

# UNA NUEVA MÉTRICA PARA UTILIZAR FILTROS DE CORRELACIÓN EN EL RECONOCIMIENTO DE OBJETOS CON EL ROBOT HUMANOIDE NAO

*A NOVEL METRIC TO USE CORRELATION FILTERS IN OBJECT RECOGNITION WITH THE NAO HUMANOID ROBOT*

**Cesar Augusto Puente Montejano**

Universidad Autónoma de San Luis Potosí, México  
*cesar.puente@uaslp.mx*

**José Ignacio Núñez Varela**

Universidad Autónoma de San Luis Potosí, México  
*jose.nunez@uaslp.mx*

**Francisco Eduardo Martínez Pérez**

Universidad Autónoma de San Luis Potosí, México  
*eduardo.perez@uaslp.mx*

**Rogelio Castillo Morquecho**

Universidad Autónoma de San Luis Potosí, México  
*morquecho.rogelio@gmail.com*

**Recepción:** 22/octubre/2019

**Aceptación:** 23/noviembre/2019

## Resumen

En este trabajo, se propone una nueva métrica para el uso del filtro de función discriminante sintética (Synthetic Discriminant Function, SDF por sus siglas en inglés) en el problema de reconocimiento de objetos. Se realiza una serie de experimentos con el filtro SDF en la plataforma de programación del robot humanoide NAO, que permiten observar un comportamiento de la nueva métrica (Peak to Neighboring Values, *PNV* por sus siglas en inglés) y predecir comportamientos futuros en situaciones similares. Con los experimentos realizados se concluye que la métrica *PNV* mejora notablemente la medición del desempeño del filtro, generando mejores resultados que las métricas convencionales, específicamente en los objetos que tienen variaciones en su apariencia, como cambios de escala y de rotación. Calificaciones altas en el desempeño brindan una mayor seguridad para determinar que el objeto ha sido reconocido.

**Palabras Claves:** Filtros de correlación, peak to neighboring values, reconocimiento de objetos, robot NAO.

## **Abstract**

*In this paper, a new metric is proposed for the use of the Synthetic Discriminant Function (SDF) in the problem of object recognition. A series of experiments are carried out with the SDF filter in the programming platform of the NAO humanoid robot, which allow observing a behavior of the new metric (Peak to Neighboring Values, PNV) and predicting future behaviors in similar situations. With the experiments carried out, it is concluded that the PNV metric significantly improves the measurement of the filter's performance, generating better results than conventional metrics, specifically on objects that have variations in their appearance, such as changes in scale and rotation. High performance ratings provide greater security to determine that the object has been recognized.*

**Keywords:** *Correlation filters, NAO robot, object recognition, peak to neighboring values.*

## **1. Introducción**

Muchas de las actividades que el humano realiza cotidianamente podrían verse beneficiadas con la asistencia de un robot humanoide. Para que el robot pueda realizar estas actividades, debería ser capaz de percibir e interactuar con el mundo que lo rodea. Una buena percepción del entorno, como por ejemplo el reconocimiento visual de objetos, hace que la realización de algunas tareas sea más eficiente.

Las cámaras situadas en el robot NAO le permiten obtener imágenes digitales, con las cuáles es posible analizar el mundo que lo rodea y realizar el reconocimiento visual de los objetos. De acuerdo con [Andreopoulos, 2013], un objeto es un conjunto de características que algo debe poseer bajo todas las perspectivas y condiciones con las que puede ser percibido. Por lo tanto, reconocer un objeto es darse a la tarea de buscar un conjunto de características en una imagen que sean similares a las características del objeto que se buscan.

El reconocimiento de objetos se compone de subtareas como la detección y localización [Tsotsos, 2005]. La detección se define como decidir si un objeto se encuentra en la imagen o no, mientras que la localización se refiere a proveer la ubicación del objeto en la imagen.

El proceso de reconocimiento de objetos presenta diversos problemas derivados de variaciones en la apariencia que tiene el objeto en la imagen. Estas variaciones, normalmente son ocasionadas por el sensor (i.e. cámara) y/o el entorno en el que se encuentra el objeto. Los problemas ocasionados por el sensor se originan debido a las características físicas de la cámara. Propiedades como la resolución de la imagen, la sensibilidad a la luz, el auto-enfoque, el acercamiento óptico y digital, y la saturación de color, determinan la calidad de la imagen que se obtiene [Hedgerow, 2004]. Con respecto al entorno donde se encuentra el objeto, las variaciones en el objeto son ocasionadas por las distintas perspectivas con la que se puede observar un mismo objeto. Esta diferencia de perspectivas es ocasionada por las diferencias en las distancias y posiciones con las que se observa el objeto. Es importante que el reconocimiento de objetos sea robusto, es decir que tenga una gran tolerancia a los factores previamente descritos, para que el robot pueda realizar actividades de forma eficiente.

El filtro de correlación de función discriminante sintética (Synthetic Discriminant Function , SDF por sus siglas en inglés), postulado por [Casasent, 1984], ha sido usado en diferentes aplicaciones donde el reconocimiento de patrones es fundamental para el desempeño de las mismas. Trabajos como el de [Savvides, 2002] en el que se utiliza el filtro de correlación para reconocimiento de rostros, o el de [Álvarez, 2003] donde se realiza el reconocimiento de organismos vivos de la especie *Acartia*, por mencionar algunos. [Martínez, 2010] desarrollaron una estimación de actividades basados en el comportamiento de un objeto. Por su rapidez en el procesamiento, recientemente se han utilizado filtros de correlación en aplicaciones para el control de vehículos autónomos. Por ejemplo [Zhao, 2018] utilizan los filtros de correlación como base para el entrenamiento de una Red Neuronal Convolutiva para identificar y localizar objetos frente al vehículo. También, [Han, 2017] proponen un algoritmo de detección de vehículos basado en un nuevo

filtro de correlación. Los filtros de correlación de función discriminante sintética se han convertido en una herramienta importante para el reconocimiento de objetos en imágenes digitales, además de poseer unos de los mejores desempeños en términos de precisión al reconocer un objeto y descartar un objeto diferente [Aguilar, 2014].

El objetivo de este trabajo es implementar en el robot humanoide NAO un sistema reconocedor de objetos, basado en filtros de correlación, que extraiga las características de un objeto para posteriormente buscarlas en una imagen. Además, se propone una nueva métrica que mejora la evaluación del reconocimiento de objetos con variaciones en su apariencia. La métrica propuesta evalúa los valores que se encuentran en un área alrededor del valor máximo para saber qué tanta diferencia de magnitud existe entre el valor máximo y estos valores. Entre mayor diferencia, existe una mayor certeza de que el objeto ha sido reconocido correctamente. La métrica propuesta se denomina Relación Pico a Valores Vecinos (*Peak to Neighboring Values*, PNV por sus siglas en inglés).

## 2. Métodos

### Filtros de Correlación

Los filtros digitales pueden ser utilizados para el reconocimiento de objetos dentro de una imagen [Kumar, 2005]. Mediante el teorema de correlación es posible comparar dos imágenes y determinar si son iguales o tienen cierta similitud. Si se compara una imagen  $H(u,v)$  donde se tiene el objeto a reconocer y otra imagen  $F(u,v)$  con la escena donde se busca el objeto, una similitud alta determina que el objeto a buscar se encuentra dentro de la escena. La ecuación 1 expresa la operación de la correlación utilizada en los filtros digitales:

$$G(u, v) = H(u, v) \otimes F^*(u, v) \quad (1)$$

Donde el símbolo  $*$  denota el complejo conjugado.

A través de los años se han propuesto distintos filtros de correlación que mejoran el desempeño en ciertas áreas del procesamiento de imágenes. Comenzando por los filtros básicos como el filtro clásico de acoplamiento (Classical Matched Filter, o

CMF) [Lugt, 1964], filtro de sólo fase (Phase Only Filter o POF) [Horner, 1984] y el filtro inverso (Inverse Filter o IF) [Ramírez, 2009]. Sin embargo, en el problema de reconocimiento de objetos es necesario tomar en cuenta los cambios de forma que aparentemente tiene el objeto cuando es observado desde diferentes perspectivas [Javidi, 1997]. Por lo tanto, es importante considerar estas variaciones para lograr un reconocimiento robusto. Para resolver esto, se han propuesto los llamados filtros de correlación compuestos. Estos filtros utilizan imágenes del objeto a reconocer tomadas desde las perspectivas más representativas del mismo. Estas imágenes son consideradas de entrenamiento.

Cada filtro puede ser entrenado para reconocer cualquier objeto siempre y cuando la variación esperada sea capturada en las imágenes de entrenamiento. Un filtro compuesto combina un conjunto de imágenes como entrenamiento para detectar un objeto. A este proceso se le denomina sintetización. Uno de los filtros de correlación compuestos más utilizados es el filtro de función discriminante sintética (Synthetic Discriminant Function o SDF). Este filtro se representa como una función  $h(x, y)$  que se define en la ecuación 2.

$$h = S(S^t S)^{-1} c^* \quad (2)$$

Donde  $S$  representa una matriz formada por las imágenes de entrenamiento,  $S^t$  representa la transpuesta conjugada de la matriz,  $( )^{-1}$  representa la inversa de la matriz,  $c$  representa un vector con los valores deseados de correlación para cada imagen de entrenamiento (denominado vector de coeficientes), y  $c^*$  su complejo conjugado. El vector  $c$  tiene como función discriminar las imágenes de entrenamiento, en donde un 1 significa que la imagen corresponde al objeto y se denomina clase verdadera, y un 0 indica que esa imagen de entrenamiento no es el objeto, y se denomina como clase falsa. Para este trabajo, sólo se utilizan imágenes de entrenamiento de clase verdadera, ya que se pretende trabajar en escenas donde se desconocen los objetos que se encuentran en ellas.

Si se desea tener una mayor tolerancia a las distorsiones del objeto en la imagen, usar filtros no lineales puede hacer estos filtros menos sensibles a los cambios de escala y rotación [Aguilar, 2014]. Para convertir el filtro SDF en un filtro no lineal, se aplica un operador de no linealidad a la matriz  $S$ , denominado Ley-k. Por lo tanto,

ahora se representa de la forma  $S^k$ , redefiniendo la ecuación 2, dando paso a la ecuación 3.

$$h^k = S^k(S^{kt}S^k)^{-1}c^* \quad (3)$$

De acuerdo al trabajo de [Aguilar, 2014], un coeficiente de  $k = 0.3$  ofrece el mejor desempeño para el reconocimiento de rostros.

La figura 1 muestra el proceso completo en el que se crea el filtro SDF y su correlación con una imagen de la escena para encontrar un objeto. Primero las imágenes son transformadas al dominio frecuencial mediante la transformada de Fourier (FT, por sus siglas en inglés), después se aplica el operador de no linealidad a cada una de ellas. Posteriormente, se realiza la sintetización que resulta en la creación de una plantilla que representa el filtro SDF. Por otro lado, la imagen que representa la escena donde posiblemente el objeto se encuentra, es transformada al dominio frecuencial y después se aplica el operador de no linealidad. Esto se correlaciona con el filtro SDF, y el resultado es devuelto al dominio espacial, mediante la transformada inversa de Fourier (Inverse Fourier Transform, IFT por sus siglas en inglés). Esto genera lo que se conoce como el plano de correlación. Si el objeto se encuentra dentro de la escena, el plano de correlación contendrá un pico alto y delgado (figura 1).

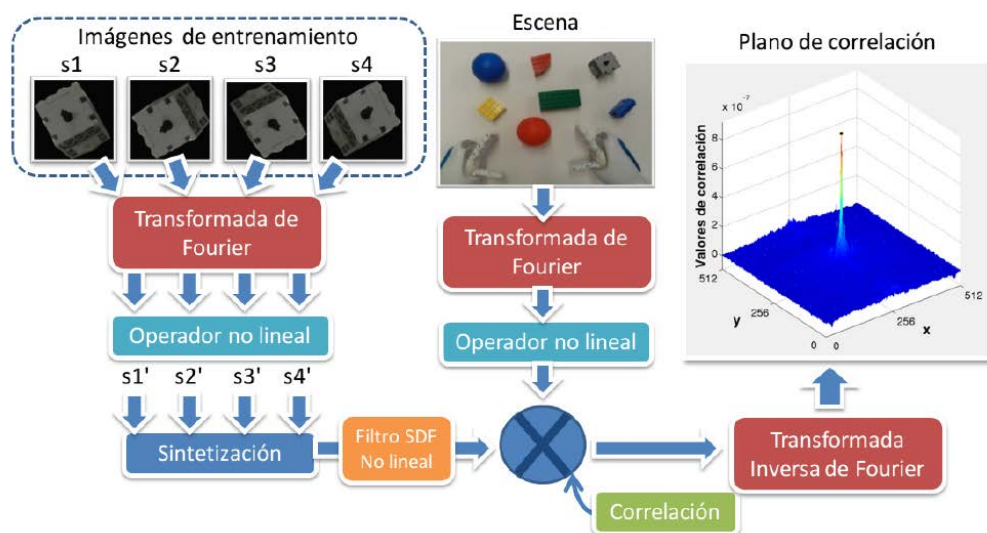


Figura 1 Metodología para la creación del filtro SDF y su correlación con una escena.

## Métricas de Desempeño Existentes y Métrica Propuesta

El plano de correlación obtenido durante la aplicación de un filtro es evaluado por cada una de las métricas que determinan una medida o valor. Este valor, puede ser comparado con otro valor de referencia para medir una similitud entre el objeto que se busca y la escena donde se busca el objeto. De tal forma, si la métrica de desempeño genera un valor similar a este valor de referencia, la respuesta del filtro es positiva para el reconocimiento de un objeto. Generalmente, en estas métricas se construye un umbral de aceptación con base en el valor de referencia.

Si el valor obtenido de la métrica se encuentra dentro de este umbral, se considera un desempeño satisfactorio del filtro y por lo tanto un reconocimiento correcto del objeto. De acuerdo con [Kumar, 2005], las siguientes son las métricas más utilizadas en el estado del arte para evaluar un filtro de correlación:

- Valor Máximo. la altura del pico de correlación es usada para determinar si una imagen es similar a la imagen de entrenamiento con la que se construyó el filtro. Mediante una función de búsqueda en el plano de correlación, se localiza el valor máximo del plano (i.e. altura del pico de correlación) y se compara con un valor o umbral previamente definido. Si el valor máximo es mayor o igual que el umbral, el objeto ha sido reconocido (Figura 2).

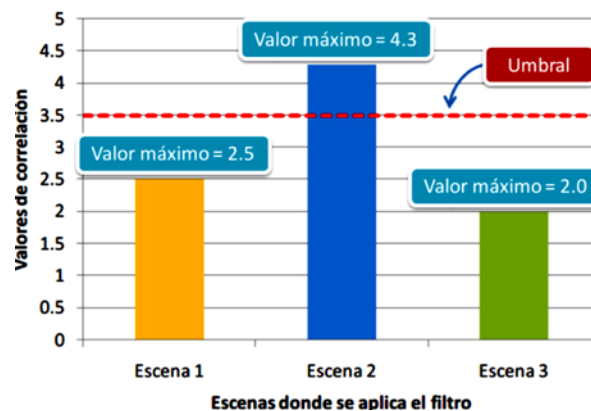


Figura 2 Métrica del valor máximo.

- Relación Pico a Lóbulo (Peak to Sidelobe Ratio, PSR). La métrica representa el contraste que existe entre el valor máximo del plano de correlación y el vecindario más cercano a dicho valor. Para ello se calcula la diferencia entre

el valor máximo y el promedio de dicho vecindario. Posteriormente se divide entre la desviación estándar. El vecindario se define como una región externa alrededor del valor máximo, excluyendo una región interna. El tamaño de las regiones interna y externa son calculados de acuerdo al tamaño de la imagen, ecuación 4.

$$PSR = \frac{valormáximo - \mu}{\sigma} \quad (4)$$

Donde  $\mu$  y  $\sigma$  son los valores de la media y desviación estándar de correlación dentro de la región alrededor del valor máximo, figura 3.

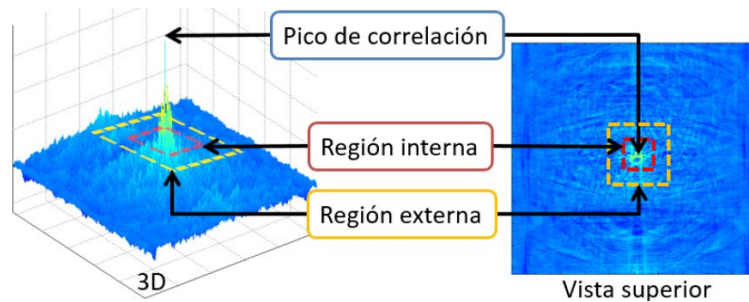


Figura 3 Métrica PSR.

- Relación Pico a Energía de Correlación (Peak to Correlation Energy o *PCE*). Se define como la relación entre la magnitud del cuadrado del valor máximo con respecto a la energía total del plano de correlación  $g(x, y)$ . La energía se define como la sumatoria de los valores absolutos al cuadrado de cada elemento del plano de correlación (Ecuación 5).

$$PCE = \frac{|valormáximo|^2}{\sum_x \sum_y |g(x, y)|^2} \quad (5)$$

Las métricas anteriores presentan dos desventajas. La primera es que los resultados obtenidos por cada métrica carecen de una estandarización y no pueden ser comparados entre sí, además de que los valores de sus respuestas cambian dependiendo del número de imágenes de entrenamiento. La segunda se da cuando la respuesta de una misma métrica presenta una diferencia muy grande entre un objeto reconocido y el mismo objeto con alguna variación en tamaño o rotación. Por



lo tanto, se propone una métrica de desempeño con una escala que determine el grado del reconocimiento del objeto en el plano de correlación y que pueda ser comparado fácilmente.

### La Métrica Relación Pico a Valores Vecinos (Peak to neighboring values o *PNV*)

La métrica *PNV* consiste en obtener el promedio de las distancias entre el valor máximo de correlación y otros valores, obtenidos dentro de un área de búsqueda en el plano de correlación. Estos valores corresponden a los valores más altos que se encuentran entre un área alrededor del valor máximo y un área alrededor del pico de correlación. La región alrededor del valor máximo se denominará región interna y la región alrededor del pico de correlación se denominará región externa. El área de búsqueda dependerá del tamaño del plano de correlación y de la respuesta que se espera del filtro, es decir qué tanta definición se espera en su pico de correlación. Una vez creada el área de búsqueda, se buscan los valores más altos (figura 4).

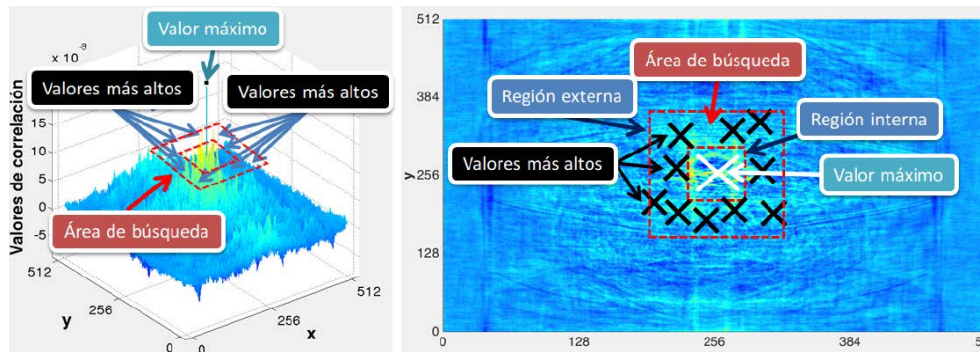


Figura 4 Métrica propuesta: *PNV*.

La cantidad de estos valores dependerá del tamaño del plano de correlación y de la precisión que se desea obtener. Una vez que se obtienen estos valores más altos, se obtiene el promedio de la diferencia entre las distancias del valor de correlación máximo y cada uno de dichos valores. Esto puede ser expresado con ecuación 6.

$$D = \frac{(v_{max} - v_1) + (v_{max} - v_2) + \dots + (v_{max} - v_n)}{n} \quad (6)$$

Donde  $v_{max}$  corresponde al valor máximo,  $v_1, v_2, \dots, v_n$  corresponden a la lista de los valores más altos y,  $n$  representa la cantidad de valores de la lista. A partir del cálculo de  $D$ , se obtiene la métrica  $PNV$  mediante la ecuación 7.

$$PNV = \frac{D \times 100}{v_{max}} \quad (7)$$

El valor resultante de la métrica  $PNV$  se encuentra definido en una escala de base 100, donde un valor de 0 corresponde al valor más bajo a obtener y un valor de 100 corresponde al valor más alto. Esto supone que una métrica  $PNV$  con valor de 100 significa que el filtro reconoce en su totalidad el objeto que se busca, sin importar los cambios que tuvo en la escena.

### **Robot Humanoide NAO**

NAO es un robot bípedo de 58 cm de altura y un peso de 4.3 kg. Tiene hasta 25 grados de libertad, que hacen referencia al número de movimientos independientes que puede realizar en sus articulaciones. El movimiento del robot NAO esta basado en motores de corriente continua y tiene una autonomía de hasta 60 minutos de uso continuo. El robot tiene un procesador Intel ATOM Z530 a 1,6 GHz de velocidad de procesamiento, con capacidad en memoria RAM de 1 GB y almacenamiento mediante una memoria flash de 2 GB, además de una ranura para tarjeta SDHC de 8 GB. Posee conectividad por medio de un puerto Ethernet y conexión Wi-Fi. La figura 5 muestra un diagrama general del robot, donde se observan los diversos sensores de los que puede obtener información de su ambiente: dos cámaras, cuatro micrófonos, un sonar, dos sensores infrarrojos, así como sensores táctiles y de presión. Las cámaras del NAO se encuentran en la parte frontal de su cabeza. La cámara superior es utilizada para ver lo que se encuentra frente al NAO, mientras que la cámara inferior es utilizada para ver el camino por dónde camina. El modelo en las cámaras del NAO es el MT9M114 y provee hasta una resolución de 1280x960 pixeles a 30 frames por segundo. En este trabajo de investigación se hace uso de la cámara superior para reconocer objetos por medio del filtro SDF.

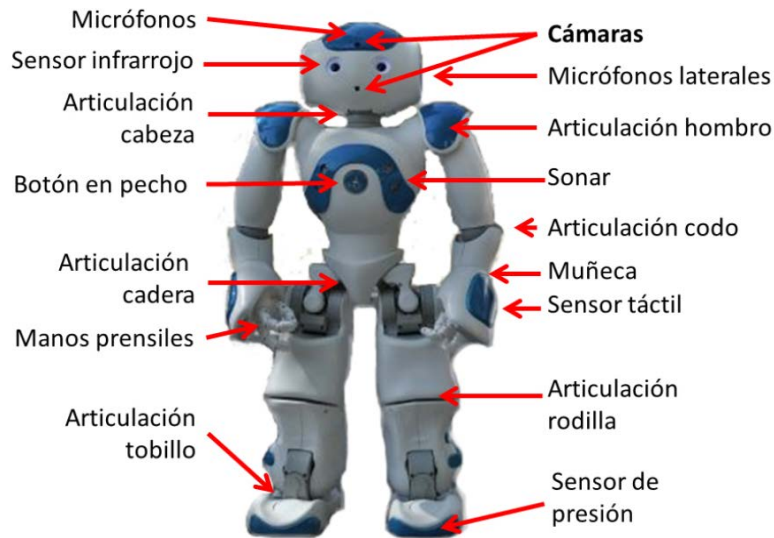


Figura 5 Diagrama del robot NAO

El software del robot se conoce como NAOqi. Para poder utilizar este software, es necesario el marco de programación NAOqi Framework, el cual permite escribir y ejecutar instrucciones en el robot NAO, así como hacer uso de los sensores que se encuentran en el robot. NAOqi Framework es multiplataforma y multilenguaje. Para trabajar en el NAOqi Framework se hace uso de la interfaz de programación de aplicaciones (NAOqi API), la cuál contiene el conjunto de subrutinas, funciones y procedimientos que ofrece NAOqi para el desarrollo de aplicaciones en el robot NAO. Se encuentra disponible para 8 lenguajes de programación, uno de ellos el lenguaje C++, que es el que se utilizó en este trabajo. Para crear los ejecutables que corren en el robot se utiliza QiBuild, que es un conjunto de librerías que permite crear el proyecto y área de trabajo para trabajar con el NAOqi Framework. Por defecto, qiBuild usa librerías provistas en el sistema operativo y está basado en el sistema de construcción CMake [Martin, 2008].

### Diseño de los Experimentos

El objetivo principal de estos experimentos fue evaluar el desempeño de la métrica PNV en el filtro SDF para el reconocimiento de objetos en el Robot NAO. Para esto, se creó un filtro para un objeto que aparece en las escenas tomadas con la cámara del robot NAO (ver objeto en la figura 6). Este filtro, fue comparados con

las escenas obtenidas del robot NAO donde se encontraba un objeto a reconocer. El objeto por reconocer presentó cambios en escala y rotación. Para evaluar el reconocimiento del objeto, se utilizaron las métricas que se mencionan en la sección anterior.

### **Pre-procesamiento de Imágenes**

El robot NAO obtuvo las imágenes de las escenas mediante su cámara superior frontal a una resolución tipo k4VGA de 1280 × 960 píxeles. Mediante la librería OpenCV se realizó una reducción de la imagen al formato más cercano de 640 × 480 píxeles, aminorando la pérdida de información mediante un método de interpolación. Uno de los requerimientos para trabajar con el filtro SDF es la utilización de imágenes del mismo tamaño par en sus renglones y columnas, por lo que la imagen se modificó a un tamaño 512×512 píxeles. Para lograr esto, se cortó la región sobrante a los costados derecha e izquierda y se agregó una región de color negro en la parte superior e inferior, mediante la función de replicado de borde en OpenCV. El formato de la imagen a color, fue necesario transformarlo a escala de grises, por lo que se aplicó la función `cvtColor` para obtener dicho formato. Para crear las imágenes con las que se construyó el filtro de cada objeto, se recuperaron las escenas guardadas en la memoria del robot y se editaron por medio del software de edición de imágenes GIMP (versión 2.8). La edición consistió en remover el fondo de la imagen dejando sólo el objeto que se deseaba reconocer y después sustituir el fondo por color negro. Una vez que se tuvieron las imágenes de entrenamiento, se aplicó el procedimiento descrito a continuación.

### **Construcción y Correlación del Filtro SDF**

Para realizar la implementación del filtro, es necesario desarrollar algoritmos para la construcción del filtro SDF y para la correlación del filtro SDF. Para llevar a cabo este procedimiento, cada imagen de entrenamiento se transforma al dominio frecuencial mediante la transformada de Fourier y se le aplica el operador de no linealidad Ley-k. Posteriormente, cada imagen de entrenamiento modificada se convierte a vector y se agrega a un arreglo de vectores que conforman la matriz

ordenada de forma lexicográfica. Después, se crea una copia de la matriz ordenada y se aplica la función de transpuesta del complejo conjugado. Las dos matrices se multiplican y al resultado de esta operación se le aplica la función inversa de la matriz. El resultado de la operación anterior se multiplica por el vector de coeficientes. La matriz resultante de esta operación corresponde al filtro SDF.

Una vez construido el filtro SDF, se procede a realizar la correlación. Se recupera la imagen de la escena de la memoria del robot. A continuación, se le aplica transformada de Fourier y el operador de no linealidad Ley-k. Después, se realiza la multiplicación elemento por elemento de la imagen resultante con el complejo conjugado del filtro SDF, es decir, la correlación. Al resultado de la multiplicación, se le transforma al dominio espacial mediante la transformada inversa de Fourier. Al final, se obtiene una matriz cuyo contenido es el plano de correlación final.

### **Experimentos con el Robot NAO**

Es necesaria una selección adecuada de imágenes de entrenamiento que tome en cuenta los cambios de escala y rotación en un objeto para construir un filtro SDF. Se construyó un filtro que fue evaluado en tres escenarios diferentes (Figura 6). Estos corresponden a una plataforma con fondo oscuro donde aparece sólo el objeto a reconocer en el escenario 1. El escenario 2 es similar, salvo porque se agregan otros objetos diferentes al que se quiere reconocer. El escenario 3 tiene un fondo con diversas imágenes que simulan una multitud de objetos que rodean al objeto a reconocer.



Figura 6 Objeto a reconocer en tres escenarios distintos para los experimentos.

Con estos tres escenarios se plantean dos experimentos que se describen a continuación:

- **Cambios de escala.** Para probar la respuesta del filtro SDF ante cambios en la escala de un objeto se propuso la prueba que se ilustra en la figura 7. Se colocó el robot en posición de descanso frente al escenario, poniendo el objeto a reconocer a una distancia inicial de 15 cm y se toma la imagen con la cámara del robot. Esta representa la primera escena. Posteriormente el objeto retrocede en distancias de 1 cm fotografiándolo cada vez para un total de 50 cm de distancia. Por tanto, se adquirieron 50 imágenes que fueron denominadas las 50 escenas en las que ocurrió el experimento. Se utilizó el prefijo e para denominar las escenas. De estas escenas, se utilizaron las imágenes e1, e3, e6, e9, e12, e15, e20, e25, e30, e40 y e50 para construir el filtro SDF para el experimento. A este filtro se le da denominación de *fPE*.

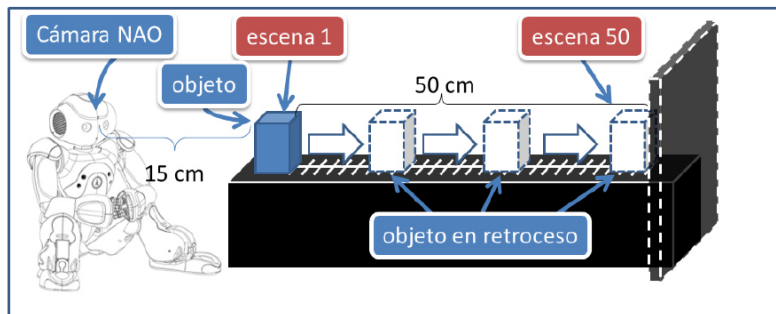


Figura 7 Esquema del experimento que evalúa cambios de traslación.

- **Cambios de rotación.** Para probar la respuesta del filtro SDF ante cambios en la rotación de un objeto se propuso la prueba que se ilustra en la figura 8. Se colocó el robot en posición de descanso frente al escenario, poniendo el objeto a reconocer sobre una plataforma giratoria que se encuentra a una distancia de 20 cm. La posición inicial del objeto se considera la primera escena y se le da el ángulo de 0 grados. Entonces el objeto se gira 5 grados a la derecha y se toma la siguiente imagen. Se repite esta operación cada 5 grados hasta llegar a los 90 grados con respecto a la cámara del robot. Ahora se regresa el objeto a la posición inicial y se gira 5 grados a la izquierda para

tomar una nueva imagen. Se repite esta operación hasta llegar a los 90 grados con respecto a la cámara del robot. En total se tomaron 37 escenas. Aquellas donde el objeto giró hacia la derecha se denominan con el prefijo *e* y aquellas donde el objeto giró a la izquierda se les agrega el prefijo *-e*. Del total de las escenas, se tomaron *-e45*, *-e25*, *e0*, *e25* y *e45* para construir el filtro SDF para el experimento. A este filtro se le da la denominación de *fPR*.

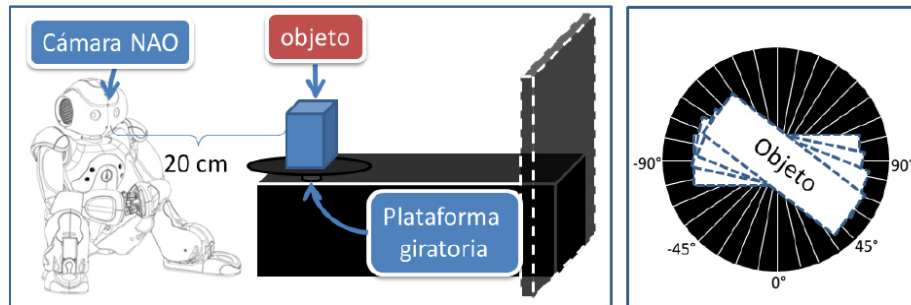


Figura 8 Esquema del experimento que evalúa cambios de rotación.

### 3. Resultados

A continuación, se presentan los resultados de los experimentos descritos en la sección anterior. La figura 9 muestra el resultado de aplicar el filtro *fPE* a las 50 escenas obtenidas para el experimento de cambios en la escala. Los resultados del escenario 1 se muestran en azul, del escenario 2 en naranja y del escenario 3 en rosa. Cada gráfica representa una de las 4 métricas presentadas anteriormente. Las métricas de desempeño Valor Máximo, *PSR* y *PCE* arrojaron resultados similares. La principal diferencia que se observa, fue una mayor repetición de valores altos en las escenas que se usaron como imágenes de entrenamiento (*e1*, *e3*, *e6*, *e9*, *e12*, *e15*, *e20*, *e25*, *e30*, *e40* y *e50*). Sin embargo, en las escenas que no lo fueron obtuvieron puntuaciones bajas. Se puede apreciar cómo la puntuación baja de manera abrupta en las demás escenas. Por otro lado, la métrica *PNV* generó valores altos, inclusive en las escenas que no se usaron como imágenes de entrenamiento. La puntuación en estas escenas baja, pero de manera suave. Nótese también que los resultados muestran las mismas tendencias para los 3 escenarios, aunque con valores ligeramente más bajos para los escenarios 2 y 3.

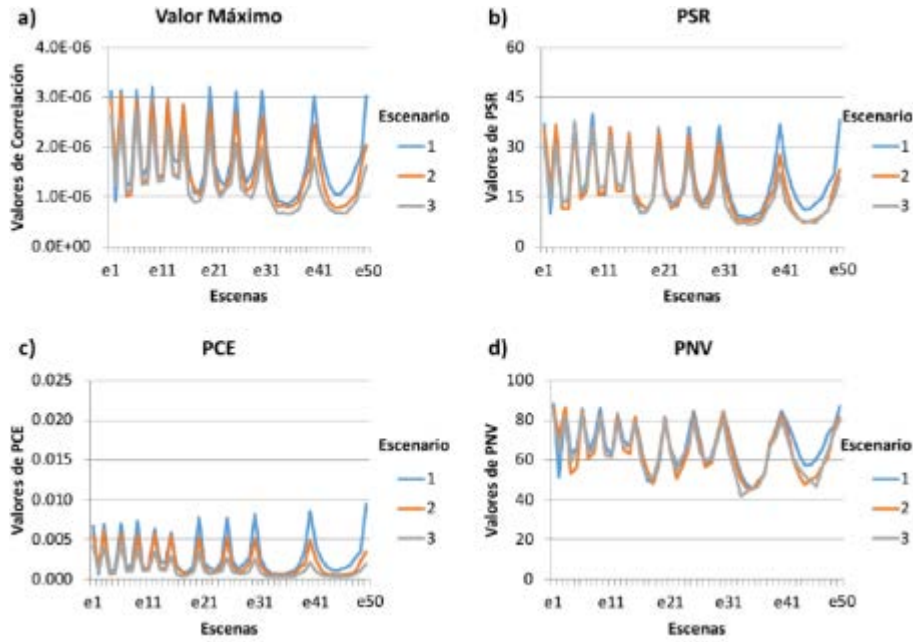


Figura 9 Resultados del experimento con cambios en la escala.

La figura 10 muestra el desempeño del filtro fPR bajo las 4 métricas mencionadas en la prueba de cambio de rotación. Se puede apreciar un comportamiento similar al del experimento anterior.

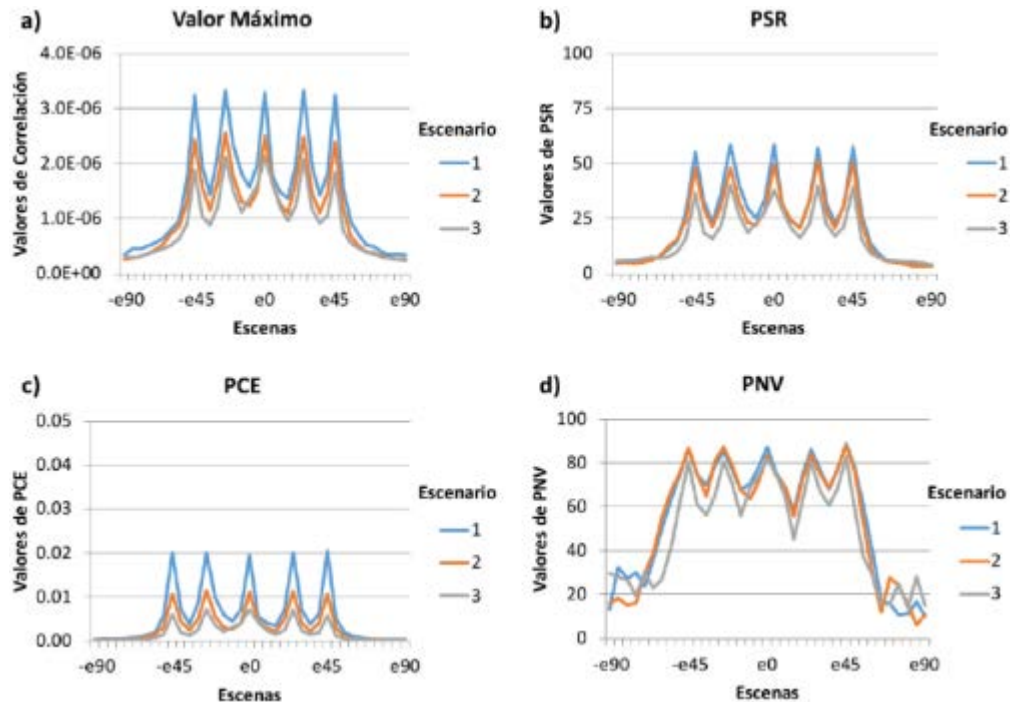


Figura 10 Resultados del experimento con cambios en la rotación.



Para las tres métricas el estado del arte, se obtiene un valor alto en las escenas utilizadas para la construcción del filtro, mientras que en las otras escenas se obtienen resultados notablemente más bajos. En el caso de la métrica *PNV*, note el comportamiento que se ha observado en el experimento anterior de una disminución menos abrupta. En la figura 10 se observa también otro comportamiento. Los valores más altos se concentraron en un área que abarcaba las escenas -e45 hasta la escena e45. Esto se debe a que sólo se eligieron escenas para construir el filtro dentro de este rango. Observe que las métricas Valor Máximo, *PSR* y *PCE* arrojaron valores muy bajos en las escenas que están fuera de dicho rango. Por otro lado, la métrica *PNV* muestra una caída de valores mucho menos abrupta que las otras métricas en este rango. Como se observa en la figura 10, los resultados obtenidos de la escena -e60 hasta la escena e60 fueron arriba de 50 en la métrica *PNV*.

#### **4. Discusión**

Como puede apreciarse en la sección de resultados, la métrica propuesta determina de una manera más clara la presencia de un objeto que se quiere reconocer ante cambios en la escala y rotación de dicho objeto. Las métricas de desempeño Valor Máximo, *PSR* y *PCE* presentan una rápida disminución de valores en las escenas que no fueron utilizadas para construir el filtro. Sin embargo, la métrica *PNV* generó una lenta y moderada disminución en estas escenas.

Otro punto a discutir consiste en la selección de las imágenes para construir un filtro. En el caso de los objetos que rotan producen un mayor cambio en su apariencia, a comparación de los objetos que se alejan (cambios de escala). Es decir, un objeto que rota 45 grados tendrá una menor similitud en su apariencia con el objeto antes de rotar. Por lo tanto, además de seleccionar las imágenes de entrenamiento que mejor representen tal cambio, se deberá de reducir el rango de rotación entre estas.

#### **5. Conclusiones**

Las principales contribuciones de este trabajo son la implementación del filtro SDF en un robot humanoide y la propuesta de una métrica de desempeño que mejoró la evaluación del filtro, denominada Relación Pico a Valores Vecinos (Peak

to Neighboring Values, *PNV* por sus siglas en inglés). La evaluación del desempeño del filtro se realizó con la métrica propuesta *PNV* y las métricas de desempeño convencionales: Valor Máximo, *PSR* y *PCE*.

Por medio de la métrica *PNV* se mejoró notablemente la medición del desempeño del filtro, generando mejores resultados que las métricas convencionales, específicamente en los objetos que tienen variaciones en su apariencia, como cambios de escala y de rotación. Calificaciones altas en el desempeño brindan una mayor seguridad para determinar que el objeto ha sido reconocido. Además, descarta el reconocimiento de un objeto diferente al que se busca reconocer, pues este obtendría una calificación baja. Además, los experimentos demostraron que el valor que genera la métrica *PNV* no depende del objeto que se busca reconocer, ni del conjunto de objetos que aparecen en la imagen donde se busca.

El siguiente paso en este trabajo de investigación consiste en evaluar el comportamiento de esta nueva métrica en escenarios no controlados, así como su implementación en video. Esto es posible ya que una vez construido el filtro para reconocer un objeto, la imagen del video se convoluciona con el filtro, obteniendo la respuesta en un tiempo rápido incluso para el poder de procesamiento del robot NAO. Esto da pie a generar pruebas dinámicas donde el robot se mueva en un entorno en tiempo real.

## **6. Bibliografía y Referencias**

- [1] Aguilar-González, P. M., Kober, V., & Díaz-Ramírez, V. H. (2014). Adaptive composite filters for pattern recognition in nonoverlapping scenes using noisy training images. *Pattern Recognition Letters*, 41, 83-92.
- [2] Álvarez-Borrego, J., & Castro-Longoria, E. (2003). Discrimination between *Acartia* (Copepoda: Calanoida) species using their diffraction pattern in a position, rotation invariant digital correlation. *Journal of Plankton Research*, 25(2), 229-233.
- [3] Andreopoulos, A., & Tsotsos, J. K. (2013). 50 years of object recognition: Directions forward. *Computer vision and image understanding*, 117(8), 827-891.

- [4] Casasent, D. (1984). Unified synthetic discriminant function computational formulation. *Applied Optics*, 23 (10), 1620-1627.
- [5] Han, S., Kim, M. J., Park, S., & Paik, J. (2017). Fast Vehicle Detection using Correlation Filters. *IEIE Transactions on Smart Processing & Computing*, 6(5), 309-316.
- [6] Hedgerow, J. (2004). *Manual de Técnica Fotográfica* (Tercera ed.). H. Blume
- [7] Horner, J. L., y Gianino, P. D. (1984). Phase-only matched filtering. *Applied Optics*, 23 (6), 812-816.
- [8] Javidi, B., Wang, W., y Zhang, G. (1997). Composite Fourier-plane nonlinear filter for distortion-invariant pattern recognition. *Optical Engineering*, 36 (10).
- [9] Kumar, B. V. K. V., Mahalanobis, A., y Juday, R. D. (2005). *Correlation Pattern Recognition* (Primera ed.). Cambridge University Press.
- [10] Lugt, A. V. (1964). Signal Detection by Complex Spatial Filtering. *Information Theory, IEEE Transactions on*, 10 (2), 139-145.
- [11] Martínez-Pérez, F. E., González-Fraga, J. A., & Tentori, M. (2010). Automatic activity estimation based on object behaviour signature. In *Applications of Digital Image Processing XXXIII* (Vol. 7798, p. 77980E). International Society for Optics and Photonics.
- [12] Ramírez, A. A. P. (2009). *Desarrollo de una aplicación de correlación digital que haga uso de filtros lineales y no lineales compuestos* (Tesis de Master no publicada). Universidad Autónoma de Baja California, Ensenada Baja California.
- [13] Savvides, M., Vijaya, B., y Khosla, V. (2002). Face Verification using correlation filters. *Proc. of the 3rd IEEE Automatic Identification Advanced Technologies*, 56-61 (62)
- [14] Tsotsos, J. K., Liu, Y., Martínez-Trujillo, J. C., Pomplun, M., Simine, E., y Zhou, K. (2005). Attending to visual motion. *Computer Vision and Image Understanding* (100), 3-40.
- [15] Zhao, D., Fu, H., Xiao, L., Wu, T., & Dai, B. (2018). Multi-object tracking with correlation filter for autonomous vehicle. *Sensors*, 18(7), 2004.