

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

ESCUELA DE POSTGRADO



PONTIFICIA
UNIVERSIDAD
CATÓLICA
DEL PERÚ

CARACTERIZACION Y RECONOCIMIENTO DE OBJETOS MEDIANTE ALGORITMOS DE VISION COMPUTACIONAL PARA LA INTERACCION DE UN ROBOT CON SU ENTORNO

Tesis para optar el grado de Magíster en Informática
con Mención en Ciencias de la Computación

AUTOR

LUIS DAVID ROBLES PIZARRO

ASESOR

DR. CÉSAR ARMANDO BELTRÁN CASTAÑÓN

MIEMBROS DEL JURADO

DR. HÉCTOR ANDRÉS MELGAR SASIETA

DR. CÉSAR ARMANDO BELTRÁN CASTAÑÓN

DR. HUGO ALATRISTA SALAS

LIMA - PERÚ

Dedicatoria

*A mis padres, Rocío y Hugo, por sus
consejos y apoyo incondicional*



Agradecimientos

Al Doctor César Beltrán por su apoyo y guía durante la realización de la presente tesis.

Al Magíster Francisco Cuéllar por facilitarme las instalaciones de CETAM.

A Alexander López por su orientación en el manejo del robot NAO.

A Andrea Cáceres por su compañía y buen humor durante los momentos de estrés.

A mis padres y a mi hermano por empujarme a ser mejor.

Tabla de contenido

RESUMEN	8
1. GENERALIDADES	10
1.1. Definición de la problemática.....	10
1.2. Estado del Arte	11
1.2.1. Trabajos de Reconocimiento de Objetos.....	12
1.2.2. Trabajos de Reconocimiento de Objetos aplicado a Robots Humanoides	12
1.3. Marco Conceptual.....	13
1.3.1. Autonomía en la robótica	13
1.3.2. Reconocimiento de Objetos.....	14
1.3.3. Método <i>SIFT – Scale Invariant Feature Transform</i>	16
1.3.4. Método <i>SURF – Speeded-up Robust Feature</i>	18
1.3.5. Procesamiento de imágenes digitales	19
1.3.6. Filtros de suavizado – <i>Smoothing Filters</i>	20
1.3.7. Filtros de nitidez – <i>Sharpening Filters</i>	21
1.3.8. Sistema robótico Nao Aldebaran	21
2. DESCRIPCIÓN Y SUSTENTACIÓN DE LA SOLUCIÓN	23
2.1. Objetivo general y objetivos específicos	23
2.2. Resultados esperados	24
2.3. Marco Lógico de la propuesta.....	24
2.4. Alcances de la solución propuesta	26
3. CARACTERIZACIÓN DE OBJETOS CON MÉTODOS DE VISIÓN COMPUTACIONAL	27
3.1. Aplicación del método <i>SIFT</i> en un conjunto de imágenes	27
3.2. Aplicación del método <i>SURF</i> en un conjunto de imágenes.....	30
3.3. Comparación de los resultados obtenidos por los métodos <i>SIFT</i> y <i>SURF</i>	32
3.4. Selección del dominio de objetos que construirán nuestro modelo predictor.	36
4. DISEÑO DE LA SOLUCIÓN	40
4.1. Flujo general del sistema.....	40
4.2. Diseño del modelo predictor utilizando un clasificador bayesiano para predecir la categoría de un objeto	41
4.2.1. Consideraciones previas	41
4.2.2. Modelo predictor <i>SIFT</i>	42

4.2.3.	Modelo predictor <i>SURF</i>	47
4.2.4.	Comparación de los modelos predictores <i>SIFT</i> y <i>SURF</i>	51
5.	CONCLUSIONES	55
6.	TRABAJO FUTUROS	57
7.	REFERENCIAS BIBLIOGRÁFICAS	58
8.	ANEXOS	60
8.1.	Anexo A: Aplicación del método SIFT y SURF para detectar puntos característicos en un dominio de flores.	61
8.2.	Anexo B: Aplicación del método SIFT y SURF para detectar puntos característicos en un dominio de flores ante cambios de iluminación	63
8.3.	Anexo C: Aplicación del método SIFT y SURF para detectar puntos característicos en un dominio de imágenes ante cambios de escala	65
8.4.	Anexo D: Aplicación del método SIFT para la clasificación de dos grupos de objetos por medio del Clasificador Bayesiano	66
8.5.	Anexo G: Captura de imágenes para el proceso de predicción	68

Índice de Figuras

Figura 1.1 Visualización del conjunto de píxeles vecinos alrededor de un pixel X para encontrar los máximos y mínimos de la Diferencia de Gaussianos. [En línea] Disponible: http://aishack.in/static/img/tut/sift-maxima-idea.jpg	17
Figura 1.2: Visualización de los puntos de interés detectados en una imagen	17
Figura 1.3 Aproximación del Laplaciano Gaussiano utilizando filtros de caja.	19
Figura 1.4: Resultado de aplicar un filtro de suavizado a una imagen.	20
Figura 1.5: Resultado de aplicar un filtro de nitidez a una imagen.....	21
Figura 1.6: Grados de libertad del robot NAO. [En línea] Disponible: http://taaone-sti2d.fr/STI2D/ETT/Les_activites/5_7_NAO/files/documents/SiteAldebaran/donnees-techniques/ALMotionJointDefinition.jpg	22
Figura 1.7: Robot nao desarrollado por la Empresa Aldebaran Robotics.....	22
Figura 3.1: A la izquierda, la imagen original. A la derecha, imagen con los puntos de interés del método SIFT	27
Figura 3.2: Cálculo de la gradiente alrededor de un Keypoint	28
Figura 3.3 A la izquierda la imagen con baja iluminación. Al medio la imagen original. A la izquierda la imagen con alta iluminación. [Desarrollado por el autor].....	29
Figura 3.4: Descriptor de un punto de interés utilizando el método SURF	30
Figura 3.5: A la izquierda la imagen original. A la derecha imagen con los puntos de interés del SURF [Desarrollado por el autor]	31
Figura 3.6: Cantidad de puntos de interés encontrados en una muestra de 20 flores utilizando método <i>SIFT</i> y <i>SURF</i>	33
Figura 3.7 Puntos de interés encontrados en una muestra de 10 imágenes utilizando métodos <i>SIFT</i> y <i>SURF</i>	34
Figura 3.8 Tiempos requeridos para encontrar los puntos de interés en imágenes oscurecidas utilizando métodos <i>SIFT</i> y <i>SURF</i>	34
Figura 3.9 Puntos de interés encontrados en una muestra de 10 imágenes utilizando métodos <i>SIFT</i> y <i>SURF</i>	35
Figura 3.10 Tiempos requeridos para encontrar los puntos de interés en imágenes oscurecidas utilizando métodos <i>SIFT</i> y <i>SURF</i>	36
Figura 3.11 Conjunto de imágenes de mandarinas	36
Figura 3.12 Conjunto de imágenes de granadillas.....	37
Figura 3.13 Conjunto de imágenes de manzanas	37
Figura 3.14 (Izquierda) Imágenes originales. (Centro) Imágenes con puntos SIFT. (Derecha) Imágenes con puntos SURF.....	37
Figura 3.15 Comparación de puntos de interés <i>SIFT</i> en imágenes de objetos del dominio elegido	39
Figura 3.16 Comparación de puntos de interés <i>SURF</i> en imágenes del dominio elegido	39
Figura 4.1 Flujo del control sistema y el sistema robótico.....	40
Figura 4.2 Selección aleatoria de puntos de interés <i>SIFT</i> del conjunto total de puntos de interés originales	42
Figura 4.3 Gráficos de barras con las cantidades de puntos de interés encontrados con el método <i>SIFT</i>	44
Figura 4.4 Gráficos de barras con las cantidades de puntos de interés encontrados con el método <i>SURF</i>	48
Figura 4.5 Gráfico del porcentaje de precisión al predecir Mandarinas.....	52
Figura 4.6 Gráfico del porcentaje de exactitud al predecir Mandarinas.....	52
Figura 4.7 Gráfico del porcentaje de precisión al predecir Granadillas	53
Figura 4.8 Gráfico del porcentaje de exactitud al predecir Granadillas	53
Figura 4.9 Gráfico del porcentaje de precisión al predecir Manzanas.....	53
Figura 4.10 Gráfico del porcentaje de exactitud al predecir Manzanas.....	54
Figura 8.1: Extracto de imágenes de la base de datos	61

Figura 8.2: Resultado de las imágenes para SIFT62

Figura 8.3: Resultado de las imágenes para SURF62

Figura 8.4: Cambios de iluminación con SIFT63

Figura 8.5: Cambios de iluminación con SURF64

Figura 8.6: Imágenes de un bus a 3 resoluciones sometido al cálculo SIFT65

Figura 8.7 Imágenes de un bus a 3 resoluciones sometido al cálculo SURF.....65

Figura 8.8: Ejemplo de clasificación correcta de la imagen de un auto.....66

Figura 8.9: Ejemplo de clasificación incorrecta de la imagen de una flor67



Resumen

En el campo de la robótica, se han desarrollado distintos algoritmos y métodos con el objetivo de mejorar la interacción de los robots con las personas y con su entorno de trabajo en tiempo real; es así, como el sistema reacciona y evoluciona constantemente ante cambios que podrían ocurrir durante su funcionamiento. Para alcanzar los objetivos mencionados, una de las habilidades que se le confiere a la máquina es la capacidad de detectar, registrar y reconocer objetos.

La presente tesis es un trabajo de investigación aplicada que tiene como objetivo desarrollar un procedimiento que permita a un sistema robótico reconocer y detectar objetos en tiempo real dentro de un entorno controlado; para ello, nos enfocamos en utilizar dos métodos conocidos de reconocimientos de objetos (métodos *SIFT* y *SURF*) con los cuales categorizaremos un objeto de un dominio predefinido y comparamos los resultados obtenidos. Se eligieron el método *SIFT* y el método *SURF* por la similitud en los pasos que siguen para obtener la información de un objeto; cabe resaltar que el método *SURF* es un método alternativo al *SIFT*.

Los resultados finales mostraron una mejor predicción en la categorización utilizando el método *SIFT*, pero ésta requería de mayor tiempo para extraer los puntos característicos de los objetos. Por otro lado, el método *SURF* generaba más puntos característicos de los objetos y en mejor tiempo. La extracción de puntos de interés se analizó en tiempo real; mientras, que la etapa de categorización no consideró este parámetro, sino la cantidad de puntos de interés necesarios para predecir con exactitud la categoría de un objeto.

En el capítulo I, presentamos el planteamiento de la investigación, estado del arte y marco teórico

En el capítulo II, detallaremos la descripción y la sustentación de la solución propuesta.

En el capítulo III, mostramos los resultados de aplicar los métodos a un conjunto de objetos y sus diferencias respectivas.

En el capítulo IV, se desarrolla los modelos del clasificador para determinar la categoría de un objeto.

En el capítulo V y VI, se encuentran las conclusiones del proyecto desarrollado y propuestas para trabajos futuros, respectivamente.

Los Anexos contienen información de las pruebas realizadas utilizando los métodos SIFT y SURF.



1. Generalidades

En el presente capítulo, se expone el contexto en que se enmarca el problema; así como, una revisión de los conceptos claves necesarios para el entendimiento de ésta y de la solución propuesta.

1.1. Definición de la problemática

El profesor Qiang Ji [2], define la Visión Artificial como una disciplina que estudia el proceso de construir, interpretar y entender una escena tridimensional a partir de imágenes bidimensionales; asimismo, clasifica procesos de la Visión Artificial en 3 niveles: un nivel bajo, que consiste en el procesamiento de una imagen para obtener características; un nivel intermedio, que involucra el reconocimiento de objetos, análisis de movimiento, y reconstrucción 3D empleando las características del nivel bajo; y finalmente, un nivel alto, que abarca la interpretación de la información obtenida en el nivel intermedio. Asimismo, según Thomas Huang [1], menciona que la Visión Artificial comprende dos objetivos principales: la elaboración de modelos capaces de simular el sistema visual humano; y la construcción de sistemas autónomos para que realicen tareas similares al del sistema visual humano.

En el área de la robótica, dotar a la máquina de capacidades sensoriales similares a las de los humanos, son motivo de investigaciones que tienen por objetivo facilitar las tareas de reconocimiento de objetos, navegación y manipulación de herramientas que realizan los robots[5].

Generalmente el proceso de reconocimiento de objetos está basado en la solución de dos grandes problemas: la percepción de los objetos y la identificación del objeto observado [4].

Con respecto al primer problema, las características que permiten percibir un objeto se presentan en los cambios de la gradiente de los puntos o píxeles en una imagen digital, éstos suelen encontrarse en los bordes, puntos de interés (como esquinas) y regiones de puntos de interés. Entre los algoritmos que permiten extraer estas características tenemos el SIFT, propuesto por David Lowe [4] y el SURF [5], una versión mejorada del SIFT [5].

Para el segundo problema, la identificación de un objeto se realiza concatenando en dos imágenes las características obtenidas durante la etapa de percepción; a pesar de la efectividad de este método, generalmente se presentan algunas limitantes como:

1. Las variaciones de rotación de un objeto puede dar como resultandos características nuevas y distintas; esto genera que el mismo objeto en dos imágenes no sea concatenado o identificado correctamente.
2. El objeto a identificarse debe poseer características idénticas al objeto original (como forma, tamaño o textura); es decir, la categorización de objetos es específica a un elemento y no a un conjunto. Por ejemplo, dos árboles pueden ser distintos visualmente, pero son agruparlos bajo la misma categoría de “árbol”.

En el presente proyecto, se desarrolla una propuesta para identificar un objeto basado en el análisis de imágenes digitales. Primeramente, se adquiere la imagen mediante una cámara, luego el proceso de detección del objeto se realiza con técnicas de procesamiento de imágenes basadas en binarización. Para la extracción de características se utilizaron los algoritmos SIFT y SURF que tienen las características de presentar un nivel de robustez a varianzas de rotación, iluminación y escala. Finalmente, se realiza la predicción del tipo de objeto de un dominio mediante aprendizaje supervisado, el cual utiliza un conjunto de objetos previamente clasificados a los cuales se les aplicó los anteriores procesos de detección y extracción de características, los cuales sirvieron para entrenar el clasificador bayesiano.

1.2. Estado del Arte

A continuación, se presenta algunos de los últimos proyectos referentes al reconocimiento de objetos en robots humanoides.

1.2.1. Trabajos de Reconocimiento de Objetos

Liu, M., Li, X., Dezert, J., & Luo, C [9] proponen un método genérico de reconocimiento de objetos utilizando la fusión de múltiples descriptores *SIFT*. El método consta de tres etapas: extracción y representación de las características, fusión de características, y diseño del clasificador. En la primera etapa, extraen las características de objetos utilizando una combinación de *2D SIFT* y *3D SIFT*; luego utilizando el método *Bag of Words* los descriptores son agrupados en una cantidad de especificada de grupos. En la segunda etapa, se diseñan clasificadores *SVM* entrenándolos con los descriptores agrupados; estos clasificadores lineales pueden predecir dos categorías; por ello, para obtener una clasificación de varias categorías se entrena 1vs1 *SVMs*. En la tercera etapa, la estrategia de fusión se realiza de dos formas; una es entrenar los clasificadores y luego realizar el reconocimiento de objetos; mientras que la otra, consiste en entrenar los *SVM* con específicos descriptores para luego intentar combinaciones como peso promedio y la regla de *Murphy*; de este modo, predice la categoría. Los resultados del método propuesto demostraron mejor rendimiento frente a descriptores *PFH (Point Feature Histogram)*, *2D SIFT*, *3D SIFT*, *Murphy*.

Bo, L., & Whangbo, T. [10] propone un descriptor de características que combina los puntos de interés del *SIFT* y la información de los *Color Moments*. El descriptor provee la información de tres *Color Moments*: el promedio, la desviación estándar y la oblicuidad. Los resultados del método demostraron un mejor rendimiento al reconocer objetos. No obstante, el método también presenta algunas limitaciones; por ejemplo, la precisión aumenta, pero la exactitud es menor; y las pruebas solo extraen información del espacio RGB.

1.2.2. Trabajos de Reconocimiento de Objetos aplicado a Robots Humanoides

Vignesh Ramanathan [11] desarrolló una técnica para la categorización de objetos utilizando modelos *Bag of Words*, y luego la implementó un robot humanoide Nao. La técnica permitía que el robot categorizara un objeto que sostenía con su mano; para ello, realizaba una serie de capturas de imágenes utilizando una cámara externa y el movimiento de la mano del Nao para tener varios ángulos de visión.

El esquema propuesto extrae las características de las imágenes utilizando el método, y luego las agrupaba a través de la técnica *K-means*. Finalmente, la imagen se representa en forma de vectores de frecuencia y son utilizadas para entrenar a los clasificadores. La eliminación de características del fondo utiliza dos imágenes: una sin

la mano del robot y otra con la mano del robot, y se realiza la reducción de ambas para obtener los puntos de interés del objeto.

Du, Mingfang [12] propone el método *Fast SURF (SSURF)* para aplicarlo en la identificación visual de un robot. El método propuesto modifica el método SURF para que en lugar de buscar características alrededor de todo el vecindario, lo hace solo en un sector del vecindario. Los experimentos demostraron que el método SSURF para la extracción de características y el método de asociación tienen robustez y buen rendimiento.

1.3. Marco Conceptual

A continuación, se presentan conceptos que se utilizarán en el presente trabajo; por ejemplo, la autonomía de los robots, reconocimiento de objetos, y los métodos *SIFT* y *SURF* utilizados en la visión artificial para detectar y describir características en una imagen.

1.3.1. Autonomía en la robótica

Lozano-Perez, T. [11] define algunas propiedades que tienen los vehículos autónomos; tales como, capacidad para detectar las características de sus alrededores; capacidad para interpretar la información recibida para mejorar el conocimiento sobre su posición y la estructura del ambiente; y finalmente, capacidad para planear una ruta hacia un objetivo esquivando obstáculos que pudiesen presentarse.

Tim Smithers [12] menciona, en su libro sobre autonomía de los robots, la importancia de crear sistemas que no solo dependan de controles de retroalimentación negativa (Teoría de Control), sino que además interactúen con su ambiente. No obstante, existen dificultades en el diseño del comportamiento del sistema, debido a que es imposible predecir todas las situaciones que un robot enfrentará. Si somos capaces de especificar y controlar el ambiente del robot, es posible diseñar sistemas robustos que puedan trabajar; algunos ejemplos son los robots de fábricas, los robots móviles y los brazos robóticos. Con esto obtendríamos ambientes correctamente estructurados y libres de personas. Por otro lado, si deseamos que los robots operen en hospitales, museos, calles u otros lugares públicos, entonces deberían ser capaces de enfrentar mayores variaciones, cambios dinámicos y situaciones impredecibles.

En el libro “*Design, control, and applications of autonomous mobile robots*” [13], los autores consideran a los robots como máquinas que solamente operan en ambientes controlados, pero encuentran el mismo problema de no poder programar todas las posibles acciones que pudieran.

Las aplicaciones concretas de la autonomía en los robots son muy variadas. E. Freund, J. RoBmann [14] presentan el “Principio de jerarquía de coordinador”, el concepto básico del sistema para el desarrollo de sistemas de control complejo en IRF; para luego aplicarlo a un sistema multi robot en un módulo de laboratorio espacial. Gryaznov, N., & Lopota [15] mencionan el esfuerzo detrás del planeamiento de una ruta para un robot móvil, estos incluyen el mapeo del área visible, el reconocimiento de objetos y el tipo de suelo; así como el costo de calcular el proceso 3D en tiempo real. Utilizar la detección de contornos de los límites es una de las técnicas que les permite simplificar la descomposición de objetos y reducir el tiempo computacional.

1.3.2. Reconocimiento de Objetos

Carlini [16], define el reconocimiento de objetos como la capacidad de percibir propiedades físicas de un objeto y aplicar atributos que le otorgan significado o interpretación, basado en: el entendimiento de su uso, las experiencias previas con el objeto, y el cómo se relaciona con otros objetos.

Según Grauman, K., & Leibe, B. [3], el desarrollo de descriptores y detectores de las características locales en una imagen son importantes en el reconocimiento de objetos. Por ejemplo, a partir de un objeto rígido se pueden tener dos vistas del mismo modelo en ángulos diferentes, luego detectar sus características locales, y finalmente, relacionar dichas características, indicando de este modo que se trata del mismo objeto. Los pasos generales de este proceso se resumen en:

1. Extraer las características locales de una imagen de entrenamiento y otra de prueba.
2. Relacionas las características de una imagen con otra para encontrar correspondencias entre ellas.
3. Verificar si las correspondencias tienen consistencia con su configuración geométrica.

Los Métodos para el reconocimiento de objetos en la Visión Artificial son muy variados debido a los distintos problemas que conlleva el proceso de visión. Algunas de estas

dificultades pueden resumirse en: Detección (presencia de un objeto ante un estímulo en particular), localización (detección y ubicación precisa del objeto), reconocimiento (localización de todos los objetos presentes frente a un estímulo) y entendimiento (reconocimiento con el rol de estímulo en el contexto de una escena). Si nos concentramos en el problema de reconocimiento, la dificultad se genera en identificar los objetos presentes en una imagen y otorgar información de los mismos [14]. Algunos de los procedimientos usados típicamente en un sistema de reconocimiento de objetos son: extracción de características, agrupación de características, generación de hipótesis y verificación de objetos.

Goad [15] publicó uno de los primeros métodos para la generación automática de programas que reconocieran objetos. Él compilaba los bordes de un objeto en el campo de visión en un árbol de interpretación. Bolles y Haroud [16] utilizaron también modelos tridimensionales de varios objetos para encontrarlos en el rango de datos; sin embargo, éstos dependían mucho de modelos basados en líneas y bordes que no siempre son adecuados para objetos que se diferencian por características como el color y la textura.

Las técnicas de organización perceptiva fueron desarrolladas en un intento de modelar la capacidad del sistema visual humano para detectar propiedades no accidentales de características de bajo nivel y agruparlas con el fin de construir representaciones de objetos más compactas. Cuando extraemos características de bajo nivel como bordes y líneas, se busca correspondencias entre ellas para compararlas con un modelo de mayor complejidad. El mayor éxito de la organización perceptiva es la detección de grupos estables en las imágenes; algunos de estos factores comunes que pueden agruparse se resumen en: Similitud (características similares), Proximidad (características cercana), Continuidad (características que conducen a curvas continuas), Paralelismo (curvas y características paralelas), Simetría (curvas que crean grupos simétricos) y Regiones Comunes (características que yacen dentro de una misma región cerrada).

Por otro lado, existen evidencias de que el cerebro utiliza varias medidas – tales como el cierre total, mediante la medición del espacio entre los contornos percibidos previamente - como pasos intermedios en la formación de formas y representación. Lowe utiliza heurísticas para “unir” líneas y bordes que probablemente pertenecen juntos y por lo tanto, disminuye la complejidad global del modelo que se ajusta al proceso.

1.3.3. Método *SIFT* – *Scale Invariant Feature Transform*

Los sistemas de reconocimiento de objetos eran efectivos en ambientes controlados, pero la viabilidad disminuía cuando el objeto percibido rotaba, cambiaba su escala o la iluminación variaba. Una alternativa para ubicar los objetos en una imagen era extraer las características invariantes; para ello se propusieron muchos métodos, incluyendo segmentación de líneas, agrupación de bordes y regiones.

David Lowe [8] propuso el método *SIFT*; un algoritmo que permite extraer puntos estables y de interés en una imagen, y que eran capaces de soportar alteraciones de condiciones tales como: cambios de escala, rotaciones y variación de iluminación. El conjunto de estos puntos de interés representan descripciones de características que son utilizadas para identificar un objeto en otras imágenes.

Según Andrea Vedaldi [17], los puntos de interés *SIFT* contienen información en una región circular de una imagen, y contiene cuatro parámetros importantes: el centro de coordenadas X e Y, la escala (radio de la región) y el ángulo de orientación expresado en radianes.

Utkarsh Sinha [18], explica la secuencia del cálculo de los puntos de interés *SIFT* en 6 etapas:

1. Construcción de un espacio de escala: Una imagen con varios detalles incluye mucha información que debe reducirse; para ello, se utiliza un conjunto de operaciones como: suavizado Gaussiano, generación de octavas de la imagen (reducir el tamaño en un octavo), y aplicación del suavizado en las octavas. El conjunto de nuevas imágenes son utilizadas en la siguiente etapa.
2. Aproximación del método *LoG*: Los puntos de interés *SIFT* se encuentran principalmente en los bordes y esquinas de una imagen, que a su vez pueden obtenerse tras aplicar el Laplaciano Gaussiano a la imagen suavizada; no obstante, este proceso es costoso computacionalmente. Lowe [8] demostró que el Laplaciano Gaussiano puede aproximarse a la Diferencia de Gaussianos de dos imágenes obtenidas a partir de la reducción de Octavas de una misma imagen. De este modo, se consigue un conjunto de puntos de interés iniciales invariantes a los cambios de escala.

3. Detección de los puntos de interés: Se obtienen los puntos de interés a partir de los máximos y mínimos de la Diferencia de Gaussianos. Lowe [8] propone buscarlos alrededor de los 26 vecinos de un pixel a partir de la segunda imagen de Octava, ya que es la primera imagen en la que los 26 vecinos existen.

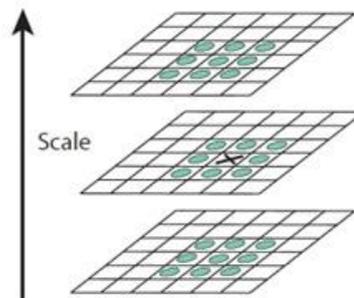


Figura 1.1 Visualización del conjunto de píxeles vecinos alrededor de un pixel X para encontrar los máximos y mínimos de la Diferencia de Gaussianos. [En línea]

Disponible: <http://aishack.in/static/img/tut/sift-maxima-idea.jpg>

4. Eliminación de puntos de interés con características de bajo contraste y ubicados en esquinas: Consta de dos posibles criterios de eliminación: El primero consiste en que si la magnitud de la intensidad de un pixel en el DoG es menor a determinado valor, es rechazado. El segundo calcula los valores de dos gradientes perpendiculares en un punto de interés, si las gradientes son pequeñas (región plana), o si una es grande y la otra pequeña (borde) son eliminadas.

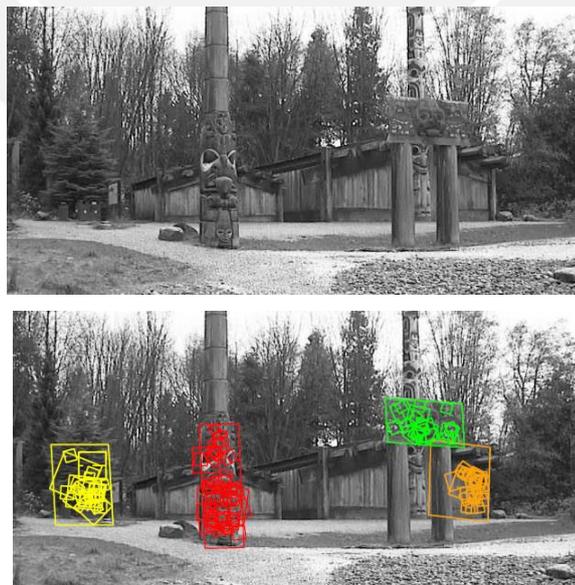


Figura 1.2: Visualización de los puntos de interés detectados en una imagen

[En línea] Disponible: <http://image.slidesharecdn.com/textura-130520161819-phpapp01/95/descriptores-de-textura-50-638.jpg?cb=1369066788>

5. Asignación de la orientación para los puntos de interés: En esta etapa, se asigna una orientación a los puntos de interés anteriores para otorgarles la propiedad “Invariancia a la Rotación”. Se calcula la dirección y la magnitud de las gradientes de cada punto de interés, se forma un histograma de orientación, y la orientación más prominente de la región es asignada al punto de interés.
6. Generación de las características SIFT: Finalmente, con la rotación y escala invariantes, se realiza una última representación que facilite la identificación de las características de los puntos de interés.

Varios trabajos se han diseñado alrededor del método *SIFT* con el fin de obtener mejores rendimientos ante problemas determinados. Abdel-Hakim y Aly A. Farag proponen el *CSIFT* [19], un método que aprovecha las ventajas del *SIFT* y le añaden al descriptor invariancia en el color; de este modo, los autores obtienen una mejor estabilidad de los puntos de interés a cambios fotométricos o de iluminación. Otro ejemplo de aplicación del *SIFT* es combinándolo con otros métodos; Zhou, H., Yuan, Y., & Shi, C [18] proponen un método para el seguimiento de objetos utilizando el método *SIFT* y la *Mean Shift*. El método *SIFT* es utilizado para encontrar regiones de interés en fotografías; mientras, que el *Mean Shifts* buscan similitud a través de histograma de color; los resultados de aplicar ambas técnicas son evaluados en un esquema de máxima expectativa para obtener un modelo matemático cercano a la realidad.

1.3.4. Método *SURF* – *Speeded-up Robust Feature*

En la visión artificial [21], el procedimiento para identificar la correspondencia entre dos imágenes incluye tres pasos importantes. Primero, la selección de los puntos de interés en una imagen. Segundo, representar el vecindario de cada punto de interés como un vector de características; el descriptor debe ser robusto a: los ruidos, deformaciones geométricas y cambios de iluminación. Tercero, los descriptores de vectores son emparejados entre imágenes diferentes.

Bay, H., Tuytelaars, T., & Van Gool, L. [21], elaboraron el método *SURF* tomando como base los conceptos del método *SIFT* [8]. La diferencia entre este método y el *SIFT*, es que el método *SURF* se enfoca en desarrollar un detector y un descriptor de vectores cuyos cálculos sean más rápidos. Las ventajas del *SURF* incluyen: capacidad de ser repetible; genera resultados robustos a las variaciones de escala, rotación e iluminación; y es capaz de realizar cálculos y comparaciones rápidas.

Las principales diferencias del método *SURF* con respecto al *SIFT* son:

1. El Construcción de un espacio de escala: El método *SURF* utiliza filtros de caja para aproximarse al Laplaciano Gaussiano de una imagen; asimismo, emplea imágenes integrales para facilitar el cálculo. En la Figura 1.3, se observa dos el resultado de estos procesos. En la imagen superior se aplica un filtro de caja en la dirección Y; mientras que en la imagen inferior el filtro actúa en las direcciones XY. Los valores que se acercan a gris tienen valor aproximado a cero.

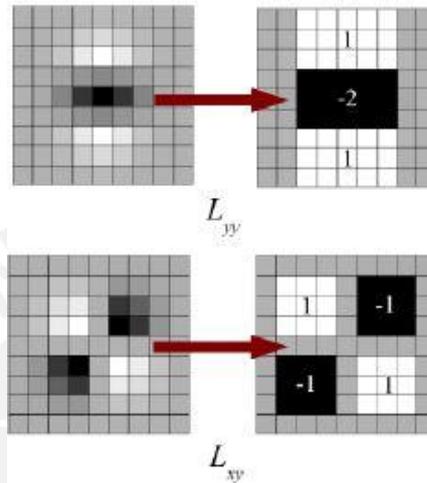


Figura 1.3 Aproximación del Laplaciano Gaussiano utilizando filtros de caja.

[En línea] Disponible: http://docs.opencv.org/3.0-beta/_images/surf_boxfilter.jpg

2. Detección de puntos de interés: El método *SURF* detecta posibles puntos de interés utilizando la matriz Hessian. El determinante de la matriz es utilizada para medir cambios locales alrededor de un punto; luego se realiza una supresión de los puntos no máximos dando como resultado los puntos de interés.

3. Asignación de la orientación de los puntos de interés: Se detecta la orientación gaussiana de cada punto de interés utilizando *wavelets* de Haar.

Jingxin Hon [22] utiliza el método *SURFT* para elaborar imágenes de mosaicos; para ello, aprovecha la robustez y el superior rendimiento del *SURF* para extraer características; asimismo, emplea métodos como *RANSAC* para una mejor unión entre las imágenes. Otras aplicaciones del *SURF* involucran la detección de espacios para el desplazamiento de vehículos en terrenos inexplorados [23].

1.3.5. Procesamiento de imágenes digitales

Se define a una imagen digital como una función $f(x, y)$, en donde x e y simbolizan coordenadas espaciales y el valor de f (intensidad) es el nivel de gris en un punto (x, y)

[24]; otro método utilizado para representar una imagen es a través de una matriz I de tamaño $M \times N$, representado de la forma:

$$I = \begin{matrix} I(1,1) & I(1,2) & \dots & I(1,M) \\ I(2,1) & I(2,2) & \dots & I(2,M) \\ \dots & \dots & \dots & \dots \\ I(N,1) & I(N,2) & \dots & I(N,M) \end{matrix}$$

Los elementos de la matriz son llamados píxeles, y en una imagen monocromática pueden adquirir valores de orden 2^8 ó 256 niveles.

En el campo de procesamiento de imágenes, R. C. Gonzalez y R. E. Woods [25], proponen un paradigma para distinguir tres tipos de procesos computadorizados, que involucran los diferentes procesamientos de imágenes: procesos de nivel bajo, medio y alto. Los procesos de bajo nivel involucran operaciones como reducción de sonido, aumento del contraste y nitidez de la imagen. Los procesos de medio nivel involucran segmentación, descripción de objetos, procesamiento y clasificación de objetos. Finalmente, los procesos de alto nivel consisten en dar sentido a una imagen; a partir de la información obtenida del nivel medio.

1.3.6. Filtros de suavizado – *Smoothing Filters*

Según R. C. Gonzalez y R. E. Woods [25], estos filtros son utilizados para remover detalles en una imagen; algunos de estos procesos consisten en: reducir pequeños detalles en una imagen y unir espacios entre líneas o curvas. Asimismo, definen dos tipos de filtros según el tipo de procedimiento matemático para calcular el suavizado de los píxeles: filtros lineales y filtros de orden estadísticos.

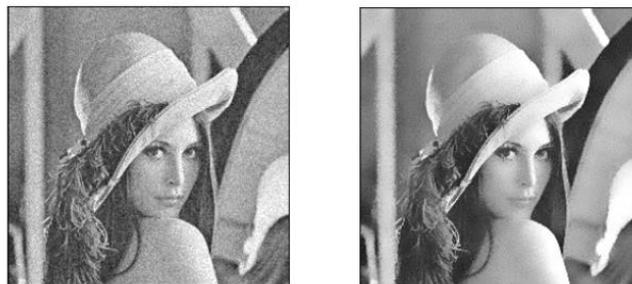


Figura 1.4: Resultado de aplicar un filtro de suavizado a una imagen.

[En línea] Disponible:

<http://cnx.org/resources/b701f12565bd730dc908767378a04d0dead25d0d/figure11.png>

La respuesta del filtro lineal es el cálculo del promedio de los píxeles contenidos en el vecindario del filtro; mientras que, la del filtro de orden estadístico, está basada valor del ordenamiento de los píxeles contenidos en el área de la imagen que rodea el filtro.

1.3.7. Filtros de nitidez – *Sharpening Filters*

Estos filtros son utilizados para realzar los detalles finos en una imagen o para mejorar el detalle de una imagen que ha sido suavizada por un error o que se originó durante el proceso de la de adquisición de la imagen. En contraposición a los filtros de suavizado, la nitidez se obtiene por la diferenciación espacial.



Figura 1.5: Resultado de aplicar un filtro de nitidez a una imagen

[En línea] Disponible: <http://studium.tobiasdeml.com/wp-content/uploads/2012/11/sharpening-mode-suggestion-dslr-controller-tobias-deml.jpg>

1.3.8. Sistema robótico Nao Aldebaran

La empresa francesa *Aldebaran Robotics* diseña y produce robots humanoides interactivos. El primero de sus proyectos fue el robot NAO que comenzó el 2006. Este robot estaba destinado para los campos de la investigación y educación.

El 2009, iniciaron el Proyecto ROMERO con el objetivo de intensificar la investigación en el apoyo a personas que han perdido autonomía. Asimismo, la empresa japonesa *SoftBank*, seleccionó a *Aldebaran Robotics* para que realizaran el diseño y creación de un robot personal al que denominaron *Pepper*.

Varias versiones de la plataforma robótica Nao están compuestas de 14, 21 o 25 grados de libertad. Cada grado de libertad indica el número mínimo de velocidades generalizadas de un sistema mecánico; en el caso del Nao, tal como se observa en la Figura 1.6, éstas se encuentran en sus brazos, rodillas, piernas, pies, muñecas, cuello, etc.

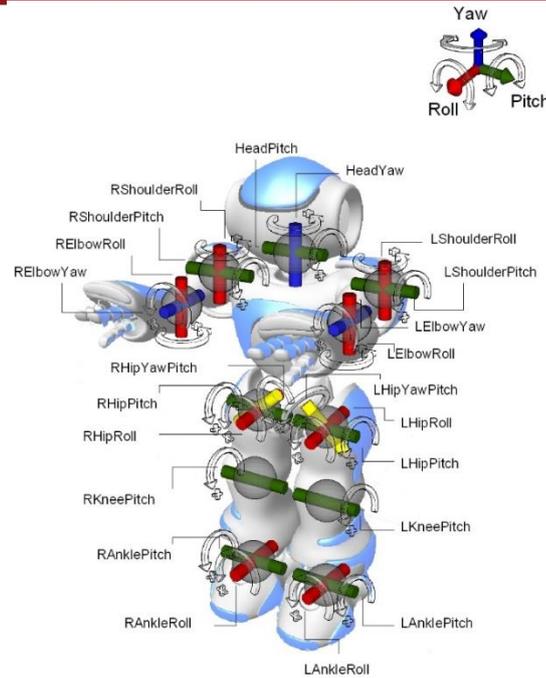


Figura 1.6: Grados de libertad del robot NAO. [En línea] Disponible: http://taaone-sti2d.fr/STI2D/ETT/Les_activites/5_7_NAO/files/documents/SiteAldebaran/donnees-techniques/ALMotionJointDefinition.jpg

Todas las versiones de Nao cuentan con unidades de medición interna como el acelerómetro, girómetro y cuatro sensores de ultrasonido que le proveen de estabilidad y posicionamiento en el espacio. Las versiones de Nao que cuentan con piernas incluyen 8 resistencias de fuerza y dos parachoques.

El robot Nao es controlado por un sistema operativo especializado y basado en Linux, denominado NAOqi. El sistema operativo potencia el sistema multimedia, que incluye cuatro micrófonos (para el reconocimiento de voz y locación de sonido), dos parlantes y dos cámaras HD.



Figura 1.7: Robot nao desarrollado por la Empresa Aldebaran Robotics.

[En línea] Disponible:

http://www.canald.com/polopoly_fs/1.1442375.1412950463!/image/aldebaran.com.jpg_gen/derivatives/cd_796_449/aldebaran.com.jp

2. Descripción y sustentación de la solución

En esta sección se presentan el marco lógico de la propuesta de solución al problema planteado; definimos el objetivo general del proyecto, los objetivos específicos, los resultados esperados y las actividades que se realizarán. Asimismo, mostramos el alcance y los límites de la propuesta de solución.

2.1. Objetivo general y objetivos específicos

Objetivo General

Desarrollar un algoritmo para la identificación de objetos y analizar su factibilidad de ser incorporarlo a un robot para que éste reconozca objetos de su entorno.

Objetivos Específicos

OE1: Definir las categorías de objetos y constituir las bases de las imágenes.

OE2: Implementar métodos de extracción de características basados en la determinación de puntos de interés (*keypoints*).

OE3: Incorporar y determinar la eficiencia de los algoritmos de extracción de características dentro del robot NAO.

OE4: Implementar un predictor probabilístico para determinar la categoría de un objeto utilizando los vectores extraídos mediante puntos de interés (*keypoints*).

2.2. Resultados esperados

Objetivo Específico 1:

- Conjunto de objetos referidos a tres categorías con aproximadamente 100 imágenes por cada categoría.

Objetivo Específico 2:

- Algoritmos de extracción de características de SIFT y SURF implementados en lenguaje Python.

Objetivo Específico 3:

- Implementación de los algoritmos de extracción de características dentro de la plataforma del robot NAO.
- Resultados de la comparación de los algoritmos respecto a eficiencia en la plataforma NAO.

Objetivo Específico 4:

- Resultados de la comparación de la precisión del clasificador.

2.3. Marco Lógico de la propuesta

En la Tabla 1, se presenta el Marco lógico de la propuesta desarrollada donde se puede apreciar el objetivo general, los objetivos específicos con sus respectivas actividades y los recursos de verificación.

Como se puede apreciar, los componentes en los cuales se colocará mayor énfasis de desarrollo son el Objetivo Específico 2 y el Objetivo Específico 4 por la cantidad de actividades que se desarrollarán. El Objetivo Específico 2, que consiste en la extracción de características de los objetos basados en puntos de interés, se constituye como el más importante de nuestra propuesta; puesto que éste sirve de base para las demás Actividades.

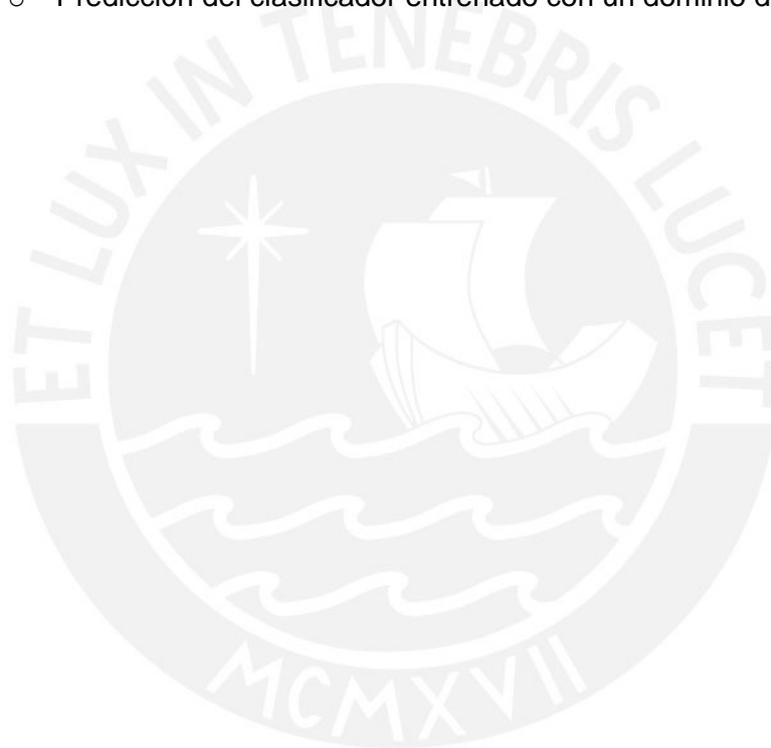
Tabla 1: Marco lógico de la propuesta desarrollada

DESCRIPCION DEL PROYECTO	ACTIVIDADES	RECURSOS DE VERIFICACION
Objetivo General - Desarrollar un algoritmo para la identificación de objetos que sea incorporado a un robot para que éste reconozca objetos de su entorno.	Actividad - Matrices de confusión de los resultados de categorización de los Modelos Clasificadores Probabilísticos	Matrices construidas al finalizar las etapas de predicción con los clasificadores
Objetivo Específico 1 - Definir las categorías de objetos y constituir las bases de las imágenes.	Actividad 1 - Realizar la captura de imágenes digitales de un dominio de objetos.	Dominio de imágenes de los objetos
	Actividad 2 - Restringir el dominio de los objetos a tres categorías.	Dominio de imágenes de los objetos
Objetivo Específico 2 - Implementar el método SIFT y el método SURF en un conjunto de objetos y comparar sus fortalezas en función de los resultados.	Actividad 1 - Comparar la cantidad de puntos de interés generados por los métodos	Gráfico de líneas con la cantidad de puntos de interés obtenidas por los métodos
	Actividad 2 - Comparar del tiempo requerido por los métodos para cada proceso de extracción	Gráfico de líneas con el tiempo requerido por los métodos
	Actividad 3 - Comparar la cantidad de puntos de interés generados por los métodos ante cambios de iluminación	Gráfico de líneas con la cantidad de puntos de interés obtenidas por los métodos
	Actividad 4 - Comparar el tiempo requerido por los métodos para cada proceso de extracción ante cambios de iluminación	Gráfico de líneas con el tiempo requerido por los métodos
Objetivo Específico 3 - Incorporar y determinar la eficiencia de los algoritmos de extracción de características dentro del robot NAO.	Actividad 1 - Comparar el tiempo requerido por los métodos SIFT y SURF para cada proceso de extracción	Gráfico de líneas con el tiempo requerido por los métodos
	Actividad 2 - Comparar el tiempo requerido por los métodos SIFT y SURF para cada proceso de extracción ante cambios de iluminación	Gráfico de líneas con el tiempo requerido por los métodos
Objetivo Específico 4 - Implementar un clasificador probabilístico para determinar la categoría de un objeto utilizando la información de los métodos SIFT y SURF.	Actividad 1 - Comparar la exactitud del clasificador utilizando el método SIFT	Matriz de confusión para el clasificador
	Actividad 2 - Comparar la exactitud del clasificador utilizando el método SURF	Matriz de confusión para el clasificador
	Actividad 3 - Comparar de la precisión del clasificador utilizando el método SIFT	Matriz de confusión para el clasificador
	Actividad 4 - Comparar de la precisión del clasificador utilizando el método SURF	Matriz de confusión para el clasificador

2.4. Alcances de la solución propuesta

Debido a la complejidad de replicar el sentido de la visión de los humanos, se ha convenido reducir el alcance a puntos específicos:

- Las capturas de imágenes para las pruebas serán tomadas en un ambiente controlado, libre de colores en el fondo.
- Se utilizará tres categorías de objetos para construir el dominio predefinido. Estos tipos de objetos se utilizarán durante la fase de entrenamiento del clasificador y la fase de predicción.
- El reconocimiento de objetos se basará en dos partes:
 - o Detección de la información de un objeto en la escena
 - o Predicción del clasificador entrenado con un dominio de objetos.



3. Caracterización de objetos con métodos de Visión Computacional

A continuación, se explica la caracterización de los objetos que serán utilizados en las etapas de prueba. El procedimiento de caracterización consta de dos etapas: En la primera, luego de aplicar el método *SIFT* y *SURF* para un conjunto de imágenes de flores, compararemos y analizaremos los resultados. En la segunda etapa, analizaremos los objetos que finalmente utilizaremos para elaborar el diseño del modelo predictor de categorización de objetos por visión computacional.

3.1. Aplicación del método SIFT en un conjunto de imágenes

Este método permite extraer características relevantes de imágenes conocidas como puntos de interés. Para esta etapa, se analizó los puntos de interés obtenidos en 20 imágenes de diferentes flores [Anexo A]; cada uno de los puntos encontrados tiene propiedades como invariante a los cambios de escala y la rotación; representados visualmente en la Figura 2.2.

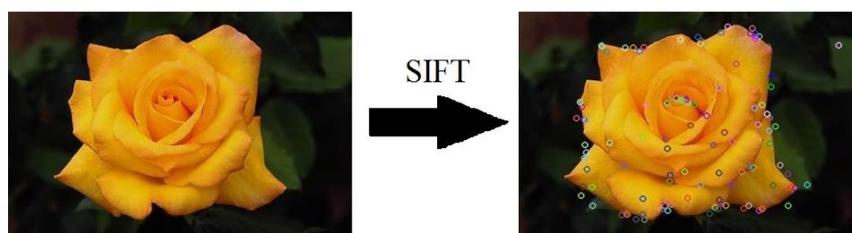


Figura 3.1: A la izquierda, la imagen original. A la derecha, imagen con los puntos de interés del método SIFT

[En línea] Disponible: <http://wang.ist.psu.edu/docs/related/>

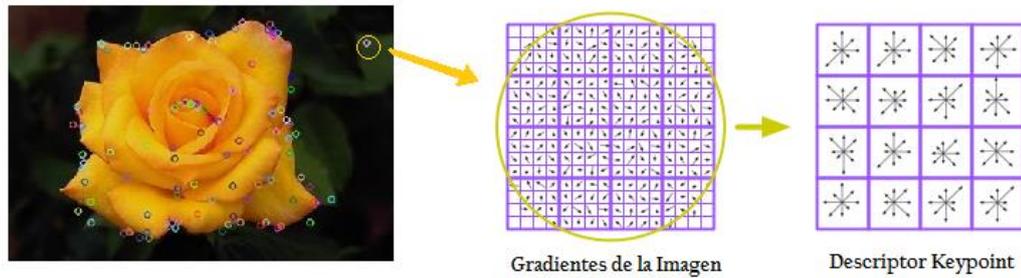


Figura 3.2: Cálculo de la gradiente alrededor de un Keypoint

[En línea] Disponible:

<http://www.cc.gatech.edu/~hays/compvision/results/proj2/rchawla8/index.html>

En la Figura 3.2, el descriptor del *SIFT* puede analizarse de la siguiente manera: la Gradiente de la Imagen está compuesto de una ventana de gradiente de tamaño 4x4; cada ventana tiene 8 direcciones, con lo cual tenemos un total de 128 direcciones. Entonces un punto de interés, representado por un vector de 1x128 valores, brinda información de direcciones de gradiente.

Tabla 2: Información de un punto de interés SIFT vectorial (1x128) con los valores de histograma de cada orientación

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16
11	4	1	15	55	4	3	18	4	3	1	32	115	24	25	20
V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29	V30	V31	V32
13	5	0	4	12	53	115	61	1	2	1	1	5	115	115	4
V33	V34	V35	V36	V37	V38	V39	V40	V41	V42	V43	V44	V45	V46	V47	V48
8	36	63	61	31	2	0	0	23	40	29	89	47	11	11	18
V49	V50	V51	V52	V53	V54	V55	V56	V57	V58	V59	V60	V61	V62	V63	V64
115	89	8	4	2	10	37	52	20	4	1	33	98	91	42	4
V65	V66	V67	V68	V69	V70	V71	V72	V73	V74	V75	V76	V77	V78	V79	V80
31	46	53	17	2	1	6	13	26	13	13	23	9	10	79	87
V81	V82	V83	V84	V85	V86	V87	V88	V89	V90	V91	V92	V93	V94	V95	V96
112	22	29	30	2	0	19	100	13	3	10	115	94	0	0	1
V97	V98	V99	V100	V101	V102	V103	V104	V105	V106	V107	V108	V109	V110	V111	V112
47	63	10	2	0	0	1	7	17	50	112	6	0	0	8	19
V113	V114	V115	V116	V117	V118	V119	V120	V121	V122	V123	V124	V125	V126	V127	V128
3	38	115	45	2	0	2	7	0	1	13	115	56	0	0	0

En la Tabla 2, observamos la información de un punto de interés de dimensión 1×128 . Los valores $V\#$, donde $\#$ es un número que va desde 1 a 128, representan las orientaciones de cada gradiente; asimismo el valor de cada orientación es del histograma; por ello, el rango del valor es $[1,255]$. Para el conjunto de 20 flores analizadas en 8.1 [Anexo A], se obtuvo un total de 4158 puntos de interés en un tiempo de procesamiento de 1.187 segundos.

Algunas de las fortalezas del SIFT frente a otros métodos de reconocimiento: son la eficiencia de su respuesta ante cambios de escala, rotación, suavizado e iluminación [31]. Se realizaron pruebas acerca de la última propiedad en un conjunto de flores para comparar los efectos de mayor y menor iluminación.

Tabla 3: Puntos de interés SIFT encontrados en 20 flores

Grupo	Keypoints	Grupo	Keypoints
Flower 1	127	Flower 11	168
Flower 2	251	Flower 12	226
Flower 3	202	Flower 13	271
Flower 4	257	Flower 14	257
Flower 5	176	Flower 15	319
Flower 6	266	Flower 16	265
Flower 7	235	Flower 17	57
Flower 8	184	Flower 18	212
Flower 9	122	Flower 19	87
Flower 10	186	Flower 20	290

En la Figura 3.3, se muestran los procedimientos de aumento y reducción de iluminación en las imágenes en un 50% respecto a la imagen original para analizar sus efectos. Se eligieron 10 flores al azar para luego calcular la cantidad de puntos de interés encontrados, así como, el tiempo de procesamiento requerido; en la Tabla 4, se observan los resultados.



Figura 3.3 A la izquierda la imagen con baja iluminación. Al medio la imagen original. A la izquierda la imagen con alta iluminación. [Desarrollado por el autor]

Tabla 4: Cuadro comparativo de los puntos detectados en imágenes oscuras e iluminadas - SIFT

Group	Original		Oscurecido		Iluminado	
	Keypoints	Keypoints	Tiempo (seg)	Keypoints	Tiempo (seg)	
Flower 1	127	53	0.10999	202	0.13499	
Flower 2	251	54	0.05999	400	0.15000	
Flower 3	202	90	0.05999	377	0.09999	
Flower 4	257	35	0.04500	536	0.10999	
Flower 5	176	66	0.05999	147	0.08000	
Flower 6	266	104	0.06200	461	0.09399	
Flower 7	235	83	0.04700	304	0.06200	
Flower 8	184	54	0.04799	338	0.07899	
Flower 9	122	14	0.04600	259	0.06299	
Flower 10	186	50	0.04699	198	0.06200	

3.2. Aplicación del método SURF en un conjunto de imágenes

El método SURF permite extraer características relevantes de imágenes; su principal diferencia con el método SIFT en resultados es que su velocidad de cálculo y robustez es mayor [21]. En la Figura 3.4, podemos observar que el descriptor interior está dividido en 8x8 ventanas; el método SURF, calcula la orientación de la gradiente de cada una de las ventanas, dando como resultado un descriptor vectorial de 64 valores.

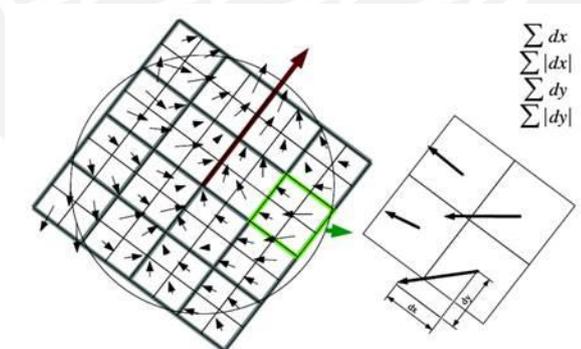


Figura 3.4: Descriptor de un punto de interés utilizando el método SURF.

[En línea] Disponible: <https://goo.gl/Rqzbvq>



Figura 3.5: A la izquierda la imagen original. A la derecha imagen con los puntos de interés del SURF [Desarrollado por el autor]

En la Tabla 5, se muestra la información de un punto de interés SURF de dimensión 1x64 de la Figura 3.5. Los valores $V\#$, donde $\#$ es un número que va desde 1 a 64, representan las orientaciones de cada gradiente; asimismo el valor de cada orientación es del histograma; en el caso del método *SURF*, el rango de valores es $[-1,1]$. Para el conjunto de 20 flores analizadas en el [Anexo A], se obtuvo un total de 3741 puntos de interés en un tiempo de procesamiento de 0.695 segundos.

Tabla 5: Información de un punto de interés *SURF* vectorial (1x64) con los valores de histograma de cada orientación

V1	V2	V3	V4	V5	V6	V7	V8
0,003474564	0,001543818	0,004035587	0,003233975	0,006128892	-0,023428194	0,05355544	0,027723268
V9	V10	V11	V12	V13	V14	V15	V16
-0,01487932	-0,005014275	0,052040219	0,018404707	0,004761568	0,000899469	0,005341252	0,001734968
V17	V18	V19	V20	V21	V22	V23	V24
0,011429873	-0,035626914	0,041108482	0,03603483	0,184777707	0,076088794	0,409126759	0,221205369
V25	V26	V27	V28	V29	V30	V31	V32
-0,020357555	0,232499987	0,214531213	0,279458702	-0,002998001	-0,006231151	0,024509613	0,012537256
V33	V34	V35	V36	V37	V38	V39	V40
0,00558805	-0,025661754	0,01648327	0,028124185	0,393814594	-0,199274465	0,420768678	0,202204391
V41	V42	V43	V44	V45	V46	V47	V48
-0,110045902	-0,206439152	0,182186946	0,206439152	0,00332528	-0,026926158	0,03517573	0,032503903
V49	V50	V51	V52	V53	V54	V55	V56
-0,033189274	0,02429381	0,035065632	0,004014444	-0,003782583	0,00475408	0,0038549	0
V57	V58	V59	V60	V61	V62	V63	V64
-0,008100849	-0,033189274	0,02429381	0,035065632	0,004014444	-0,003782583	0,00475408	0,0038549

Tabla 6: Puntos de interés *SURF* encontrados en cada una de las 20 flores

Grupo	Keypoints	Grupo	Keypoints
Flower 1	168	Flower 11	139
Flower 2	244	Flower 12	182
Flower 3	175	Flower 13	229
Flower 4	244	Flower 14	223
Flower 5	136	Flower 15	219
Flower 6	248	Flower 16	207
Flower 7	205	Flower 17	105
Flower 8	169	Flower 18	180
Flower 9	164	Flower 19	124
Flower 10	135	Flower 20	245

Igual que el procedimiento anterior, se utilizaron las imágenes de diez flores sometidas al procesamiento de aumento y reducción de iluminación para calcular los puntos de interés y el tiempo de ejecución. Los resultados se observan en la Tabla 7.

Tabla 7: Cuadro comparativo de los puntos detectados en imágenes oscurecidas e iluminadas - *SURF*

Group	Original <i>SURF</i>	Oscurecido <i>SURF</i>		Iluminado <i>SURF</i>	
	Keypoints	Keypoints	Tiempo (seg)	Keypoints	Tiempo (seg)
Flower 1	168	70	0.04600	204	0.06200
Flower 2	244	92	0.04699	305	0.09399
Flower 3	175	67	0.04700	242	0.07800
Flower 4	244	107	0.04699	323	0.09299
Flower 5	136	83	0.04699	135	0.04700
Flower 6	248	125	0.06200	322	0.09399
Flower 7	205	113	0.05800	249	0.07800
Flower 8	169	87	0.04699	220	0.07800
Flower 9	164	33	0.04699	258	0.06199
Flower 10	135	73	0.03100	139	0.06300

3.3. Comparación de los resultados obtenidos por los métodos *SIFT* y *SURF*

A partir de las pruebas iniciales realizadas en 3.1 y 3.2 se puede obtener conclusiones importantes para las etapas posteriores.

1. El predictor del método *SIFT* genera puntos de interés con dimensión vectorial 1x128 para cada punto; mientras que el método *SURF* genera puntos de interés con dimensión vectorial 1x64 valores. Sin embargo, la efectividad de los métodos se

observarán en la sección 4.2, durante la predicción utilizando los modelos *SIFT* y *SURF*.

2. En la Tabla 8 y Figura 3.6, se observa que el método *SIFT* encontró mayor cantidad de puntos en comparación con el *SURF* para casi todas las imágenes originales.

Tabla 8: Cuadro comparativo de la cantidad de puntos de interés encontrados utilizando método *SIFT* y *SURF*

Group	Keypoints SIFT	Keypoints SURF	Group	Keypoints SIFT	Keypoints SURF
Flower 1	127	168	Flower 11	168	139
Flower 2	251	244	Flower 12	226	182
Flower 3	202	175	Flower 13	271	229
Flower 4	257	244	Flower 14	257	223
Flower 5	176	136	Flower 15	319	219
Flower 6	266	248	Flower 16	265	207
Flower 7	235	205	Flower 17	57	105
Flower 8	184	169	Flower 18	212	180
Flower 9	122	164	Flower 19	87	124
Flower 10	186	135	Flower 20	290	245

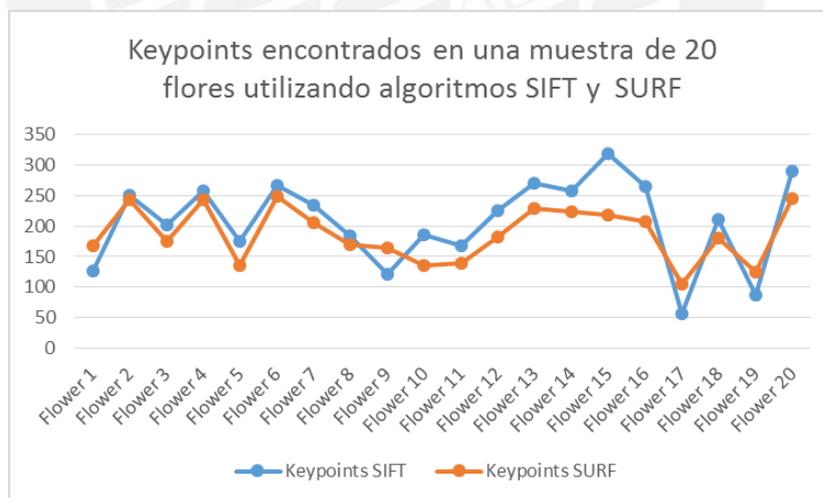


Figura 3.6: Cantidad de puntos de interés encontrados en una muestra de 20 flores utilizando método *SIFT* y *SURF*

3. En la Tabla 9 y en la Figura 3.7, se observa el resultado de comparar los puntos de interés encontrados en imágenes oscurecidas utilizando *SIFT* y *SURF*. Los resultados muestran que se detectaron más puntos de interés con el método *SURF* y tendiendo a utilizar menor tiempo.

Tabla 9: Comparación de puntos de interés encontrados en imágenes oscurecidas al utilizar método *SIFT* y *SURF*

Group	Oscurecido SIFT		Oscurecido SURF	
	Keypoints	Tiempo (seg)	Keypoints	Tiempo (seg)
Flower 1	53	0.10999	70	0.04600
Flower 2	54	0.05999	92	0.04699
Flower 3	90	0.05999	67	0.04700
Flower 4	35	0.04500	107	0.04699
Flower 5	66	0.05999	83	0.04699
Flower 6	104	0.06200	125	0.06200
Flower 7	83	0.04700	113	0.05800
Flower 8	54	0.04799	87	0.04699
Flower 9	14	0.04600	33	0.04699
Flower 10	50	0.04699	73	0.03100

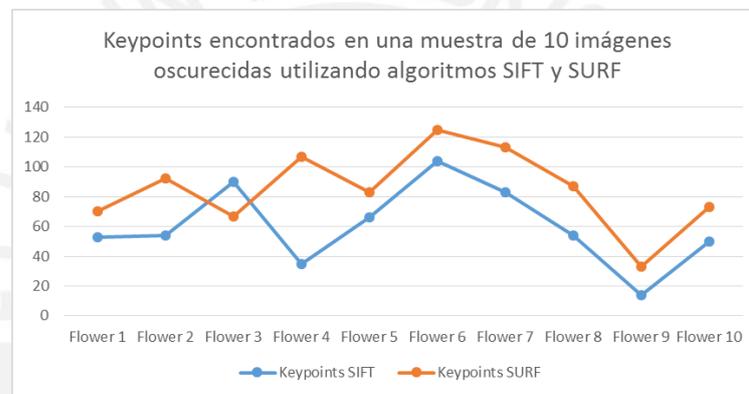


Figura 3.7 Puntos de interés encontrados en una muestra de 10 imágenes utilizando métodos *SIFT* y *SURF*

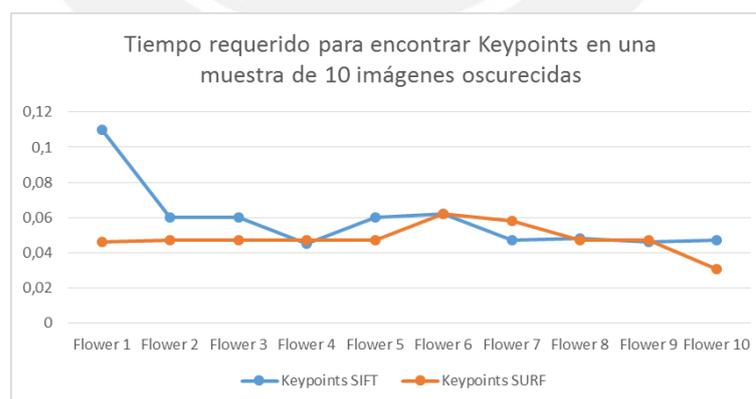


Figura 3.8 Tiempos requeridos para encontrar los puntos de interés en imágenes oscurecidas utilizando métodos *SIFT* y *SURF*

4. En la Tabla 10 y en la Figura 3.9, se observa que el cuadro comparativo de los puntos de interés encontrados en imágenes iluminadas utilizando *SIFT* y *SURF*. Los

resultados muestran que se detectaron más puntos de interés con el método *SIFT*, pero utilizando mayor tiempo.

5. En conclusión, en varias de las pruebas, el método *SIFT* genera más puntos de interés que el método *SURF*, a excepción de las imágenes oscurecidas; entonces, el método *SIFT* suele brindar mayor información de una imagen por la cantidad puntos de interés y por las dimensiones vectoriales de las mismas. Esto puede ser útil para la fase de entrenamiento del modelo predictor; sin embargo, un análisis posterior en la siguiente sección lo definirá.

Tabla 10: Comparación de puntos de interés encontrados en imágenes iluminadas al utilizar método *SIFT* y *SURF*

Group	Iluminado SIFT		Iluminado SURF	
	Keypoints	Tiempo (seg)	Keypoints	Tiempo (seg)
Flower 1	202	0.13499	204	0.06200
Flower 2	400	0.15000	305	0.09399
Flower 3	377	0.09999	242	0.07800
Flower 4	536	0.10999	323	0.09299
Flower 5	147	0.08000	135	0.04700
Flower 6	461	0.09399	322	0.09399
Flower 7	304	0.06200	249	0.07800
Flower 8	338	0.07899	220	0.07800
Flower 9	259	0.06299	258	0.06199
Flower 10	198	0.06200	139	0.06300

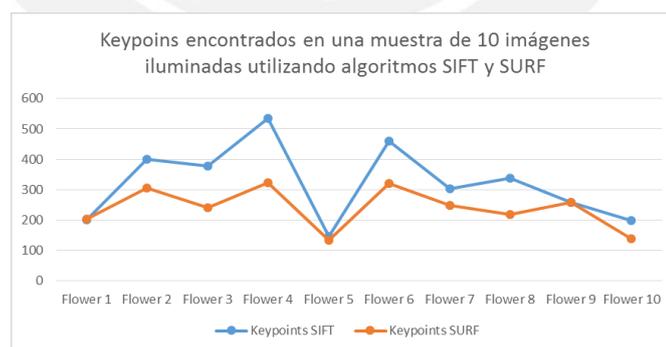


Figura 3.9 Puntos de interés encontrados en una muestra de 10 imágenes utilizando métodos *SIFT* y *SURF*

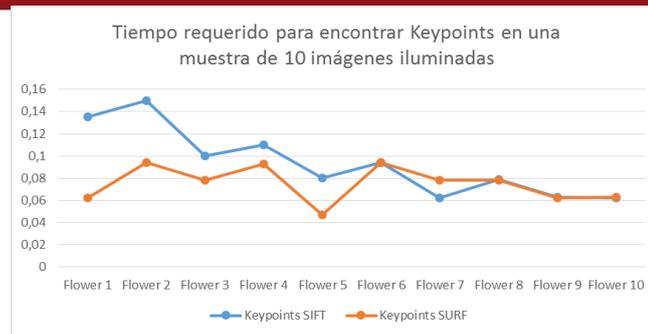


Figura 3.10 Tiempos requeridos para encontrar los puntos de interés en imágenes oscurecidas utilizando métodos *SIFT* y *SURF*

3.4. Selección del dominio de objetos que construirán nuestro modelo predictor.

El dominio de objetos que se utilizará consiste en tres frutas (manzanas, mandarinas y granadillas); éstos fueron elegidos por las siguientes razones:

1. Cada uno de los objetos tiene propiedades visuales importantes para el método como: superficie uniforme en todo el cuerpo y su forma no varía considerablemente según el ángulo de visión.
2. El tamaño de los objetos es adecuado para que la cámara del robot los perciba en su totalidad y no solo una parte de las mismas.

Durante el procedimiento de capturar las imágenes del dominio de objetos, se consideró que debe hacerse sobre un fondo oscuro y unicolor; de este modo, se evita que ruido de fondo genere puntos de interés errados. Para la etapa de entrenamiento y pruebas del predictor se dividió el conjunto de imágenes en seis grupos: 80 manzanas, 80 mandarinas y 80 granadillas para el entrenamiento; 40 manzanas, 40 mandarinas y 40 granadillas para la predicción.



Figura 3.11 Conjunto de imágenes de mandarinas



Figura 3.12 Conjunto de imágenes de granadillas

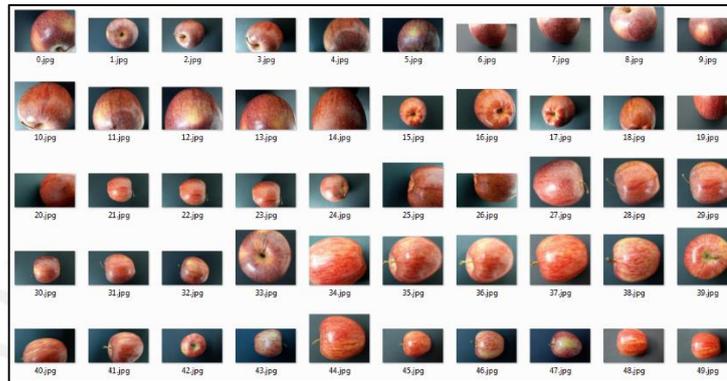


Figura 3.13 Conjunto de imágenes de manzanas

A partir de las imágenes seleccionadas para crear el dominio de objetos, se realizó un análisis de dichos grupos para encontrar propiedades comunes que pudieran influenciar en el desarrollo del modelo predictor.

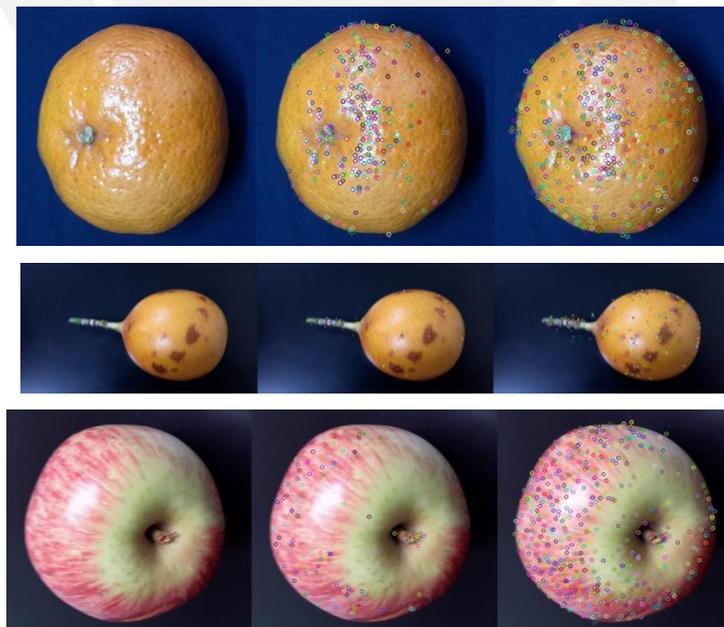


Figura 3.14 (Izquierda) Imágenes originales. (Centro) Imágenes con puntos SIFT.
(Derecha) Imágenes con puntos SURF

Tabla 11: Comparación de los puntos de interés SIFT y SURF en 50 imágenes de tres tipos de objetos

	SIFT	SURF	SIFT	SURF	SIFT	SURF
	Mandarina	Mandarina	Granadilla	Granadilla	Manzana	Manzana
Imagen 1	407	251	54	137	119	280
Imagen 2	219	195	235	179	269	209
Imagen 3	398	219	235	244	122	235
Imagen 4	88	129	718	321	198	247
Imagen 5	99	140	841	411	222	321
Imagen 6	61	68	971	656	198	299
Imagen 7	45	51	850	372	339	267
Imagen 8	395	252	269	368	212	173
Imagen 9	346	228	209	261	392	300
Imagen 10	273	184	594	328	453	278
Imagen 11	327	360	231	383	183	253
Imagen 12	510	426	449	362	308	295
Imagen 13	92	243	663	379	230	304
Imagen 14	127	242	221	163	209	309
Imagen 15	238	469	584	253	312	311
Imagen 16	189	306	496	224	806	462
Imagen 17	276	319	495	210	1100	703
Imagen 18	137	287	781	319	681	836
Imagen 19	318	385	758	294	705	575
Imagen 20	375	320	615	264	306	517
Imagen 21	454	291	433	234	783	635
Imagen 22	72	85	767	340	413	450
Imagen 23	71	236	536	271	322	384
Imagen 24	70	122	173	441	68	235
Imagen 25	458	324	522	262	347	446
Imagen 26	398	350	652	440	294	383
Imagen 27	328	360	244	422	333	393
Imagen 28	345	286	253	399	397	513
Imagen 29	176	224	376	276	459	538
Imagen 30	280	182	631	255	277	542
Imagen 31	473	374	798	392	381	616
Imagen 32	386	316	474	241	356	520
Imagen 33	272	240	753	317	353	527
Imagen 34	260	251	807	393	202	444
Imagen 35	292	270	309	208	191	289
Imagen 36	365	336	210	208	151	267
Imagen 37	417	309	67	57	197	280
Imagen 38	318	316	366	292	107	261
Imagen 39	444	345	248	171	289	317
Imagen 40	323	338	108	150	248	357
Imagen 41	495	349	305	238	292	335
Imagen 42	342	289	177	221	248	312
Imagen 43	402	282	223	242	65	225
Imagen 44	406	372	197	172	234	296
Imagen 45	202	333	417	244	161	284
Imagen 46	378	494	228	233	403	496
Imagen 47	76	257	284	306	226	375
Imagen 48	45	237	598	339	193	376
Imagen 49	139	288	108	122	207	405
Imagen 50	134	168	385	168	64	294

A partir de las Figura 3.15 y Figura 3.16 se obtienen los siguientes resultados:

- Con el método SIFT, el grupo de Granadillas tiende a tener más puntos de interés que los otros grupos.
- Con el método SURF, el grupo de Manzanas tiende a tener más puntos de interés que los otros grupos.
- La cantidad de puntos de interés es muy variada incluso en imágenes de un mismo tipo de objeto. Por ejemplo; en la Tabla 10, la columna de SIFT manzana tiene 68 puntos de interés en la Imagen 24 y 1100 puntos de interés en la Imagen 17.

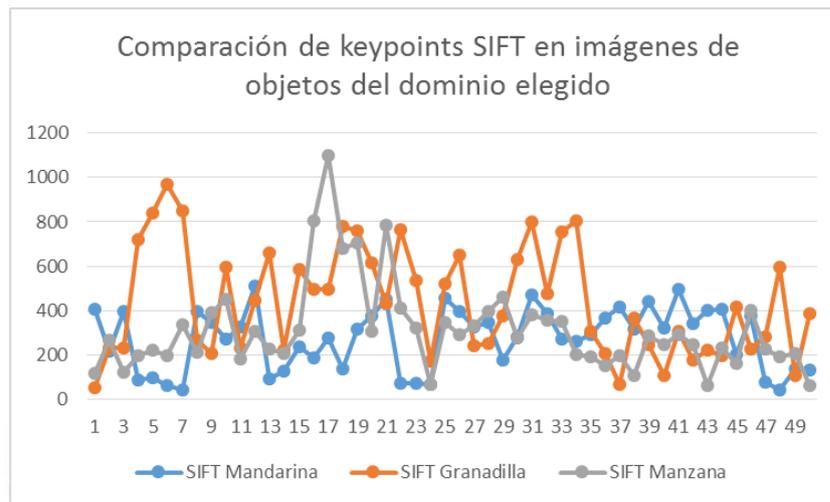


Figura 3.15 Comparación de puntos de interés *SIFT* en imágenes de objetos del dominio elegido

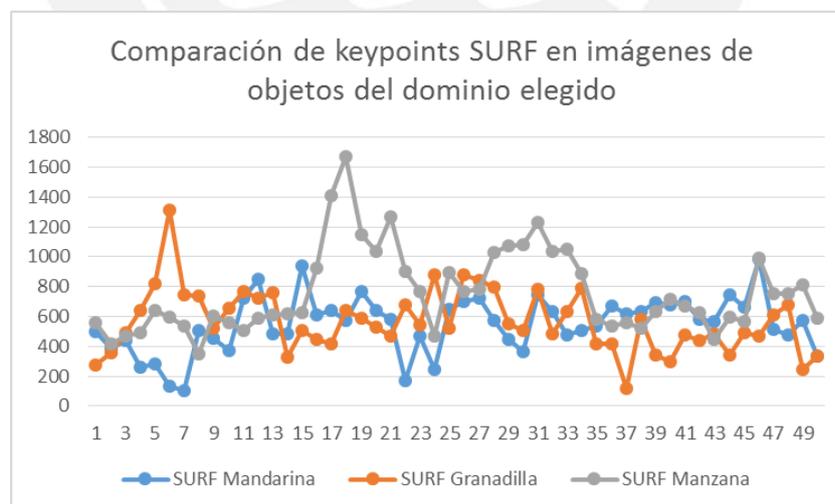


Figura 3.16 Comparación de puntos de interés *SURF* en imágenes del dominio elegido

4. Diseño de la solución

En el presente capítulo se aborda la solución en función de los resultados obtenidos en la etapa anterior.

4.1. Flujo general del sistema

Tal como se mencionó en la Sección 0, el objetivo del presente trabajo consiste en implementar dos métodos de reconocimiento de objetos para que el robot pueda diferenciar objetos en un entorno específico.

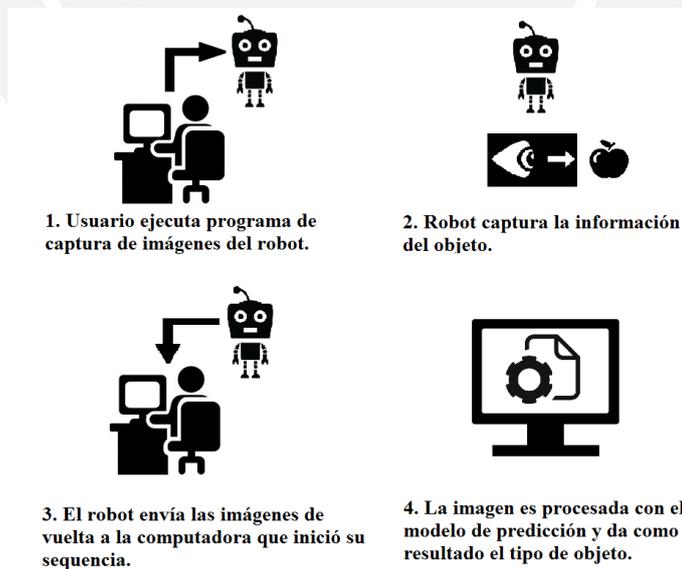


Figura 4.1 Flujo del control sistema y el sistema robótico

En la figura Figura 4.1, se muestra una propuesta para el flujo del funcionamiento del sistema:

1. El usuario inicia el funcionamiento del Sistema Robótico a través del Nao.
2. El sistema robótico comienza a captar información a través de su cámara y envía las imágenes capturadas a la computadora.
3. La computadora recibe las imágenes y comienza a procesarlas. Analiza la información y predice según el modelo seleccionado la categoría del objeto que ha recibido.
4. La computadora envía el resultado al sistema robótico y ésta ejecuta el resultado.

4.2. Diseño del modelo predictor utilizando un clasificador bayesiano para predecir la categoría de un objeto

En esta sección se explica el procedimiento para la construcción del clasificador bayesiano; para ello, se realizan pruebas y verificaciones de modelos predictores que utilizan el método SIFT y el método SURF. Luego, se comparan la efectividad de los modelos, en función de su precisión para predecir la categoría de un tipo de objeto.

4.2.1. Consideraciones previas

Para generar el modelo predictor, primero construimos conjuntos de puntos de interés con las siguientes consideraciones:

- a. Cada imagen genera un conjunto de puntos de interés; éstos los dividimos en dos grandes grupos que dependen del método utilizado para encontrar los puntos de interés. Los grupos son: Conjunto de puntos SIFT y Conjunto de puntos SURF.
- b. Las imágenes pertenecen a una de las categorías de objetos elegidos para este trabajo (mandarinas, manzanas o granadillas).
- c. Utilizaremos 80 imágenes de entrenamiento por cada categoría de objeto para el Clasificador Bayesiano; es en total 240 imágenes de entrenamiento.
- d. Las pruebas de predicción del Clasificador Bayesiano, se realizarán con 40 imágenes por cada categoría.
- e. Cada imagen puede generar distintas cantidades de puntos de interés; así que, además de utilizar la misma cantidad de imágenes para cada categoría, extraemos puntos de interés aleatoriamente para mantener uniformidad en la cantidad de puntos de interés por imagen. Esta selección aleatoria no considera criterios como relevancia de un punto de interés respecto de otros.

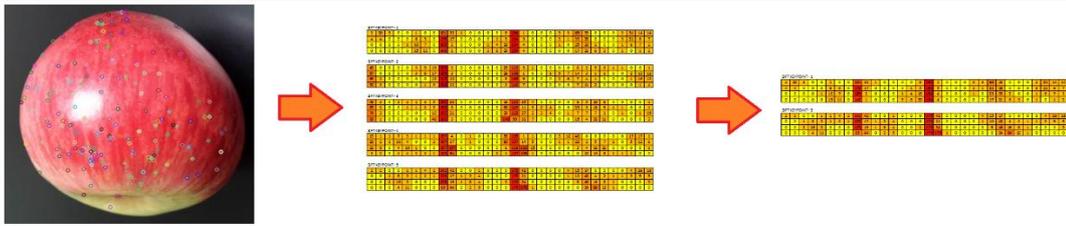


Figura 4.2 Selección aleatoria de puntos de interés *SIFT* del conjunto total de puntos de interés originales

4.2.2. Modelo predictor *SIFT*

La cantidad de puntos de interés obtenida por el método *SIFT* para las 240 imágenes puede observarse en la Tabla 11 y en la Figura 4.3. A partir de ellos, obtenemos la siguiente información:

1. La cantidad de puntos de interés es muy variada en los tres tipos de objetos.
2. Los valores promedios de puntos de interés para el grupo de Mandarinas, Granadillas y Manzanas es 376.86, 422.80 y 447.708, respectivamente.

Debido a la variedad de puntos de interés, necesitamos saber cómo la cantidad de ellos afecta la etapa de predicción de los objetos; para ello, consideramos lo siguiente:

1. Construir 4 modelos considerando 50, 100, 150 y 200 puntos de interés en cada imagen. Si una imagen cuenta con menor cantidad de puntos, se consideran todos sus puntos de interés; caso contrario, se eligen aleatoriamente por cada imagen.
2. Analizar la precisión y exactitud de los modelos y cómo afecta la cantidad de puntos de interés elegidos.

Tabla 12: Cantidad de puntos de interés encontrados con el método *SIFT*

MANDARINAS		GRANADILLAS		MANZANAS	
303	323	187	108	667	248
407	495	54	305	119	292
219	342	216	177	269	248
398	402	235	223	122	65
88	406	718	197	198	769
148	202	841	417	222	161
483	378	971	228	198	403
417	76	850	284	339	226
1304	1153	269	598	212	193
346	391	209	108	392	207
273	134	594	385	453	64
327	51	231	219	183	525
510	414	449	118	308	299
92	432	663	279	230	605
725	271	221	93	209	811
238	145	584	279	312	324
189	185	496	118	806	829
276	254	495	216	1100	573
137	295	781	1999	681	347
318	208	758	67	705	772
375	382	615	219	306	1083
454	987	433	446	783	975
72	940	767	710	413	888
385	406	536	997	322	900
984	1023	173	114	68	580
220	788	522	231	654	487
398	191	652	667	294	1119
328	345	244	92	333	350
345	227	253	263	397	404
176	389	376	1258	459	791
280	269	631	371	277	695
473	337	798	1031	381	686
386	54	474	99	356	942
272	101	753	191	353	916
260	194	807	346	202	205
292	415	309	15	191	428
365	573	210	446	151	238
417	405	67	131	197	861
318	586	98	643	107	407
444	538	248	348	289	642

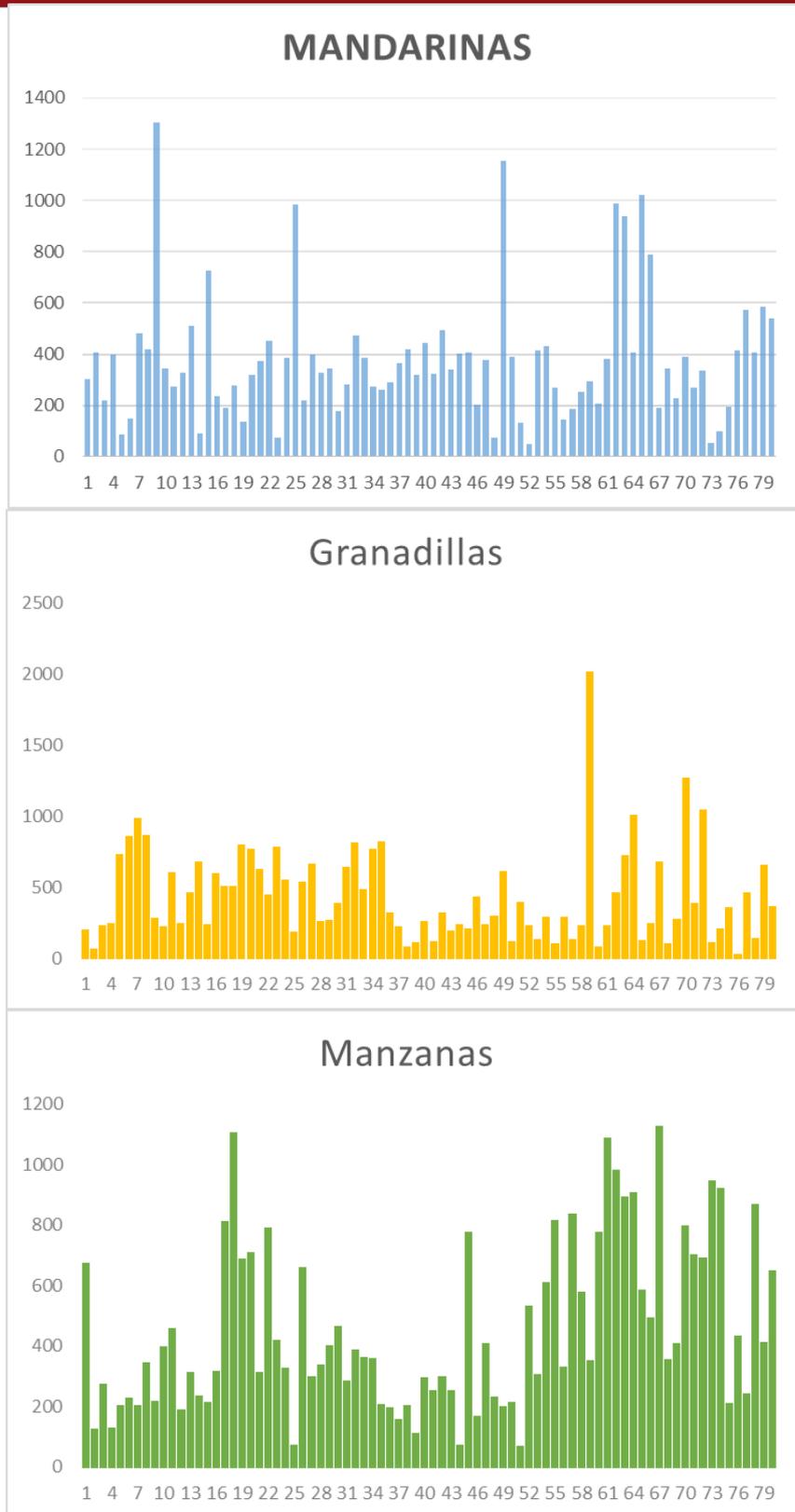


Figura 4.3 Gráficos de barras con las cantidades de puntos de interés encontrados con el método *SIFT*

Tabla 13: Resultados estadísticos de la cantidad de puntos de interés encontrados con el método *SIFT*

	Mandarinas	Granadillas	Manzanas
Promedio	377	423	448
Maximo	1304	1999	1100
Minimo	51	15	64

En la Tabla 14, Tabla 15, Tabla 16 y Tabla 17, se observan los resultados del proceso de predicción; a partir de éstos podemos sacar las siguientes conclusiones:

- El modelo que utiliza 50 puntos de interés tiene la mayor precisión de predicción; mientras que el modelo con 200 puntos de interés, tiene menor precisión.
- Se observa que al incrementar los puntos de interés, disminuye la precisión del modelo.
- El modelo con 50 puntos de interés tiene la mayor precisión al predecir Mandarinas.
- El modelo con 50 puntos de interés tiene la mayor precisión al predecir Granadillas.
- Los modelos con 150 y 200 puntos de interés tienen la mayor precisión al predecir Manzanas.
- El modelo con 100 puntos de interés tiene la mayor exactitud al predecir Mandarinas, pero no la mayor precisión.
- Los modelos con 150 y 200 puntos de interés tienen la mayor exactitud al predecir Granadillas, pero es el modelo de 150 el que tiene también mayor precisión.
- El modelo con 50 puntos de interés tiene la mayor exactitud al predecir Manzanas, pero no la mayor precisión.

Tabla 14: Precisión del modelo utilizando el algoritmo SIFT con 50 keypoints

MODELO SIFT CON 50 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	35	1	4	87,5
Granadillas	2	37	1	92,5
Manzanas	4	5	31	77,5
Exactitud (%)	85,37	86,05	86,11	85,83
Precisión del modelo (%)	85,83			

Tabla 15: Precisión del modelo utilizando el algoritmo SIFT con 100 keypoints

MODELO SIFT CON 100 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	30	2	8	75
Granadillas	2	34	4	85
Manzanas	0	6	34	85
Exactitud (%)	93,75	80,95	73,91	81,67
Precisión del modelo (%)	81,67			

Tabla 16: Precisión del modelo utilizando el algoritmo SIFT con 150 keypoints

MODELO SIFT CON 150 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	16	0	24	40
Granadillas	2	27	11	67,5
Manzanas	0	0	40	100
Exactitud (%)	88,89	100,00	53,33	69,17
Precisión del modelo (%)	69,17			

Tabla 17: Precisión del modelo utilizando el algoritmo SIFT con 200 keypoints

MODELO SIFT CON 200 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	18	0	22	45
Granadillas	2	18	20	45
Manzanas	0	0	40	100
Exactitud (%)	90,00	100,00	48,78	63,33
Precisión del modelo (%)	63,33			

4.2.3. Modelo predictor SURF

La cantidad de puntos de interés obtenida por el método SURF para las 240 imágenes puede observarse en la Tabla 18 y la

Figura 4.4; de igual modo obtenemos la siguiente información:

1. El método SURF genera más puntos de interés que el método *SIFT*.
2. Los valores promedios de puntos de interés para el grupo de Mandarinas, Granadillas y Manzanas es 806.45, 488.50 y 982.75, respectivamente.

Tabla 18: Cantidad de puntos de interés encontrados con el método SURF en una porción de las imágenes

MANDARINAS		GRANADILLAS		MANZANAS	
568	677	165	301	1156	715
502	699	274	477	560	670
390	578	269	442	418	624
438	565	488	484	470	450
259	744	642	345	495	975
549	666	822	489	643	568
1431	988	1312	466	598	993
1202	515	744	612	534	751
1869	1780	737	678	347	753
457	1554	523	245	600	810
369	336	656	336	556	589
721	277	767	295	507	849
852	972	725	174	591	959
487	1190	758	212	609	1057
1675	1046	327	228	618	1314
938	323	507	212	623	942
613	572	449	174	925	1175
638	715	420	269	1407	894
575	725	638	1055	1672	759
771	1065	589	131	1151	1059
641	1033	528	295	1035	1227
582	1033	469	302	1270	1689
170	1135	681	355	901	1361
749	884	785	607	1232	1329
633	453	482	376	1040	1318
480	870	634	345	1055	1256
503	896	787	661	888	921
540	705	416	340	578	921
672	1132	416	279	535	936
618	898	115	547	561	1558
632	1351	173	303	523	1036
691	1526	343	255	634	1508

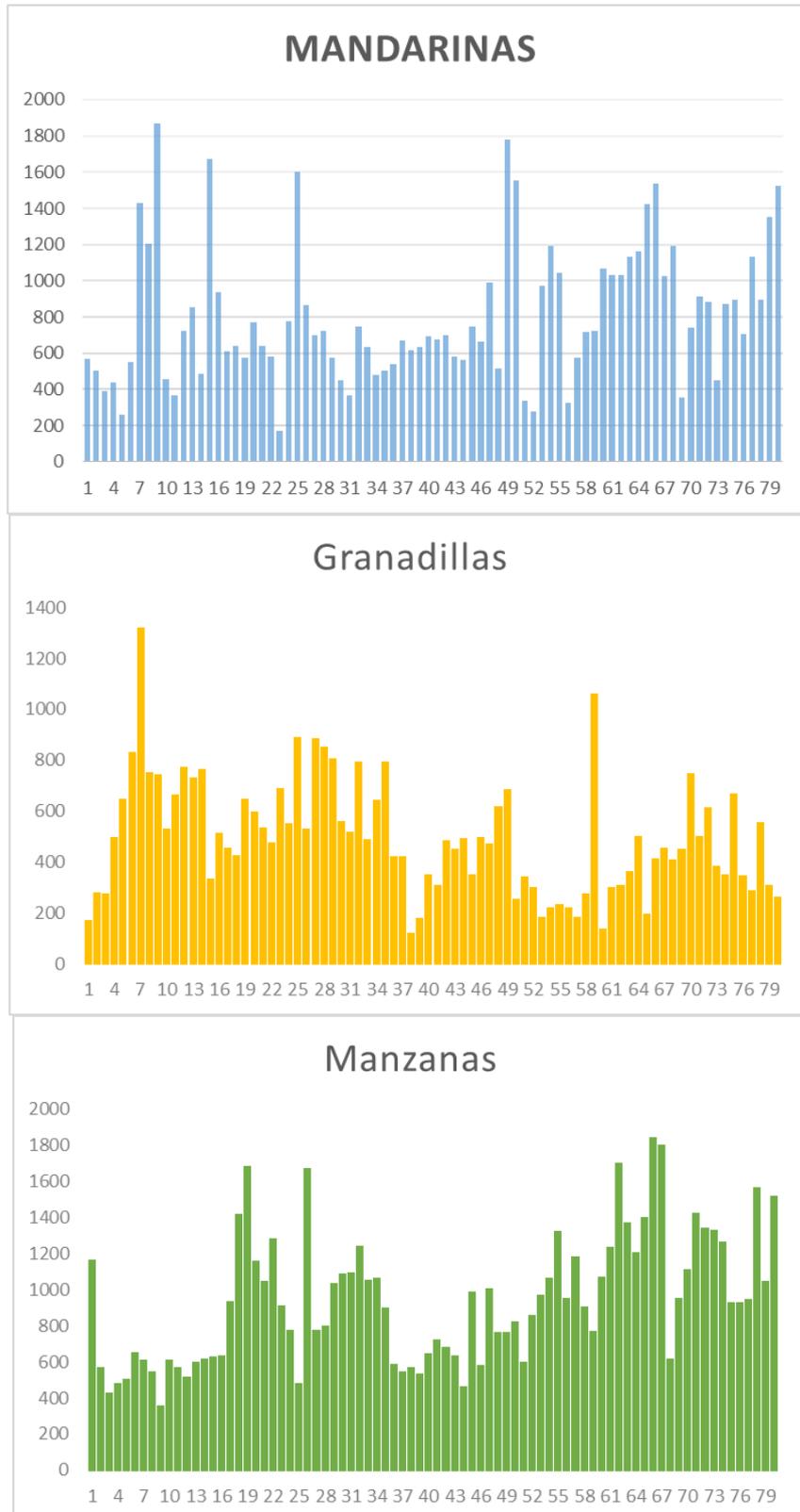


Figura 4.4 Gráficos de barras con las cantidades de puntos de interés encontrados con el método *SURF*

Tabla 19: Resultados estadísticos de la cantidad de puntos de interés encontrados con el método *SURF*

	Mandarinas	Granadillas	Manzanas
Promedio	806	488	932
Máximo	1869	1312	1672
Mínimo	170	115	347

De igual modo que los modelos *SIFT*, necesitamos conocer cómo la cantidad de puntos de interés afecta la construcción de los modelos *SURF* para predecir objetos del dominio; por ello, consideramos lo siguiente:

1. Construir 4 modelos considerando 50, 100, 150 y 200 puntos de interés en cada imagen. Si una imagen cuenta con menor cantidad de puntos, se consideran todos sus puntos de interés; caso contrario, se eligen aleatoriamente por cada imagen.
2. Analizar la precisión y exactitud de los modelos y cómo afecta la cantidad de puntos de interés elegidos.

En la Tabla 20, Tabla 21, Tabla 22 y Tabla 23, se observan los resultados del proceso de predicción; a partir de éstos podemos sacar las siguientes conclusiones:

- El modelo que utiliza 50 puntos de interés tiene la mayor precisión de predicción; mientras que el modelo con 100 puntos de interés, tiene menor precisión.
- Se observa que al incrementar los puntos de interés, tiende a disminuir la precisión del modelo.
- El modelo con 50 puntos de interés tiene la mayor precisión al predecir Mandarinas; sin embargo, ésta es baja: 15%.
- Todos los modelos predicen adecuadamente las Granadillas.
- Los modelos con 50, 150 y 200 puntos de interés tienen la mayor precisión al predecir Manzanas.
- El modelo con 50, 100 y 150 puntos de interés tiene la mayor exactitud al predecir Mandarinas. Por otro lado, el modelo de 200 puntos de interés no pudo clasificar correctamente a ninguna Mandarina.
- Los modelos con 50 y 200 puntos de interés tienen la mayor exactitud al predecir Granadillas y ambos predicen correctamente el grupo.
- El modelo con 50 puntos de interés tiene la mayor exactitud al predecir Manzanas y la mayor precisión.

Tabla 20: Precisión del modelo utilizando el algoritmo SURF con 50 keypoints

MODELO SURF CON 50 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	6	11	23	15
Granadillas	0	40	0	100
Manzanas	0	3	37	92,5
Exactitud (%)	100,00	74,07	61,67	69,17
Precisión del modelo (%)	69,17			

Tabla 21: Precisión del modelo utilizando el algoritmo SURF con 100 keypoints

MODELO SURF CON 100 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	1	17	22	2,5
Granadillas	0	40	0	100
Manzanas	0	7	33	82,5
Exactitud (%)	100,00	62,50	60,00	61,67
Precisión del modelo (%)	61,67			

Tabla 22: Precisión del modelo utilizando el algoritmo SURF con 150 keypoints

MODELO SURF CON 150 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	1	14	25	2,5
Granadillas	0	40	0	100
Manzanas	0	3	37	92,5
Exactitud (%)	100,00	70,18	59,68	65,00
Precisión del modelo (%)	65,00			

Tabla 23: Precisión del modelo utilizando el algoritmo SURF con 200 keypoints

MODELO SURF CON 200 KEYPOINTS

	Mandarinas	Granadillas	Manzanas	Precision (%)
Mandarinas	0	9	31	0
Granadillas	0	40	0	100
Manzanas	0	3	37	92,5
Exactitud (%)	#iDIV/0!	76,92	54,41	64,17
Precisión del modelo (%)	64,17			

4.2.4. Comparación de los modelos predictores *SIFT* y *SURF*

A continuación, presentamos la comparación de los modelos que se desarrollamos previamente en las secciones 4.2.2 y 4.2.3. La comparación se realizará sobre los siguientes puntos:

- Comparación de la precisión en los modelos que utilizan distintas cantidades de puntos de interés.
- Distinción de en qué modelos se obtiene la mejor precisión y exactitud en predecir cada uno de los Grupos (Mandarinas, Granadillas y Manzanas).
- Definir el modelo final que tenga la mayor precisión y al mismo tiempo la mayor exactitud al predecir los objetos.

Las comparaciones de los resultados de los métodos pueden observarse en la Tabla 24, Tabla 25 y Tabla 26 y desde la Figura 4.5 a la Figura 4.10.

Tabla 24: Tabla comparativa de la precisión y la exactitud para distintas cantidades de puntos de interés utilizando el método *SIFT* y *SURF* a predecir Mandarinas

CANTIDAD DE PUNTOS	Modelos para predecir Mandarinas			
	MODELO SIFT		MODELO SURF	
	PRECISIÓN (%)	EXACTITUD (%)	PRECISIÓN (%)	EXACTITUD (%)
50	87,5	85,366	15	100
100	75	93,75	2,5	100
150	40	88,889	2,5	100
200	45	90	0	0

Tabla 25: Tabla comparativa de la precisión y la exactitud para distintas cantidades de puntos de interés utilizando el método *SIFT* y *SURF* a predecir Granadillas

CANTIDAD DE PUNTOS	Modelos para predecir Granadillas			
	MODELO SIFT		MODELO SURF	
	PRECISIÓN (%)	EXACTITUD (%)	PRECISIÓN (%)	EXACTITUD (%)
50	92,5	86,047	100	74,074
100	85	80,952	100	62,5
150	67,5	100	100	70,175
200	45	100	100	76,923

Tabla 26: Tabla comparativa de la precisión y la exactitud para distintas cantidades de puntos de interés utilizando el método *SIFT* y *SURF* a predecir Manzana

CANTIDAD DE PUNTOS	Modelos para predecir Manzanas			
	MODELO SIFT		MODELO SURF	
	PRECISIÓN (%)	EXACTITUD (%)	PRECISIÓN (%)	EXACTITUD (%)
50	77,5	86,111	92,5	61,667
100	81,67	73,913	82,5	60
150	100	53,333	92,5	59,677
200	100	48,78	92,5	54,412

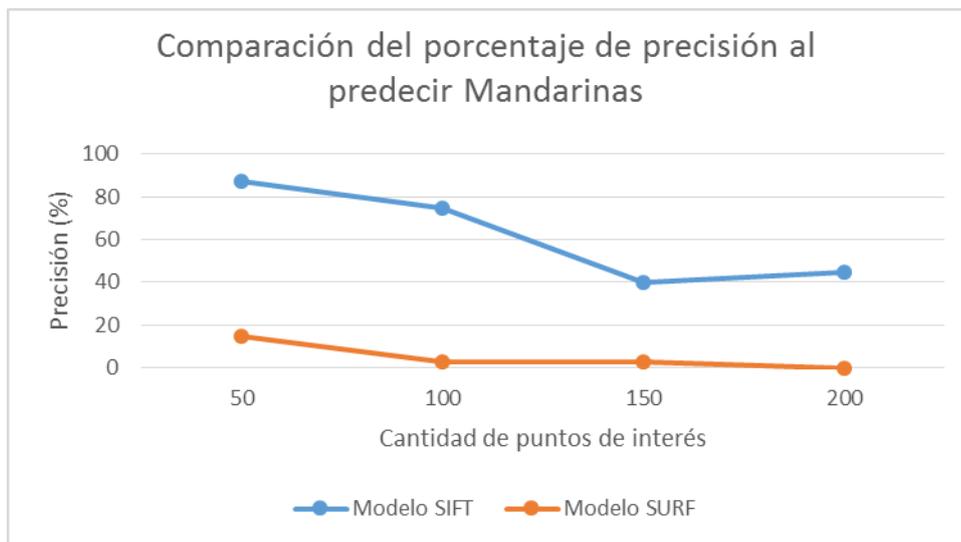


Figura 4.5 Gráfico del porcentaje de precisión al predecir Mandarinas

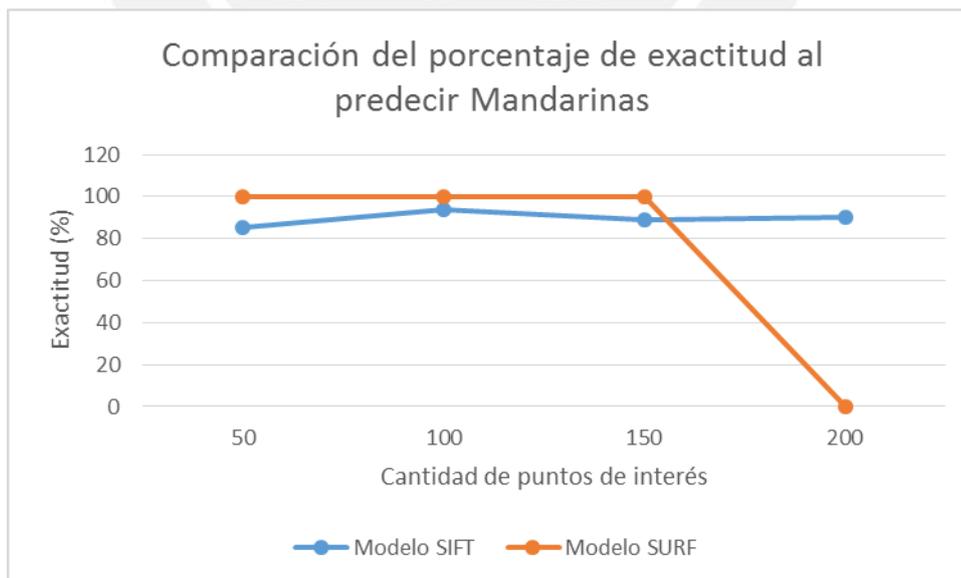


Figura 4.6 Gráfico del porcentaje de exactitud al predecir Mandarinas

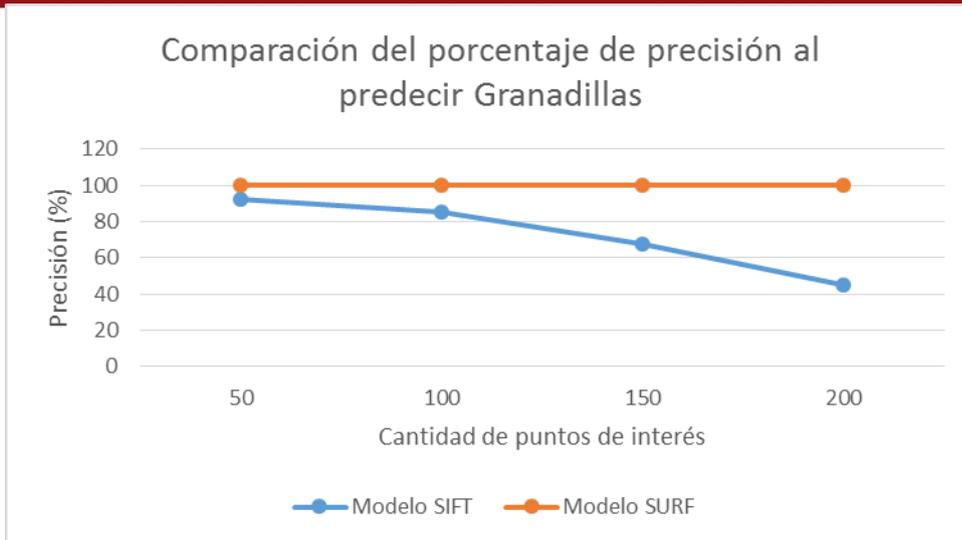


Figura 4.7 Gráfico del porcentaje de precisión al predecir Granadillas

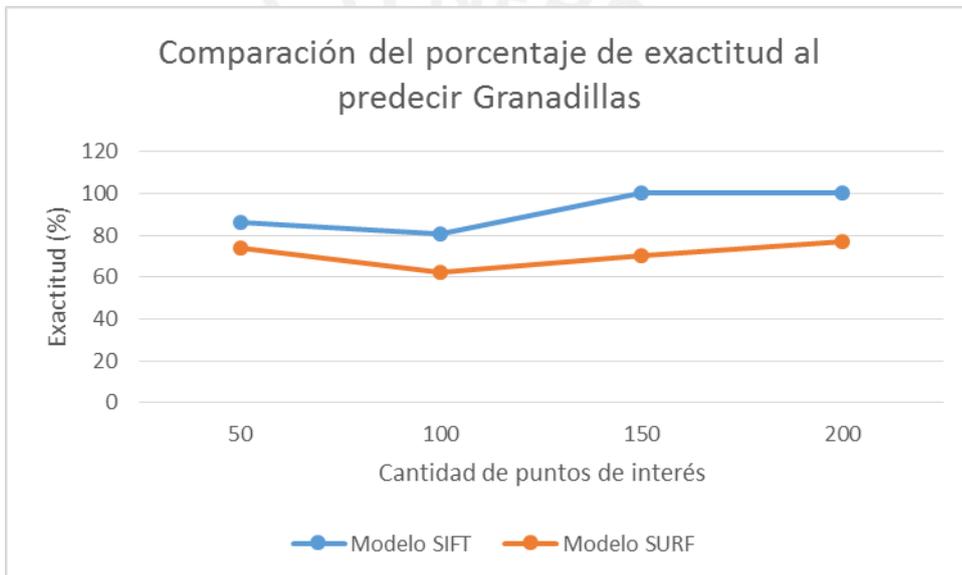


Figura 4.8 Gráfico del porcentaje de exactitud al predecir Granadillas

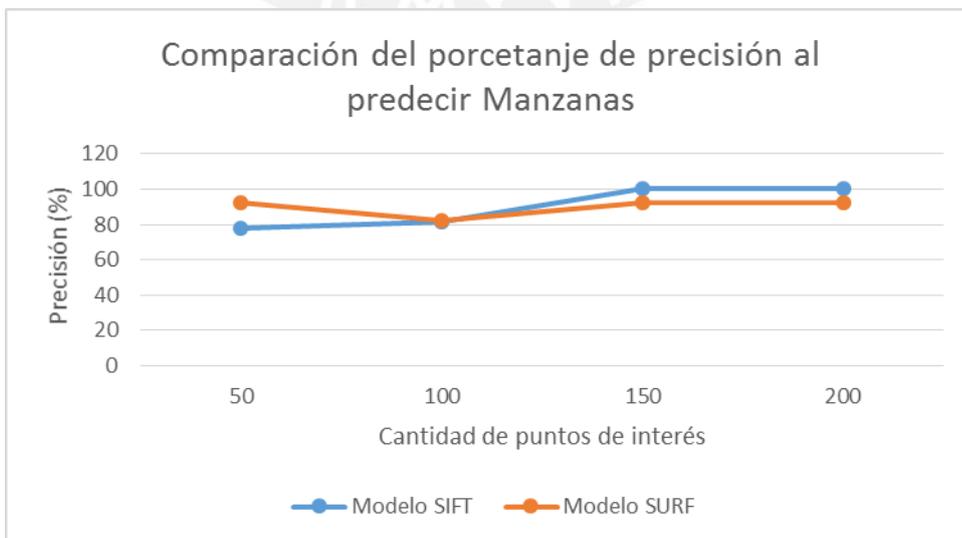


Figura 4.9 Gráfico del porcentaje de precisión al predecir Manzanas

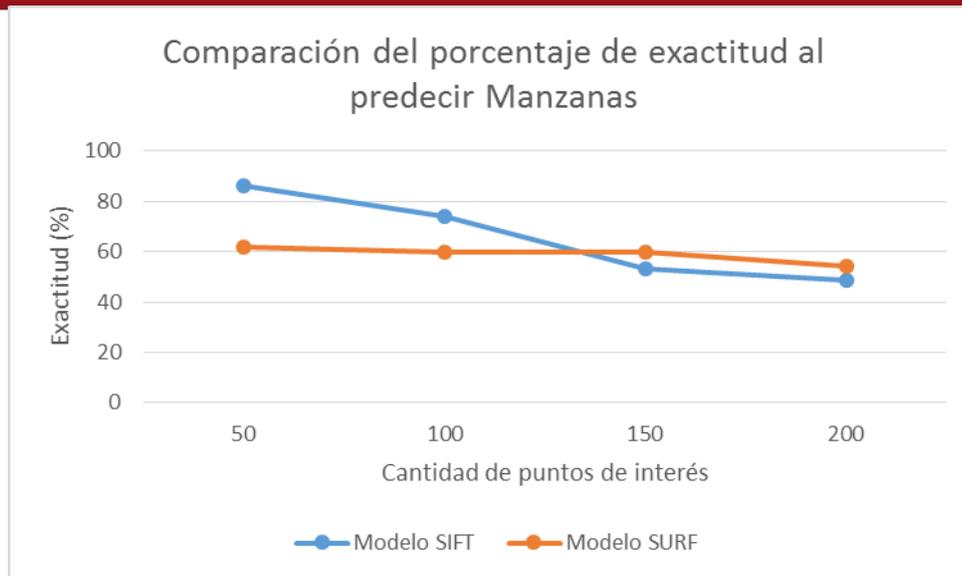


Figura 4.10 Gráfico del porcentaje de exactitud al predecir Manzanas

Los resultados de la comparación pueden resumirse en:

1. Para la predicción de Mandarinas, los modelos *SIFT* con 50 y 100 puntos de interés tienen una alta precisión y a su vez, alta exactitud.
2. Para la predicción de Granadillas, el modelo *SIFT* con 50 puntos de interés y los modelos *SURF* con 150 y 200 puntos de interés tienen alta precisión y alta exactitud.
3. Para la predicción de las Manzanas, los modelos *SIFT* con 50 y 100 puntos de interés y los modelos *SURF* con 50 y 150 puntos de interés tienen alta precisión y alta exactitud.
4. Los modelos *SURF* de 50, 100, 150 y 200 puntos de interés tienen las menores precisiones para predecir Mandarinas.
5. Los modelos *SIFT* de 50 y 100 puntos de interés (Tabla 25) tienen las mayores precisiones y a su vez buenas exactitudes al predecir los tres tipos de objetos.

5. CONCLUSIONES

A continuación se presenta las conclusiones del proyecto en función de los objetivos que planteamos al inicio.

Conclusiones

De la implementación del método SIFT y SURF en un conjunto de objetos y comparar sus fortalezas en función de los resultados:

- En la sección 3.1, los resultados obtenidos indican que los puntos de interés *SIFT* tienen dimensión vectorial de 1×128 ; asimismo, los valores de los puntos varían entre $[0, 255]$.
- En la sección 3.2, los resultados obtenidos indican que los puntos de interés SURF tienen dimensión vectorial de 1×64 ; asimismo, los valores de los puntos varían entre $[-1, 1]$.
- La información que proporciona el método *SIFT* acerca de los puntos de interés de una imagen es mayor por la dimensión (aproximadamente el doble de valores) y por ende más detallada. Sin embargo, esto último no significa que permite una mejor clasificación.
- En la sección 3.4, se definió como nuestro dominio de objetos a tres tipos de frutas: Mandarinas, Granadillas y Manzanas.
- De la primera comparación, en condiciones normales, obtuvimos que el método *SIFT* generaba más puntos de interés que el método *SURF*.
- En la segunda comparación, se oscureció las imágenes y se obtuvo que el método SURF generaba más puntos de interés, y a su vez, el tiempo requerido para los cálculos era menos que *SIFT*.
- En la tercera comparación, se aumentó la iluminación de las imágenes y se obtuvo que el método *SIFT* producía más puntos de interés; y además, el tiempo para realizar los cálculos era también mayor.

De la implementación del clasificador probabilístico para determinar la categoría de un objeto utilizando la información de los métodos SIFT y SURF:

- En la sección 4, se desarrollaron 3 modelos de predicción para cada método (*SIFT* y *SURF*), utilizando un clasificador bayesiano.
- De la comparación de los resultados, obtuvimos un modelo clasificador del método *SIFT* utilizando 50 puntos de interés de cada imagen de la etapa de entrenamiento. La precisión de clasificación fue de 85.83%.
- Incrementar la cantidad de puntos de interés no siempre aumentaba la precisión de los clasificadores.



6. TRABAJO FUTUROS

Durante el proyecto, se consideró que las imágenes no debían tener ningún otro tipo de imagen en el fondo que nos generase falsos puntos de interés durante la predicción; una mejora que se propone es:

- Incluir métodos de detección de regiones para definir los límites de un objeto y considerar sólo los puntos de interés en dichas regiones.
- Dispositivos como el *Kinect*, permiten detectar regiones de acuerdo a la lejanía de un objeto respecto de la cámara. Esto podría ser útil para facilitar la interacción con el entorno; ya que, las imágenes observadas por la cámara no solo serían 2D, sino que también tendrían la característica de proximidad.

Para mejorar la interacción del robot con el entorno, la captura y proceso de predicción de los objetos debe ser más rápida. Sería adecuado investigar más técnicas de predicción de imágenes además de los métodos *SIFT* y *SURF*; algunos métodos que se han propuesto son el *PCA-SIFT* de Ke, Y., & Sukthankar, R. [29], que es una técnica basada en el mismo procedimiento del *SIFT*, pero que reduce el tamaño de la dimensión de los puntos de interés.

7. Referencias Bibliográficas

- [1] T. Huang, «Computer vision: evolution and promise. In Imaging science and technology, Evolution and promise,» de *5th International conference on high technology*, 1996.
- [2] Q. Ji, «ECSE @ Rensselaer,» 2 Febrero 2015. [En línea]. Available: https://www.ecse.rpi.edu/Homepages/qji/CV/3dvision_intro.pdf. [Último acceso: 15 Febrero 2016].
- [3] M. E. Munich, P. Pirjanian, E. Di Bernardo, L. Goncalves, N. Karlsson y D. Lowe, «SIFT-ing through features with ViPR,» 2006.
- [4] R. C. Gonzalez y R. E. Woods, *Digital image processing*, 2002.
- [5] D. G. Lowe, «Object recognition from local scale-invariant features.,» *Computer Vision, 1999. The proceeding of the seventh IEEE international conference on*, vol. II, pp. 1150-1157, 1999.
- [6] H. Bay, T. Tuytelaars y L. Van Gool, «SURF: Speeded up Robust features,» *Computer Vision - ECCV 2006*, pp. 404-417, 2006.
- [7] M. L. X. D. J. & L. C. Liu, «Generic object recognition based on the fusion of 2D and 3D SIFT descriptors,» de *Information Fusion (Fusion), 18th International Conference*, 2015.
- [8] L. & W. T. Bo, «A SIFT-Color Moments Descriptor for Object Recognition,» de *IT Convergence and Security*, 2014.
- [9] a. A. P. Vignesh Ramanathan, «Active object categorization on humanoid platform,» de *Vignesh Ramanathan, and Axel Pinz*, Portugal, 2011.
- [10] M. W. J. L. J. C. H. C. G. F. J. Du, «Robot robust object recognition based on fast SURF feature matching.,» de *Chinese Automation Congress*, 2013.
- [11] T. Lozano-Perez, *Autonomous Robot Vehicles*, G. T. W. Ingemar J. Cox, Ed., Springer Science & Business Media, 2012.
- [12] T. Smithers, *Autonomy in robots and other agents*, 1997, pp. 88-106.
- [13] D. Floreano, J. Godjevac, A. Martinoli, F. Mondada y J. D. Nicoud, «Design, control, and applications of autonomous mobile robots,» *Advances in Intelligent Autonomous Systems*, pp. 159-186, 1999.
- [14] E. Freund y J. Rossmann, «Intelligent autonomous robots for industrial and Space applications,» *Intelligent Robots and Systems' 94. Advanced Robotic System and the Real World'*, vol. II, pp. 1072-1079, 1994.
- [15] N. Gryaznov y A. Lopota, «Computer Vision for Mobile On-Ground Robotics,» *Procedia Engineering*, pp. 1376-1380.
- [16] V. P. Carlini, «The Object Recognition Task: A New Proposal for the Memory Performance Study.,» *INTECH Open Access Publisher*, 2011.
- [17] K. Grauman y B. Leibe, *Visual object recognition. Synthesis lectures on artificial intelligence and machine learning*, 2011.
- [18] A. Andreopoulos y J. K. Tsotsos, «50 Years of object recognition: Directions forward,» *Computer Vision and Image Understanding*, pp. 827-891, 2013.
- [19] C. Goad, «Special purpose automatic programming for 3D model-based vision,» *Computer Vision*, pp. 371-381, 1987.
- [20] R. Bolles y R. Horaud, «3DPO: a three-dimensional part orientation system,» *International Journal of Robotics Research*, 1986.
- [21] A. Veldadi, B. Fulkerson, K. Lenc, D. Perrone, M. Perdoch, M. Sulc y H. Sabotova, «Documentación C-API,» [En línea]. Available: <http://www.vlfeat.org/api/sift.html>. [Último acceso: 13 Abril 2015].

- [22] U. Sinha, «AI Shack,» [En línea]. Available: <http://aishack.in/tutorials/sift-scale-invariant-feature-transform-introduction/>. [Último acceso: 09 03 2016].
- [23] A. E. Abdel-Hakim y A. Farag, «CSIFT: A SIFT descriptor with color invariant characteristics,» *Computer Vision and Pattern Recognition.*, vol. II, pp. 1978-1983, 2006.
- [24] H. Zhou, Y. Yuan y C. Shi, «Object tracking using SIFT features and mean shift,» *Computer vision and image understanding*, pp. 345-352, 2009.
- [25] J. Hong, W. Lin, H. Zhang y L. Li, «Image mosaic based on SURF feature matching,» *Information Science and Engineering*, pp. 1287-1290, 2009.
- [26] A. Milella, B. Nardelli, D. Di Paola y G. Cicirelli, «Robust Feature Detection and Matching for Vehicle Localization in Uncharted Environments,» *Proceeding of the IEEE/RSJ IROS Workshop Planning, Perception and Navigation for Intelligent Vehicles*, 2009.
- [27] A. Carcedo y Franco, Programa de segmentación de regiones en imágenes médicas en MATLAB, 2004.
- [28] L. Juan y O. Gwun, «A comparision of sift, pca-sift and surf.,» *International Journal of Image Processing*, pp. 143-152, 2009.
- [29] Y. Ke y R. Sukthankar, «PCA-SIFT: A more distinctive representation for local image descriptors.,» de *Proceedings of the 2004 IEEE Computer Society Conference*, 2004.
- [30] L. Wang y W. Gio, «James Z. Wang Research Group,» [En línea]. Available: <http://wang.ist.psu.edu/docs/related/>. [Último acceso: 15 Abril 2015].
- [31] F. Fang y S. He, «Cortical responses to invisible objects in the human dorsal and ventral pahtway,» vol. 8, n° 10, 2005.
- [32] S. Edelman, «Computational theories of object recognition,» *Trends in cognitive sciences*, pp. 296-304, 1997.
- [33] P. F. Felzenswalb, R. B. Girshck, D. McAllester y D. Ramanan, «Object deteection with discrminatively trained part-based models.,» *Pattern Analysis and Machine Intelligence, IEEE Transactions*, pp. 1627-1645, 2010.
- [34] Y. Hirano, C. Garcia, R. Sukthankar y A. Hoogs, «Industry and object recognition: Applications, applied research and challenges.,» *In Toward Category-Level Object Recognition*, pp. 49-64, 2006.
- [35] J. J. Z. D. & R. N. C. DiCarlo, «How does the brain solve visual object recognition?,» *Neuron*, 2012.