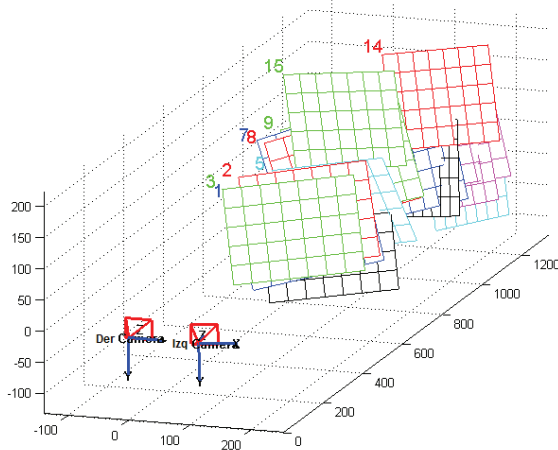


Figura 4: Cámaras calibradas en configuración estéreo, donde se conoce la transformación rígida del sistema de coordenadas de la cámara izquierda a la cámara derecha.

4.3 Calibración entre sistemas de referencia cámara-robot

Para establecer la matriz de transformación entre el sistema de coordenadas de las cámaras y el sistema de coordenadas del robot, el sistema de visión estéreo es calibrado como primer paso. Enseguida, el robot se lleva a una configuración inicial en un punto $p_0^R = (x_0^R, y_0^R, z_0^R)$ considerando un punto particular del gripper y se determinan las coordenadas de este respecto a la cámara izquierda al tomar una foto del efector y triangular la posición del punto del gripper, que es el que utilizamos como referencia para el sistema de visión $p_0^v = (x_0^v, y_0^v, z_0^v)$. Después, para obtener los vectores unitarios del sistema del robot respecto del sistema de coordenadas de la cámara, desplazamos 100 milímetros el robot a lo largo de cada uno de sus 3 ejes cartesianos y obtenemos el siguiente sistema de referencia para la cámara:

$$\begin{aligned} p_{100x}^R &= (x_0^R + 100, y_0^R, z_0^R), \\ p_{100y}^R &= (x_0^R, y_0^R + 100, z_0^R), \\ p_{100z}^R &= (x_0^R, y_0^R, z_0^R + 100), \end{aligned}$$

y sus correspondientes coordenadas p_{100x}^v , p_{100y}^v y p_{100z}^v , en coordenadas de las cámaras. Luego entonces, los elementos de la matriz de transformación se calculan de la siguiente manera:

$$\hat{a}_x^v = \frac{p_{100x}^v - p_0^v}{\|p_{100x}^v - p_0^v\|}, \quad (22)$$

$$\hat{a}_y^v = \frac{p_{100y}^v - p_0^v}{\|p_{100y}^v - p_0^v\|}, \quad (23)$$

$$\hat{a}_z^v = \frac{p_{100z}^v - p_0^v}{\|p_{100z}^v - p_0^v\|}. \quad (24)$$

tal que la matriz de transformación lineal

$$T_v^r = \begin{bmatrix} \hat{a}_x^v \\ \hat{a}_y^v \\ \hat{a}_z^v \end{bmatrix}^T \quad (25)$$

permite que el sistema de referencia del robot (Poole, 2006) sea referido al sistema de referencia de la cámara, o bien, gracias a la existencia de la matriz inversa $(T_v^r)^{-1}$ es posible la transformación inversa.

4.4 Reconocimiento del objeto por color y forma

Para identificar el color de la pieza a reconocer se utilizó el modelo de color HSV que independiza el color de la iluminación y nos permite segmentar la imagen por el color deseado. Esta técnica presenta menor costo computacional que el presentado por (Martínez, 2012) y se utilizó con:

$$b(x, y) = \begin{cases} 1 & \text{si } (h_1 > H > h_0) \text{ y } (s_1 > S > s_0) \text{ y } (v_1 > V > v_0) \\ 0 & \text{en otro caso} \end{cases} \quad (26)$$

El reconocimiento de forma del objeto se basa en los muy conocidos siete invariantes de Hu (Hu, 1962): $[\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7]$ derivados de los segmentos $b(x, y)$ mostrados en el Apéndice A. Estos indicadores permiten que siluetas sean reconocidas aunque se encuentren en diferentes partes de la imagen, rotadas o presentadas en diferente tamaño. A esto le llamamos *invariancia* a transformaciones afines. A la silueta segmentada del objeto la definimos en términos de los invariantes de Hu de la siguiente manera:

$$Ob = \begin{cases} \alpha_{1i} \leq \varphi_1 \leq \alpha_{1s} \\ \alpha_{2i} \leq \varphi_2 \leq \alpha_{2s} \\ \alpha_{3i} \leq \varphi_3 \leq \alpha_{3s} \\ \alpha_{4i} \leq \varphi_4 \leq \alpha_{4s} \\ \alpha_{5i} \leq \varphi_5 \leq \alpha_{5s} \\ \alpha_{6i} \leq \varphi_6 \leq \alpha_{6s} \\ \alpha_{7i} \leq \varphi_7 \leq \alpha_{7s} \end{cases} \quad (27)$$

donde: Ob representa al objeto buscado, α_{ki} es el límite inferior de φ_k y α_{ks} el límite superior para el k -ésimo descriptor, con $k = 1, 2, \dots, 7$.

4.5 Otras características del objeto y el robot

A partir del cálculo de centroides del objeto deseado en ambas imágenes $C_{izq} = (u_{izq}^c, v_{izq}^c)$ y $C_{der} = (u_{der}^c, v_{der}^c)$, junto con los valores de las matrices de proyección perspectiva de cada cámara P_{izq} y P_{der} (Barranco, 2011) y (Hartley y Zisserman, 2004), se triangula el centroide para obtener una estimación de sus coordenadas en 3D respecto de la cámara izquierda $C^v = (\hat{x}, \hat{y}, \hat{z})$, para después expresar ese punto en el sistema de coordenadas del robot:

$$C^R = (T_v^r)^{-1} C^v \quad (28)$$

Finalmente la planeación y ejecución de la trayectoria es delegada al controlador del robot.

5. Experimentos y resultados

Para validar nuestra propuesta, se estableció el siguiente escenario: Un objeto de color rojo fue colocado en el espacio de trabajo del robot manipulador Mitsubishi RV-2AJ (Mitsubishi, 2002) con el fin de que el sistema de visión lo identificara y localizara en 3D, para luego enviar la información al robot para que procediera a sujetarla y llevarla a una configuración arbitraria (ver Figura 5). Tanto para el procesamiento de la imagen como de la comunicación de la cámara con el robot, se desarrolló un programa en C++ con ayuda de OpenCV (Bradski, 2008) y el protocolo TCP/IP para envío de comandos de control del robot (Donahoo M. and Calvert K., 2009), extraídos de la herramienta COSIMIR, para la manipulación del mismo, vía sniffer (Sedano *et al*, 2012).



Figura 5: Robot RV-2AJ Mitsubishi a manipular una pieza roja con su gripper.

Considerando que la pieza objetivo es de color rojo se propusieron los siguientes valores $h_0 = 101, h_1 = 103, s_0 = 148, s_1 = 255, v_0 = 0$ y $v_1 = 255$, luego de analizar los rangos típicos de la pieza en el laboratorio. Para la clasificación de la pieza por

forma con los invariantes de Hu, los rangos de clasificación utilizados fueron obtenidos de 10 imágenes distintas de la misma pieza y son mostrados en la Tabla 1.

Tabla 1: Parámetros de Hu de la pieza usada en el experimento.

	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7
Min	-1.72E-01	-2.56E-03	-1.25E-06	-2.80E-08	-7.73E-12	-1.06E-07	-1.05E-12
Max	-1.82E-01	-5.48E-03	-1.49E-05	-1.59E-06	-9.26E-14	-9.01E-09	-1.15E-12

En el experimento se consideró una función binaria que recoge la información éxito del experimento en diferentes posiciones dentro del espacio de trabajo del robot, donde el valor verdadero representa que el objeto deseado fue reconocido, localizado y capturado por el robot; mientras que el valor falso representa que no se logró concretar la tarea de captura automática del objeto. En la Figura 6, se muestra el espacio de trabajo del robot en el plano x-y y con una región en color verde que representa la zona donde el sistema funcionó satisfactoriamente.

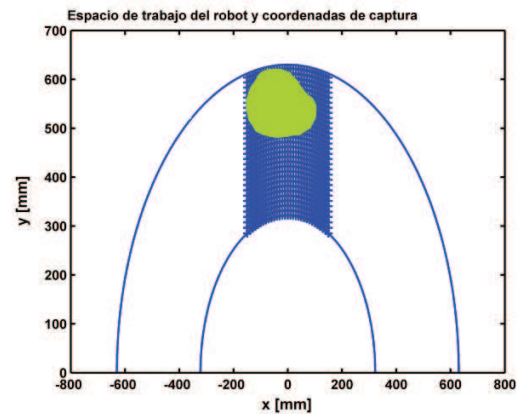


Figura 6: El plano x-y del espacio de trabajo del robot en color azul y la zona verde muestra el lugar donde el robot tomó la pieza, comandado por el sistema de visión.

6. CONCLUSIONES

Se ha presentado un esquema de visión para la manipulación de objetos con un robot, donde el reconocimiento y localización de la pieza a manipular es llevado a cabo por dos cámaras convencionales y un algoritmo que presenta un menor costo computacional que otras propuestas que realizan la misma tarea. Una ventaja de nuestra propuesta es que la ubicación de las cámaras es arbitraria y la única restricción es que el objeto a

manipular se encuentra en el espacio de trabajo del robot, y ambos sean visibles para las cámaras, mientras que otras propuestas dependen de la posición de las cámaras. Para lograr un mejor desempeño, es necesario contar con una buena resolución de las cámaras, un eficiente control del robot y del cálculo del centroide de la pieza a partir de colocarla en diferentes posiciones. Nuestra propuesta puede ser útil para la manipulación de objetos en entornos desconocidos, donde el objeto a manipular es conocido.

RECONOCIMIENTOS

Agradecimiento al Instituto Tecnológico de La Paz y al CONACyT por el apoyo otorgado a través de fondos para consolidación de grupos de investigación, así como PROMEP.

APÉNDICE A. Invariantes de Hu.

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02}\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11} \\ \phi_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} - \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} - 3\eta_{12})^2 - (\eta_{12} + \eta_{03})^2] \\ \phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{12} + \eta_{03})^2] \\ &\quad + (3\eta_{12} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} - \eta_{12})^2 - (\eta_{12} + \eta_{03})^2] \end{aligned}$$

donde:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma}, \text{ con } \gamma = \frac{p+q}{2} + 1 \text{ para } p + q = 2,3, \dots$$

$$\begin{aligned} &\sum_{x=-\infty}^{\infty} \sum_{y=-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) \\ \mu_{pq} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \end{aligned}$$

$$\begin{aligned} \bar{x} &= \frac{m_{10}}{m_{00}}, & \bar{y} &= \frac{m_{01}}{m_{00}} \\ \mu_{pq} &= \sum_x \sum_y x^p y^q f(x, y) \end{aligned}$$

REFERENCIAS

Barranco, A. y Medel J. J. (2011). Reconocimiento automático de objetos basado en la relación dimensional, *Computación y Sistemas*, vol. 15, pp. 267-272.

Bradski G. y Kaehler A. (2008). Learning openCV, *O'Reilly Media*.

Colombo C., Allotta B., y Dario P. (1995). Affine visual servoing: A framework for relative positioning with a robot. *International Conference on Robotics and Automation*, pp. 464-471.

Corke, P. (2011). Robotics, vision and control, *Springer Tracts in Advanced Robotics*, vol. 73, pp. 3-4.

Donahoo M. y Calvert K. (2009). TCP/IP Sockets in C: Practical Guide for Programmers, *Elsevier*, págs. 11-97.

Hartley, R. y Zisserman, A. (2004). Multiple View Geometry in Computer Vision, *Cambridge University Press*, Second Edition, pp. 239-249.

Hu M. (1962). Visual Pattern Recognition by Moment Invariants. *IRE Trans. Info. Theory*, vol. IT-8, pp. 179-187.

López, I. y Osorio, R. (2006). Invariant object recognition robot vision system for assembly, *Electronics, Robotics and Automotive Mechanics Conference, IEEE*, pp. 30-36.

Martínez S. (2012). Nonlinear filters for distortion-invariant pattern recognition. *2012 La investigación científica en el Instituto Tecnológico de La Paz*, pp. 15-28.

Mitubishi. (2002). RV-1A/2AJ Series Instruction manual. *Mitsubishi electric corporation*

Poole, D. (2006). Linear Algebra, a modern introduction, *Thomson*, pp. 209-224.

Sedano A., Sandoval J., Castro M., Martínez S. (2012). Laboratorio remoto basado en web para el control de un brazo robótico, *Congreso Internacional de Investigación, ISSN 1946-5351 Online, Vol. 4, No.3, Celaya, Guanajuato*.

Yang. (2011). Vision-based robot manipulator design. *Next Wave in Robotics, Communications in Computer and Information Science*, vol. 212, pp. 124-131.

Zhang Z. (2000). A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334.