

Manipulación del robot Mitsubishi RV-2A por medio de visión artificial

José Antonio Tumialán Borja* / Milton Alexander Camargo Amado** / Luis Fernando Fajardo Pérez***

Fecha de envío: 3 de mayo de 2009
Fecha de aceptación: 28 de julio de 2009

RESUMEN

En este artículo se presenta una metodología para la manipulación del robot Mitsubishi RV-2A. Ésta involucra el diseño e implementación de un *software* para la selección de eventos por medio visual con trayectorias previamente programadas en el robot industrial, con base en movimientos oculares de una persona. La captura de las imágenes de los ojos incluye el uso de filtros para, posteriormente, tratar dicha imagen. Mediante técnicas de inteligencia artificial, se entrena una red neuronal para el reconocimiento de las características de la imagen capturada y su posición; con esta información el *software* activa el sistema de control del robot para posicionarlo. Las pruebas fueron realizadas experimentalmente con cinco objetos diferentes distanciados 15 cm, cuyas respuestas fueron satisfactorias.

Palabras claves: manipuladores robóticos, redes neuronales artificiales, tratamiento de imagen.

MANIPULATION OF THE MITSUBISHI ROBOT RV-2 WITH AN ARTIFICIAL VISION

ABSTRACT

This article presents a methodology for the manipulation of the Mitsubishi RV-2A robot. The manipulation involves the design and implementation of software for the selection of events of visual means previously scheduled with trajectories on the industrial robot through eye movement of a determined person. Capturing eye movements includes the use of a filter to manage the image after and using artificial intelligence techniques and then training a neural network for the recognition of the characteristics of the captured image and its position and with this information the software activates the control system of the robot to position it. Tests were carried out experimentally using five different objects 15 cm apart, noting satisfactory answers.

Keywords: robot manipulators, artificial neural networks, image manage.

* Ingeniero de Sistemas, Magíster en Ingeniería Mecánica. Doctor en Ing. de la Universidad Federal de Uberlandia. Brasil. Profesor del programa de Ingeniería y Diseño Automatización Electrónica, Universidad de La Salle. Correo electrónico: jtumialan@lasalle.edu.co

** Ingeniero de Diseño y Automatización Electrónica, Universidad de La Salle. Gerente General SAV Ingenieros Ltda. Correo electrónico: alexander.camargo@savingenieroslt.com

*** Ingeniero de Diseño y Automatización Electrónica, Universidad de La Salle. Ingeniero Open Services Solutions Colombia. Correo electrónico: fernando.fajardo@osscolombia.com

INTRODUCCIÓN

La visión artificial es una de las áreas de mayor interés dentro del estudio de la computación y la interacción del hombre con la máquina. El propósito de este proyecto consiste en utilizar la pupila humana como dispositivo de movimiento de un teclado, un ratón, un *joystick* de computador, a fin de controlar un artefacto mecánico. Para manipular imágenes es necesario capturarlas y procesarlas sistemáticamente; a esto se le conoce como procesamiento de imágenes que consiste en mejorarlas hasta lograr la plena identificación de objetos en su interior.

En este proyecto se utilizó una cámara que digitaliza la imagen del ojo irradiado por una luz que es imperceptible para éste. Esta luz es reflejada y en varias etapas de filtrado y visión artificial, el programa obtiene una imagen de la posición de la pupila. Posteriormente, el sistema procede a enviar la señal al controlador del robot, para realizar una de las trayectorias programadas para esa ubicación de la pupila.

Para el objetivo de esta investigación se implementó una metodología basada en el reconocimiento de imágenes a partir de redes neuronales mediante el análisis de la pupila humana, el procesamiento de la información y substrayendo los recursos necesarios con el propósito de definir el punto de observación y realizar el control del robot industrial. El *software* encargado de la captura, reconocimiento y control de movimiento de las redes neuronales fue implementado en LabVIEW-IMAQ. Cabe resaltar que este proyecto no está dirigido únicamente a personas con discapacidades en sus miembros superiores, pues también está habilitado para la reducción de contacto directo de un operario en su entorno de trabajo, minimizando así la posibilidad de sufrir accidentes laborales y aumentando el control de ciertos procesos.

ANÁLISIS METODOLÓGICO

El proyecto se propició por iniciativa del grupo de investigación en bioingeniería de la Universidad de La Salle. El propósito consistió en generar un aplicativo para reconocimiento ocular de la pupila humana; para ello se implementó una arquitectura por etapas y cada una debería cumplir una parte del proceso, el cual se dividió en cinco fases: captura, procesamiento, toma de decisiones, salidas y aplicación.

La adquisición de las imágenes se hace con una interfaz RCA (conector de video estándar) por puerto USB 2.0, para obtener una secuencia de imágenes. La cámara para desarrollar este proceso es Sony con lente infrarrojo (tipo seguridad) y un filtro físico de luz Kodak (tipo gelatinoso) a fin de eliminar los ruidos provenientes de la luz ambiente.

PROCEDIMIENTO Y ADQUISICIÓN DE IMÁGENES

La calidad de las imágenes adquiridas quizá constituye el punto más importante del proyecto, ya que de éstas dependen los procesos posteriores; por ejemplo, si se obtiene una imagen de baja calidad, el resultado final no sería el adecuado. Para una adquisición de imágenes de alta calidad, se hace uso de filtros especiales, que permitan obtener datos de cualquier rostro, sin importar el color o tamaño del ojo. Se puede contar con cualquier tipo de cámara (sensor CCD o CMOS). Se debe tener en cuenta que la iluminación puede generar ruido visual para el procesamiento de las imágenes; por ello se debe contar con una fuente de iluminación infrarroja externa para dilatar aún más la pupila y ser más brillante. Este proyecto se desarrolló con una cámara de sensor CCD (figura 1), a la cual se le agregó un filtro físico de estado gelatinoso, para mejorar la calidad de la

imagen; asimismo, cuenta con un arreglo propio de emisores infrarrojos, los cuales se encargarán de irradiar la pupila del ojo.

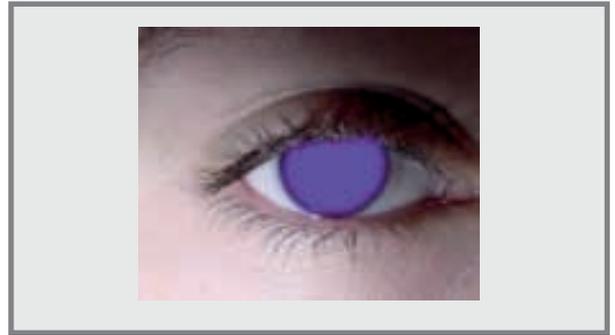
Figura 1. Cámara de sensor CCD con arreglo de emisores infrarrojos.



Fuente: Elaboración de los autores.

La cámara se aprovecha para identificar el posicionamiento del ojo en el espacio visual. Según el color del ojo, la luz reflejada cambiará en rango de luminosidad, lo que influye en la calidad de captura. En la figura 2 se observa la imagen capturada de un ojo que ha sido irradiado con luz infrarroja lo que genera una buena resolución.

Figura 2. Imagen del ojo irradiado por una luz infrarroja.

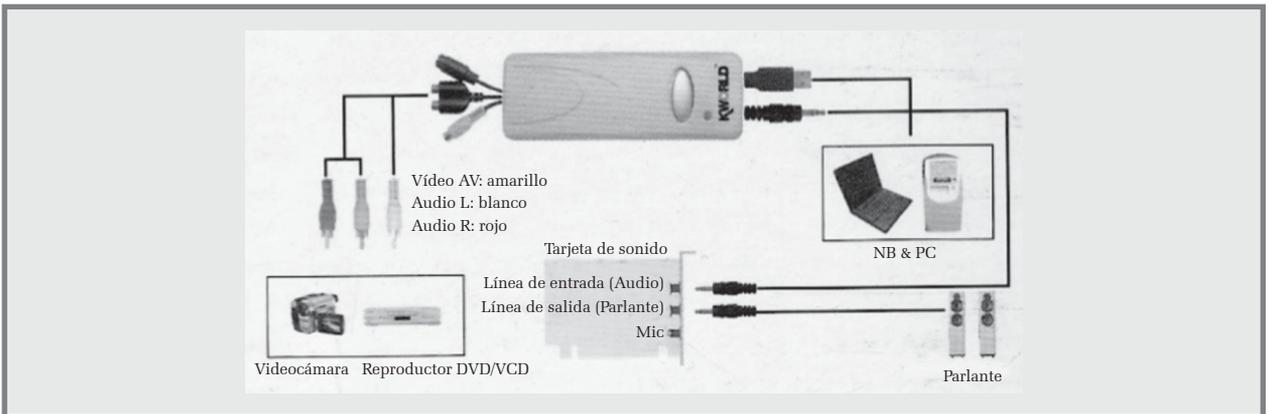


Fuente: Elaboración de los autores.

Para ingresar la imagen al computador se debe contar con un *hardware* especializado en captura de imágenes; existen varias posibilidades, como tarjetas de captura internas o externas. Por la modularidad del proyecto se decidió trabajar con una externa por el puerto USB; la resolución fue de 720 x 480 píxeles, cuya imagen produjo alta resolución y calidad superior a la de un formato de video común en computador.

En la figura 3 se puede observar la conexión de la tarjeta de adquisición externa, la cual utiliza un puerto USB para la entrada de datos.

Figura 3. Diagrama de conexión.



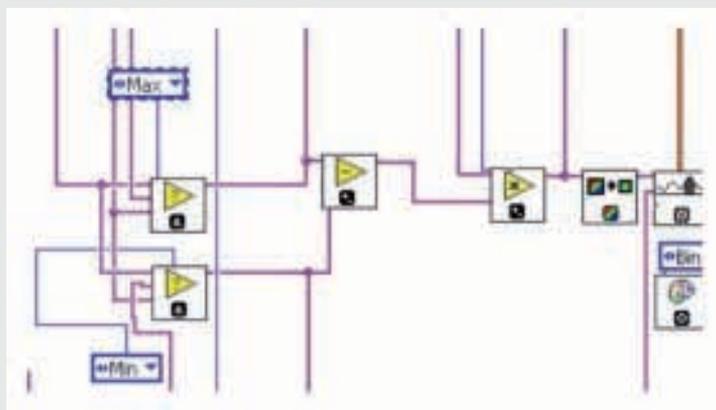
Fuente: Elaboración de los autores.

PROCESAMIENTO DE IMÁGENES

La secuencia para el procesamiento de las imágenes es la siguiente: aumentar brillo de la imagen, separar contornos borrosos, extraer características positivas –para ello se toma la alta luminosidad de la pupila irradiada–; luego, comparar con los rangos máximo y mínimo de luminosidad para extraer el rango oscuro y brillante, respectivamente; estos resultados se restan para obtener una imagen estándar, la cual

se multiplica por una constante a fin de ampliar los rasgos característicos. Este proceso se hace para filtrar la imagen y obtener rasgos brillantes altos para distinguir la pupila de otras regiones de la cara que pudieran poseer brillos por agentes externos. Con la librería del IMAQ, se extrae el círculo del ojo y se elimina la información sobrante de la imagen puesto que en la aplicación es irrelevante el rostro; únicamente es importante la posición donde el usuario esté mirando.

Figura 4. Etapa de procesamiento.



Fuente: Elaboración de los autores.

PELÍCULA KODAK INFRARROJA DE ALTA SENSIBILIDAD

Esta película negativa de sensibilidad alta y contraste moderadamente alto, es sensible a la luz infrarroja y está diseñada para penetrar la bruma en fotografía de paisajes y lograr efectos especiales en fotografía comercial y arquitectónica. Además, si se modifica el revelado y utilizan diversos filtros, también pueden darse aplicaciones científicas y médicas y copiado de documentos, así como emplear en trabajos fotomecánicos y de fotomicrografía. Esta película es sensible a la radiación cercana a la luz ultravioleta,

la luz visible y la infrarroja, hasta aproximadamente 900 nm; su sensibilidad máxima abarca de 750 a 840 nm. La película Kodak infrarroja de alta sensibilidad [7] tiene dos presentaciones: una con base ESTAR de 0,102 mm (0,004”) y otra con base ESTAR gruesa de 0,178 mm (0,007”). Su poder de resolución es de 80 líneas/mm para alto contraste y de 32 líneas/mm para bajo contraste. Asimismo, esta película se puede usar sin luz complementaria para obtener registros fotográficos en condiciones de iluminación desfavorables –como en investigaciones policiales– o en estudios sobre reacciones del público en las que sólo se necesite registrar la expresión facial.

TOMA DE DECISIONES

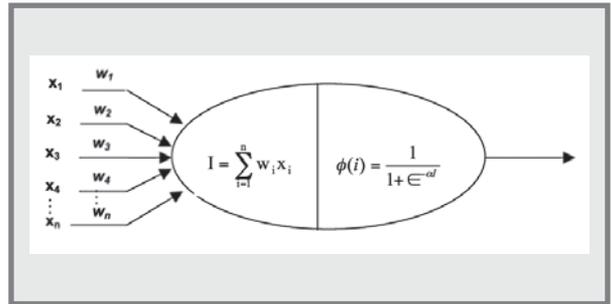
La red neuronal se encarga de accionar el actuador enviando la posición (x, y) para maniobrar el brazo. La red neuronal debe ser entrenada con gran cantidad de información para escoger la respuesta más exacta.

REDES NEURONALES ARTIFICIALES (ANN)

La arquitectura de la ANN utilizada es un perceptrón multicapas (MLP), que consiste en una capa de entrada, una de salida, una o más capas ocultas, y n neuronas en cada capa. La forma como las neuronas están interconectadas (topología o arquitectura de la red) causa un enorme efecto y una operación de la red. Las capas de una red neuronal están interconectadas a través de parámetros internos denominados pesos (w). El algoritmo *backpropagation*, con la función de activación sigmoideal, se utiliza para entrenar la red que emplea el método interactivo del gradiente descendiente el cual minimiza el error medio cuadrático entre la salida deseada y la salida de la red.

El modelo típico de la neurona utilizada en el algoritmo *backpropagation* se representa en la figura 5, en la cual las entradas x_n son conectadas por sus respectivos pesos w_n para el procesamiento de su salida por intermedio de la función de activación $\phi(i)$

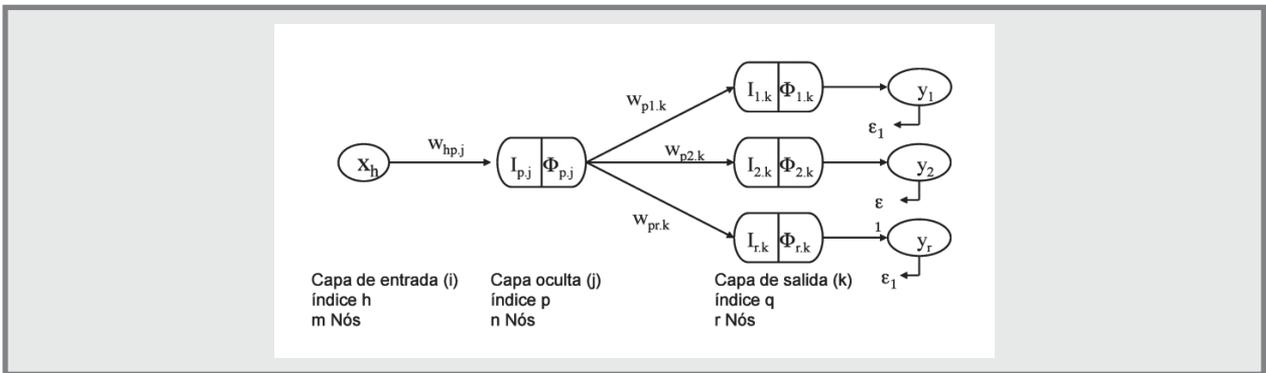
Figura 5. Neurona de procesamiento con función de activación sigmoideal.



Fuente: Elaboración de los autores.

Para comprender mejor la implementación, supóngase una red neuronal con una capa de entrada, una oculta y una de salida, tal como se muestra en la figura 6.

Figura 6. Esquema de red neuronal con capa oculta que muestra símbolos e índices.



Fuente: Elaboración de los autores.

Una vez propagadas las entradas de la red, se define el error medio cuadrático, así:

$$\epsilon^2 = [T_q - \Phi_{qk}]^2 \quad (1)$$

Donde se producen las salidas deseadas y las salidas propagadas de la red, la actualización de pesos en la capa de salida. Éstas son mudadas en proporción al error cuadrático medio en relación con los respectivos pesos. La ecuación (2) corresponde al

cálculo de la derivada para una red de dos capas (oculta y salida).

$$\Delta w_{pq.k} = -\eta_{p,q} \frac{\partial \varepsilon^2}{\partial w_{pq.k}} \quad (2)$$

Donde $\eta_{p,q}$ *learning* es una constante de proporcionalidad asociada a una tasa de aprendizaje de los pesos.

Para resolver esta derivada parcial se utiliza la regla de la cadena que genera la ecuación (3).

$$\Delta w_{pq.k} = -\eta_{p,q} \cdot \frac{\partial \varepsilon^2}{\partial \Phi_{q,k}} \cdot \frac{\partial \Phi_{q,k}}{\partial I_{q,k}} \cdot \frac{\partial I_{q,k}}{\partial w_{pq.k}} \quad (3)$$

Donde:

$$\delta_{pq.k} = 2 \left[T_q - \Phi_{q,k} \right] \Phi_{q,k} \left[1 - \Phi_{q,k} \right] \quad (4)$$

Luego, los pesos son actualizados a partir de la ecuación de recurrencia (5):

$$w_{pq,k}(N+1) = w_{pq,k}(N) + \eta_{p,q} \cdot \delta_{pq,k} \cdot \Phi_{p,j} \quad (5)$$

Obsérvese que N corresponde al número de interacciones o épocas.

Para el cálculo de los pesos que están entre la capa de entrada y la capa oculta se utilizan los siguientes derivados. Donde $\eta_{h,p}$ es otra constante proporcional de aprendizaje relativa a la actualización de los pesos. Utilizando nuevamente la regla de la cadena, la ecuación (6) puede ser escrita así:

$$\Delta w_{hp.j} = -\eta_{h,p} \frac{\partial \varepsilon^2}{\partial w_{hp.j}} \quad (6)$$

$$\Delta w_{hp.j} = -\eta_{h,p} \frac{\partial \varepsilon^2}{\partial \Phi_{q,k}} \cdot \frac{\partial \Phi_{q,k}}{\partial I_{q,k}} \cdot \frac{\partial I_{q,k}}{\partial \Phi_{p,j}} \cdot \frac{\partial \Phi_{p,j}}{\partial I_{p,j}} \cdot \frac{\partial I_{p,j}}{\partial w_{hp.j}} \quad (7)$$

Donde:

$$\delta_{hp.j} = \delta_{pq,k} w_{pq,k} \frac{\partial \Phi_{p,j}}{\partial I_{p,j}} \quad (8)$$

Luego, los pesos $w_{hp.j}$ son actualizados según la ecuación de recurrencia (9):

$$w_{hp.j}(N+1) = w_{hp.j}(N) + \eta_{hp} x_h \delta_{hp.j} \quad (9)$$

En las dos ecuaciones aparece el término η referente a la tasa de aprendizaje. Esta variable determina pequeñas modificaciones en los pesos de la red de una interacción a otra. Otra manera de evitar oscilaciones bruscas sin necesidad de escoger una tasa de aprendizaje pequeña, es la inserción del término *momentum*, que restringe las ecuaciones así:

$$w_{pq,k}(N+1) = \alpha \cdot w_{pq,k}(N) + \eta_{p,q} \cdot \delta_{pq,k} \cdot \Phi_{p,j} \quad (10)$$

$$w_{hp,j}(N+1) = \alpha \cdot w_{hp,j}(N) + \eta_{hp} x_h \delta_{hp,j} \quad (11)$$

SISTEMA DE POSICIONAMIENTO

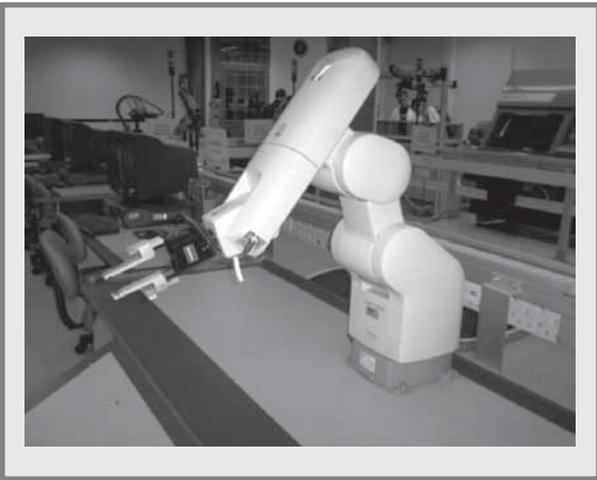
Para dar a conocer la posición a la cual el usuario dirige la mirada, el programa señala una salida gráfica, representada por cinco indicadores –uno para cada posición. Cada vez que la captura se ubica sobre un punto, la red neuronal confirma la posición y el indicador visual muestra la posición analizada previamente, y genera una instrucción, la cual es enviada al manipulador robótico a fin de ejecutar la trayectoria predeterminada para esa posición.

MANIPULADOR ROBÓTICO

En la figura 7 se muestra el robot Mitsubishi RV-2A de tipo industrial. Este brazo robótico se controla por medio del *software* COSIROB [6], el cual es muy similar al entorno de CNC y cuenta con comunicación

TCP/IP. Se programaron diez rutinas específicas: cinco para liberación y cinco para transporte. La combinación de estas rutinas abarca todas las posibilidades de movimiento de un punto a otro. El programa en LabVIEW [9] envía instrucciones para localizar y ejecutar las líneas programadas en COSIROB, que servirán para guiar al manipulador de un punto inicial a otro punto de llegada, liberando el elemento transportado al final del recorrido. El robot se puede mover inicialmente a cinco puntos, en los cuales puede o no estar ubicado un objeto. Si se observa un punto donde no hay objeto, el robot no se mueve; pero si al observar el punto allí existe un objeto, el robot se mueve hasta la posición indicada visualmente, toma el objeto y se traslada hasta el punto de llegada, correspondiente al sitio de almacenamiento y lo suelta.

Figura 7. Robot manipulador Mitsubishi RV-2A.

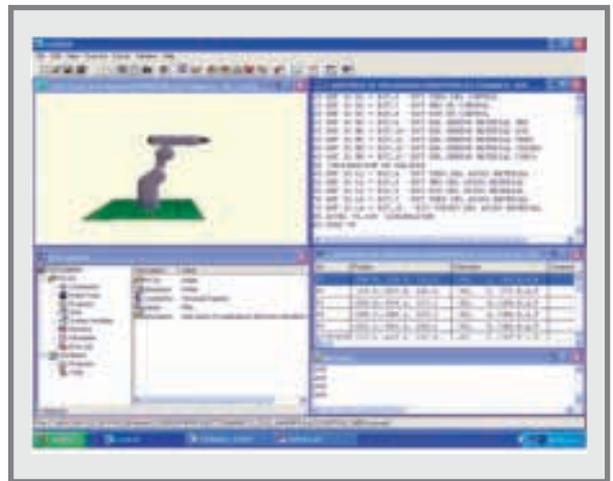


Fuente: Elaboración de los autores.

El *software* COSIROB se divide en cinco paneles, tal como se puede ver en la figura 8. En el panel superior izquierdo se pueden simular las posiciones del robot que cuenta con una ventana donde aparecen éstas. En la parte inferior izquierda se encuentra el panel explorador y de comunicación, el cual se encarga de configurar la comunicación computador-manipulador robótico; asimismo, muestra los pro-

gramas guardados en el controlador, *slots* ejecutados, errores enviados, pantallas de indicadores, variables y programas realizados. En la parte inferior derecha se encuentra el panel de mensajes de comunicación; por ejemplo, cuando una orden es enviada o recibida del controlador, el panel envía los mensajes de estado. En la parte central derecha se encuentra un panel en el que se configuran las posiciones y orientaciones de cada uno de los estados finales definidos por el usuario. Finalmente, en la parte superior derecha se encuentra el panel de programación; en éste se escriben las instrucciones que junto a las posiciones se almacenarán en el controlador del robot.

Figura 8. Software de control.



Fuente: Elaboración de los autores.

ANÁLISIS DE RESULTADOS

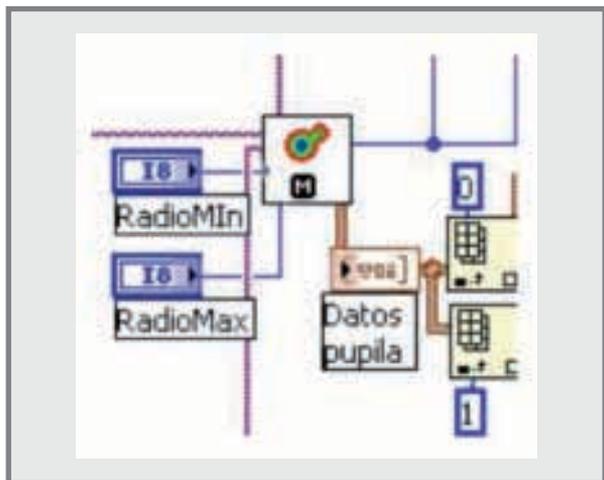
Este trabajo presentó resultados representativos de los análisis y pruebas que se han realizado, en las etapas implementadas en el *software*: captura, filtrado, procesamiento y toma de decisión mediante la red neuronal para la manipulación de un robot industrial.

En el proceso de captura, el problema principal es el elevado precio de la tarjeta de adquisición; por tanto, se buscó una interfaz común para cualquier computador; a través del puerto USB resulta más económi-

ca sin sacrificar calidad de imagen (resolución). Con respecto a las cámaras, se realizaron pruebas de irradiación con distintos tipos de cámaras: las de vídeo con un arreglo externo de *leds* infrarrojos, de seguridad blanco y negro, como de color, pero la cámara que más se acomodó al sistema fue Sony con arreglo interno de *leds* infrarrojos y alta resolución. Con esta cámara, la influencia de la luz externa es muy pequeña y la calidad de imagen es bastante buena.

El procesamiento, debido a la resolución de la cámara se vio afectado, pero con modificaciones en cuanto a las limitantes de calidad, se logró buena respuesta, al separar los centroides del ojo de las demás partes de la imagen. El programa para la separación se muestra en la figura 9.

Figura 9. Extractor de círculos de IMAQ.



Fuente: Elaboración de los autores.

CONCLUSIONES

Debido a que este proyecto se encuentra en una línea amplia de investigación, presenta resultados parcia-

les satisfactorios. En la etapa de captura de imágenes se tiene una irradiación de alta calidad –aun con luz día– pero lo óptimo se logra con baja luz natural. El procesamiento de imágenes presenta buenos resultados, y esto comprueba la efectividad de la red neuronal para el control del robot. Los resultados con la interacción del robot y el computador son satisfactorios, pues se logra precisión durante la manipulación del robot.

Las tarjetas de adquisición, debido a que son diseñadas para tal fin, son muy costosas y deben ser montadas internamente sobre un PC lo que resta movilidad a un proyecto. La decisión de captura de información por puertos USB fue acertada, pues no se requirió compra ni instalación de ningún tipo de tarjeta.

La continua evolución de la visión artificial permite que ingenieros e investigadores no se limiten a los puntos clásicos de control (modelos matemáticos), sino a mirar más allá del entorno diario, mediante el uso de información como la que se obtiene con los sentidos básicos del ser humano. El proyecto se queda en un prototipo funcional, en espera de su continuidad, por parte de otros investigadores interesados en esta temática, para así seguir fortaleciéndolo y consolidándolo aún más. Esto será posible con grupos de bioingeniería cuyo objetivo sea facilitar y equilibrar aún más la vida laboral de las personas que tengan discapacidades físicas.

AGRADECIMIENTOS

Este trabajo ha sido patrocinado por Laboratorio de Investigaciones del programa de Ingeniería de Diseño y Automatización Electrónica de la Universidad de La Salle.

REFERENCIAS

- Awcock, G. y Thomas, R. 1996. *Applied Image Processing*. McGraw-Hill.
- Ballard, P. y Stockman, G. 1995. *Controlling a Computer via Facial Aspect*. *IEEE Trans Sist Man Cyber* 25(4), pp. 669-677.
- Baluja, S. y Pomerlau, D. 1994. Non-Intrusive Gaze Tracking Using Artificial Neural Networks. En: *Advances in Neural Information Processing Systems*. Cowan JD, Tesauro G, Alspector J (Eds).
- Ben-Yacoub, S. 1997. *Fast Object Detection using MLP and FFT*. IDIAP-RR 11, IDIAP.
- De la Escalera, A. *Visión por computador fundamentos y métodos*, Prentice Hall.
- Digital Toy Shop. Filtro Infrarrojo para Kodak EasyShare Z7590. http://www.digitaltoyshop.com/product.asp?pt_id=481&p_id=5604&lg=2c=FILTRO_INFRRARROJO_para_Kodak_EasyShare_Z7590&c=FILTRO_INFRRARROJO_para_Kodak_EasyShare_Z7590. Página consultada el 15 de octubre de 2009.
- Galer, M. y Les Horuat. Tratamiento digital de imágenes, Anayamultimedia.
- Hemmerling, R. Design and simulation of working cells and motion sequences with COSIROB,
- Kim, K y Ramakrishna, R. 1999. Vision-Based Eye-Gaze Tracking. *IEEE International Conf. on Systems, Man, and Cybernetics*, Tokyo, Japan.
- Martin, Hall T. y Martin, Meg L. *LabVIEW for Automotive, Telecommunications, Semiconductor, Biomedical and Other Applications (National Instruments Virtual Instrumentation Series)*, Prentice Hall.
- National Instruments Corp. 1999. *IMAQ Vision User Manual*.
- National Instruments Corp. LabVIEW. Entorno de programación gráfica. <http://www.ni.com/LabVIEW/esa/>.
- Paton, Barry E. *Sensors, Transducers and LabVIEW: An Application Approach To Virtual Instrumentation*, Dalhousie University.
- Rojas S. A., Rubiano L., y Cantor D. "Detección de rasgos oculares mediante visión artificial", en *Memorias del II Simposio de Investigación Universidad de La Salle*. Bogotá, Septiembre, 2001.
- Russell, S. y Norvin, P. *Artificial Intelligence and Modern Approach*, Prentice Hall.