

# Visión para Robots en Tareas de Ensamble

M. Peña,\*I López-Juárez,\*J. Corona,\*K.Ordaz

\*CIATEQ,A.C.

IIMAS-UNAM

e-mail: [mario@leibniz.iimas.unam.mx](mailto:mario@leibniz.iimas.unam.mx)

[ilopez@ciateq.mx](mailto:ilopez@ciateq.mx)

## Resumen

*En el trabajo se presenta un sistema de visión que le permite a un robot manipulador KUKA de 6 grados de libertad (DOF), obtener información del POSE (localización del objeto) para realizar tareas de ensamble en línea (tiempo real).*

*El sistema de visión forma parte de una celda de manufactura inteligente que se esta integrando en el Laboratorio de Manufactura Inteligente del CIATEQ en Querétaro, México.*

*Se presenta la arquitectura del sistema diseñado, y la descripción del método utilizado para obtener información de la posición y la orientación en tiempo real de imágenes 2D obtenidas con técnicas de segmentación basándose en el análisis de histogramas 1D y 2D. La base de funcionamiento con formas de imágenes binarias, hace al sistema robusto y flexible, lo que permite utilizarlo también en aplicaciones de “visual servoing”. El proceso de adquisición de la imagen, proceso de calibración y procesamiento de los datos para generar las siluetas (imagen binaria) y extraer características (features) es mostrado.*

*También se presenta la integración sensorial y descripción general de los algoritmos de inteligencia utilizados en el modo de visión en la celda de manufactura.*

Palabras Clave: Sistema de Visión, ensamble, sensor CCD, histograma, celda de manufactura inteligente, robot manipulador, cámara progresiva.

## 1. Introducción

Recientemente, se ha generalizado la idea de contar con celdas inteligentes de manufactura que utilicen robots con capacidades sensoriales y algoritmos inteligentes basadas en las últimas investigaciones y desarrollos en Inteligencia Artificial (IA) como son: lógica difusa y redes neuronales. Utilizar estas técnicas, simplifica el manejo de modelos matemáticos complejos que requieren de bastante tiempo de cómputo para su implementación. Utilizar sistemas multimodales que obtengan información de diferentes modos sensoriales con el propósito de manufactura, nos hace pensar en sistemas más robustos, auto-adaptables y parecidos al comportamiento humano que necesariamente mejoran el desempeño, y flexibilidad en aplicaciones comunes con robots manipuladores [1].

Diferentes sensores en sistemas de manufactura, han sido utilizados para realizar diferentes tareas muy específicas y por separado como son: guía de trayectorias para soldar piezas, sistemas de visión para aplicaciones de inspección y control de calidad y sensores de fuerza/torque para realizar tareas de ensamble. La integración de robots con capacidades sensoriales para realizar tareas de ensamble en una celda flexible de manufactura, representa un tema de vanguardia en la investigación y desarrollo de sistemas de manufactura inteligente. La mayoría de los usuarios y fabricantes que utilizan sistemas automatizados en la fabricación, han puesto su empeño en mejorar la precisión y velocidad de los sistemas robotizados, pero a un lado de las

habilidades cinemáticas de un robot, sus capacidades sensoriales no han sido desarrolladas y son necesarias para proveer al robot con un alto grado de flexibilidad y autonomía en sus tareas de manufactura. Para lograr este objetivo es necesario cumplir con los requerimientos de un ambiente de producción, cuyas características pueden resultar un factor crucial en el proceso de fabricación como son:

- *incertidumbre en la posición de las herramientas o piezas de trabajo.* Para solventar esto, se utilizan partes y accesorios que garantizan posiciones fijas para realizar los ensambles pero resultan costosas, muy particularizadas y generalmente con necesidad de ser ajustadas en ciertos intervalos de tiempo de uso.
- *operación de los manipuladores con piezas de trabajo que se mueven.*
- *evaluación de fuerzas de interacción y torques.*

La idea entonces, es contar con robots auto-adaptables con capacidades sensoriales y de aprendizaje de habilidades en la fabricación en línea. Para ello, se necesita una integración de diferentes modos de percepción sensorial, de manera que el desempeño se parezca al comportamiento humano en sus tareas de aprendizaje y ejecución en procesos de manufactura. (operaciones de ensamble)[2].

Los modos que han sido más utilizados son: reconocimiento de voz, visión y sensado de fuerza.

En CIATEQ, A.C., Centro de Tecnología Avanzada, se han desarrollado por muchos años diseños y fabricación de máquinas especializadas para procesos de manufactura específicos, actualmente se ha concebido la necesidad de desarrollar celdas inteligentes de manufactura con capacidades sensoriales. En el trabajo, se presenta la arquitectura de la celda con los modos

sensoriales que se han integrado, haciendo énfasis en el sistema de visión. El brazo-robot tiene la habilidad de realizar operaciones de ensambles (*peg-hole operation*), que pueden ser mejorados con la experiencia. Para este fin se utilizan redes neuronales artificiales (ANN) ART (*Adaptive Resonance Theory*) Fuzzy ARTMAP en conjunto con una base de conocimientos dinámica. La parte medular en la celda es un brazo-robot de 6 grados de libertad que realiza tareas de ensamble y obtiene información sensorial de reconocimiento de voz, visión y sensado de fuerza.

## 2. Arquitectura del Sistema

Como se muestra en la figura 1 y 2, la arquitectura del sistema de celda de manufactura, está formada por una PC maestra industrial en donde residen una tarjeta de sensado F/T basada en DSP, el software de reconocimiento de voz y el NNC (Neural Network Controller). En una computadora dedicada se implementa el sistema de visión y reside el "*frame grabber*". Otros elementos de la celda son: el manipulador industrial KUKA KR15, controlador KRC2, el Panel de Control KCP, sensor de fuerza torque JR3 F/T, cámara progresiva con sensor CCD y micrófono. El brazo-robot y su controlador están interconectados con dos cables especializados para obtener la potencia e intercambio de datos, la computadora maestra se conecta al sistema de visión y controlador del robot mediante puertos seriales usando el protocolo Xon/Xoff.

Reconocimiento de voz (*voice recognition*), sistema de visión (*machine vision*) y sensado de fuerza (*force sensing*), son los modos sensoriales que actualmente se están integrando y han sido utilizados exitosamente para proporcionar al robot habilidades para manipular objetos y realizar tareas de ensamble. Dentro del objetivo principal se encuentra la implementación en la celda de un planificador de tareas y una arquitectura

neuronal multimodal para percepción sensorial en robots industriales. [3], [4].

remoción de percepciones, considerando así el utilizar M<sub>2</sub>ARTMAP para la implementación.

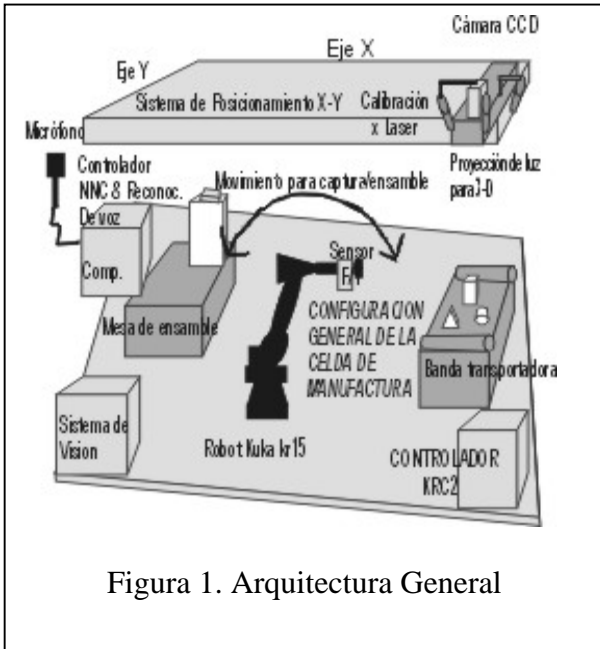


Figura 1. Arquitectura General



Figura 2. Celda de Manufactura CIATEQ

## 2.1 Multimodalidad

La multimodalidad está restringida a las interacciones que comprenden más de una modalidad en la percepción como:

Percepción Sensorial	Organo	Modalidad
Sentido de la vista	Ojos	Visual
Sentido del Oído	Oídos	Auditiva
Sentido del Tacto	Piel	Táctil
Sentido del Olfato	Nariz	Olfativa
Sentido del Gusto	Lengua	Gustativa
Sentido del equilibrio	Organo de equilibrio	Vestibular

El desarrollo de la celda SIM (Sistema integrado de Manufactura), se basa en las investigaciones de López-Juárez de aprendizaje maquina en un sistema robótico de ensamble que utiliza sensado de fuerzas de contacto, esta investigación ha sido extendida para incorporar vision artificial y reconocimiento de voz y proporcionar mayor autonomía al sistema robótico, con esto surge la necesidad de una arquitectura neuronal multimodal que facilite la integración o

El trabajo hace énfasis en el modo sensorial de visión y el objetivo es proporcionar al robot la información necesaria por medio de un *vector descriptivo* que le permite aproximarse a tomar una pieza que se quiera ensamblar dentro de un volumen específico de trabajo y pasar luego a realizar tareas autónomas de ensamble. La información de que objeto se trata, es aprendida por el sistema utilizando un modelo de red neuronal M<sub>2</sub>ARTMAP, desarrollado por el grupo de investigación en mecatrónica y manufactura inteligente del CIATEQ.

El sistema inteligente de manufactura hasta el momento cuenta con cuatro modos sensoriales: reconocimiento de voz, visión, reconocimiento de escritura y sensado de fuerza. Cada modo sensorial se encuentra en una etapa de desarrollo diferente y la idea es alcanzar una integración de modos de percepción sensorial para utilizarlos con robots industriales en procesos de manufactura.

## 2.2 Reconocimiento de Voz

La interfase hombre-máquina de alto nivel se lleva a cabo por reconocimiento de voz, utilizando una interfase de programa de

aplicación comercial (API) de Dragon Systems. El control Active-X se ha incorporado al controlador principal del brazo-robot, de manera que puede ser instruido a moverse en: movimientos X-Y, velocidad y selección del cuadro de coordenadas del robot por medio de voz.

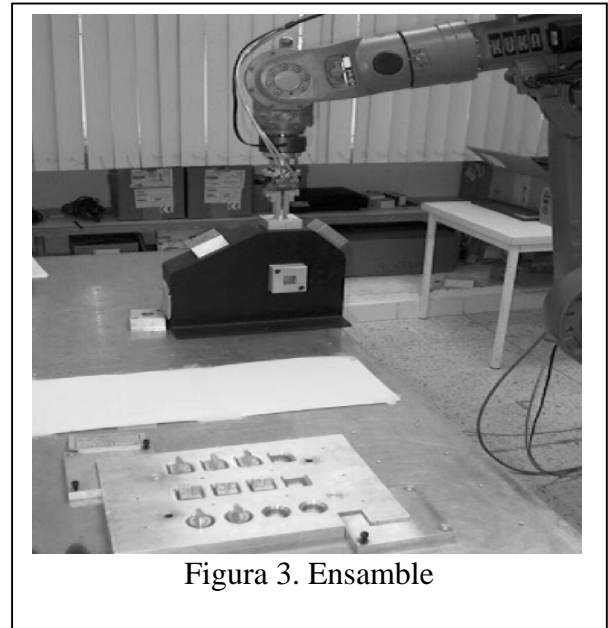
### 2.3 Sensado de Fuerza y Habilidad de Aprendizaje

El sistema que habilita el aprendizaje para el ensamble, está implementado en la computadora maestra con un sistema de sensado de fuerza que provee al NNC (Neural Network Controller) de la información requerida para aprender la operación de ensamble en línea. El diseño del NNC está fundamentado en la robustez de ART para el aprendizaje incremental combinándolo con una base de conocimiento primitiva dinámica (PKB), que inicialmente provee al robot con un conjunto mínimo de información de condiciones de fuerza de contacto y alineación de movimientos. El conocimiento es mejorado en línea basado en el suceso de predicción de movimientos que reducen las fuerzas que impidan realizar ensambles más rápidos mejorándose así el desempeño [5].

### 2.3 Ensamble

Típicamente, los sistemas de ensamble robotizados trabajan con una gran incertidumbre asociada a diferentes factores durante el proceso de ensamble que pueden llevar a una operación a fallar, como lo son: errores de posición, perturbaciones durante el ensamble, sacudidas en los engranes de las juntas del robot, ruido, etc. Tener un sistema de visión que realimente los movimientos iniciales del brazo-robot para posicionarse en el lugar y realizar la operación de ensamble proporciona robustez y certeza en los procesos de este tipo en líneas de manufactura. La idea es dar al manipulador capacidades parecidas a como lo hacen los humanos, primeramente se

obtiene la información de la localización del objeto que se tomará para el ensamble y en donde se va a ensamblar, con información visual (modo visual) se dirigen los movimientos para aproximarse al objeto (*pose*) y tomarlo (*grasping*), luego aproximarse al lugar de ensamble y “cambiar” de modo sensorial (modo sensado de fuerza) para obtener información de fuerza (tacto) y realizar el detalle fino del ensamble. Las figura 3. muestra esta idea.



### 3. Sistema de Visión

El modo sensorial de visión proporciona al brazo-robot, la información espacial en la localización (*POSE*) y captura de partes (*grasping*). La figura 4 muestra al sistema de visión integrado por: una cámara de televisión de alta velocidad del tipo progresiva con sensor CCD/B&W, PULNIX 6710 [6], colocada en un sistema de posicionamiento “X-Y” en la parte superior de la celda de manufactura, lo que permite tener escenas de “partes” en diferentes localidades dentro del área de trabajo.



Figura 4.- Localización de cámara en sistema de posicionamiento X-Y.

La cámara de televisión tiene una resolución de 648x482 pixeles y puede adquirir “pixeles” hasta velocidades de 40Mhz con 32 bits en modo LVDS y 20MHz. con la interfase RS-422, está conectada a un “*frame grabber*” PC-DIG de Coreco Imaging [7], instalado en una computadora PC. La imagen es adquirida y procesada para obtener información confiable dentro de un rango de iluminación considerable en tareas de manufactura dentro de un ambiente industrial. Implementar un sistema de iluminación para optimizar estas primeras fases es una etapa que se encuentra en desarrollo, con lo que se espera también aplicar métodos originales para la obtención de información en 3D basados en técnicas de segmentación de imágenes de proyecciones del objeto.

La imagen es procesada para obtener una imagen binaria sobre la cual se aplican

algoritmos que permiten transformar la forma en una colección ordenada de pares numéricos la cual es analizada y utilizada para obtener información de las coordenadas del centroide y puntos importantes de la forma que permiten el cálculo de la orientación del objeto para obtener el POSE [8]. Por el análisis de éstos pares numéricos se forma un vector descriptivo del objeto que se integra a la información del POSE para formar el CFD (*current frame descriptor/on line*), el formato de este vector puede ser alimentado al modelo de red neuronal que se utiliza y desarrollado por K. Ordaz e I. López. La información del vector descriptivo es adquirida por la computadora maestra a través de un puerto serie para instruir al controlador los movimientos necesarios para alcanzar la parte que se quiere atrapar y luego realizar la tarea de ensamble específica que le mande el sistema planificador de tareas desarrollado por J. Corona e I. López. La figura 5. muestra la arquitectura general del sistema de visión en donde se plantea el esquema que se utiliza dentro de la celda de manufactura, trabajos en desarrollo contemplan llevar al sistema de visión a realizar la adquisición de la imagen de modo estereoscópico o con una configuración de cámara en mano en un futuro.

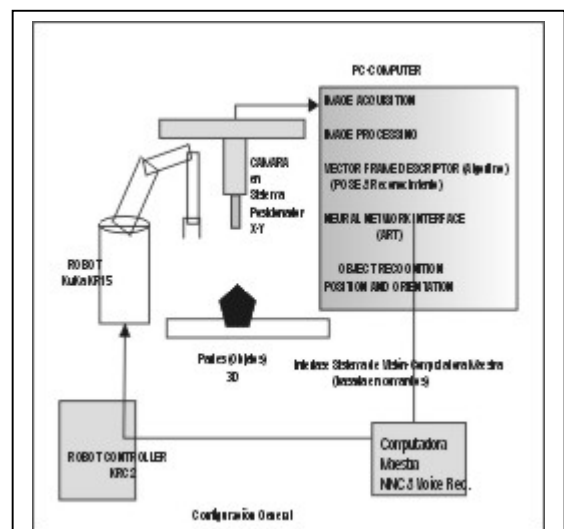
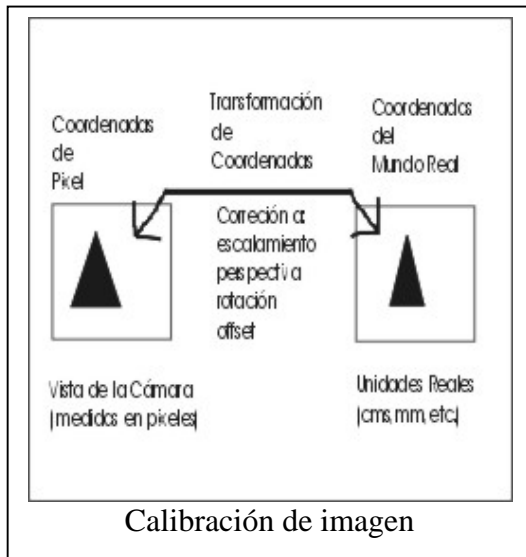
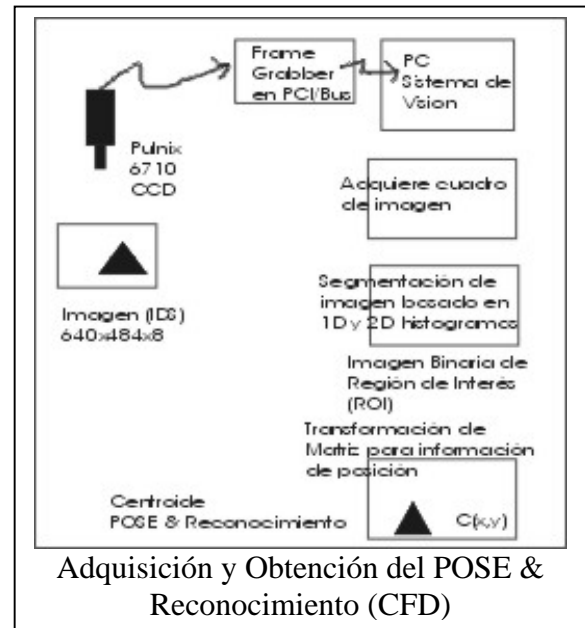


Figura 5.- Configuración General del Sistema de Visión

El método de calibración utilizado para obtener las coordenadas de imagen (en píxeles) con las coordenadas de mundo real (en escena real de trabajo), se resume en el diagrama:



deseados, el método y algoritmo que se utiliza se muestra a continuación:



En el método se siguen los siguientes pasos:

- Se establece un sistema de coordenadas en 2D con cuatro puntos proyectados en el área de trabajo.
- Se establece la localización de los 4 puntos dentro de la imagen
- Se asocian las posiciones de los 4 puntos en la imagen con medidas de las coordenadas del mundo real
- Se aplica el algoritmo de corrección para escalamiento, perspectiva y rotación
- Se obtienen las transformaciones de medidas en píxeles a unidades reales
- Cada vez que se cambie de localidad en la cámara se recalibra el sistema

Una vez calibrado el sistema se procede a adquirir la imagen en cuestión para obtener la información de POSE & reconocimiento

Ejemplos de imágenes antes y después de aplicar el algoritmo para la obtención con la información para formar el vector descriptivo (CFD) con dimensiones de 32x32 que contiene la información de las propiedades del objeto, se muestran en la figura 6.

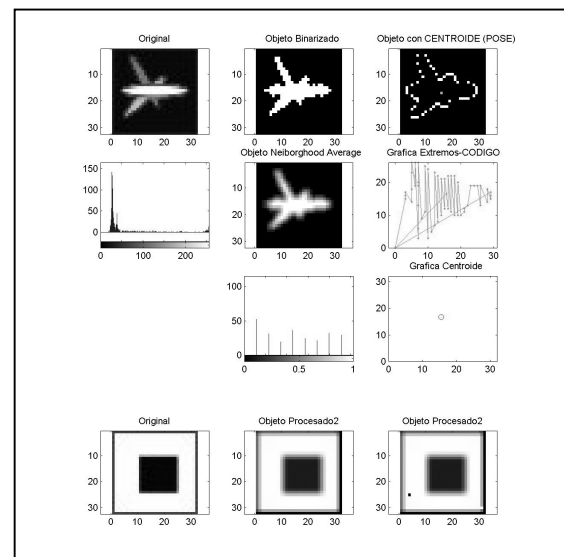


Figura 6. Imágenes del Proceso de Obtención del POSE

#### 4. Resultados y conclusiones

Al momento se han obtenido resultados satisfactorios con el sistema de visión para la localización de partes en línea, con la restricción de conocer previamente la información de la altura de los objetos, se han realizado experimentos con figuras básicas en 2D como cuadros, círculos, rectángulos, triángulos y polígonos en general, para luego realizarlos con polyhedros como cubos, pirámides, y conos. En un porcentaje alto el resultado es satisfactorio para la obtención del POSE, en cuanto al reconocimiento el resultado es satisfactorio actualmente con figuras en 2D, la implementación del sistema de iluminación que permitirá obtener la información de 3D está en desarrollo y no se han obtenido resultados. La metodología para obtener las mediciones se realiza en base a la lectura de mediciones repetitivas para el mismo objeto varias veces y para diferentes objetos y se establecen cálculos estadísticos básicos como medias y variancias para tomar conclusiones (los métodos estadísticos no son presentados en este trabajo). Las mediciones se hacen considerando un cierto rango de iluminación fuera del cual el sistema es propenso a equivocaciones. Resultados esperados se han obtenido al poder obtener información del POSE en tiempo real, lo que permite realizar aplicaciones en línea y proporcionar a un sistema planificador de tareas de ensamble la información necesaria para alcanzar la parte requerida, capturarla e iniciar el proceso de ensamble con otros modos sensoriales como el sensado de fuerzas y el reconocimiento de voz.

#### 5. Referencias

[1] I.López-Juárez, On-Line Learning for Robotic Assembly using artificial neural networks and contact force sensing, PhD. Thesis, Nottingham Trent University, Inglaterra, 2000.

[2] L. Wu, S. L. Oviatt, P. R. Cohen, “Multimodal Integration – A Statical View”, IEEE Transactions on Multimedia, vol 1 , Num. 4, pp 334-341, 1999.

[3] I. López-Juárez, M. Howarth. Learning Manipulative Skills with ART. IEEE/RSJ International Conference on Intelligent Robots and Systems ( IROS´2000), Takamatsu, Japan, Vol 1, pp 578-583 ISBN 0-7803-6351-5.

[4] K.Ordaz, I. Lopez-Juarez, “Hacia M<sub>2</sub>ARTMAP: Una Arquitectura Neuronal Multimodal para Percepción Sensorial en Robots Industriales”, TAM2003, Taller de Aprendizaje de Máquinas dentro de la 4ª. Conf. Int. en Ciencias Computacionales 2003 (ENC´03), Tlaxcala , México, 8-12 Sept.2003.

[5] I. López-Juárez, M. Howarth. Knowledge Acquisition and Learning in Unstructured Robotic Assembly Environments, The International Journal of Information Sciences, 2002.

[6] [www.pulnix.com](http://www.pulnix.com)

[7] [www.imaging.com](http://www.imaging.com)

[8] M. Peña, I. López-Juárez, R. Osorio. Invariant Object Recognition Applied to Machine Robot Vision (In Spanish). 17<sup>th</sup>. Conference of the Mexican Society for Instrumentation´2002, October 2002.