



UNIVERSIDAD NACIONAL DE COLOMBIA

Diseño de un sistema que determine regiones de agarre de objetos cilíndricos en un entorno semi-estructurado basado en visión

Daniel Francisco Ochoa Morón

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica
Bogotá D.C., Colombia
2024

Diseño de un sistema que determine regiones de agarre de objetos cilíndricos en un entorno semi-estructurado basado en visión

Daniel Francisco Ochoa Morón

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magister en Ingeniería - Automatización Industrial

Director(a):

PhD. Victor Hugo Grisales Palacio

Co-director(a):

PhD. Pedro Fabian Cardenas Herrera

Línea de Investigación:

Robótica Industrial y Grasping

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería Eléctrica y Electrónica

Bogotá D.C., Colombia

2024

Agradecimientos

A mi familia: mi Padre, mi Madre y mis Hermanos. A mis tíos(as), primos(as), amigos(as) y allegados, a todas esas personas que siempre han estado presentes en mi formación como persona, que creen en mi y aprecian el proceso educativo.

A mis Directores, educadores, y todas aquellas personas que directa o indirectamente siempre estuvieron allí presentes impulsándome a avanzar.

A la educación pública, que ha sido la formadora en todas las etapas recorridas de mi formación como profesional.

Gracias

Declaración de obra original

Yo declaro lo siguiente:

He leído el Acuerdo 035 de 2003 del Consejo Académico de la Universidad Nacional de Colombia. «Reglamento sobre propiedad intelectual» y la Normatividad Nacional relacionada al respeto de los derechos de autor. Esta disertación representa mi trabajo original, excepto donde he reconocido las ideas, las palabras, o materiales de otros autores.

Cuando se han presentado ideas o palabras de otros autores en esta disertación, he realizado su respectivo reconocimiento aplicando correctamente los esquemas de citas y referencias bibliográficas en el estilo requerido.

He obtenido el permiso del autor o editor para incluir cualquier material con derechos de autor (por ejemplo, tablas, figuras, instrumentos de encuesta o grandes porciones de texto).

Por último, he sometido esta disertación a la herramienta de integridad académica, definida por la universidad.



.....
Nombre

Resumen

Diseño de un sistema que determine regiones de agarre de objetos cilíndricos en un entorno semi-estructurado basado en visión

La presente tesis de maestría-investigación se focaliza en el desarrollo de un sistema destinado a determinar las regiones de agarre de objetos cilíndricos, específicamente botellas plásticas, en un entorno semiestructurado utilizando visión por computadora. A pesar de la diversidad de formas, tamaños y colores que presentan las botellas, se asume un tamaño promedio de 500 ml para la investigación. El proyecto tiene como objetivo abordar desafíos en la manipulación robótica y la automatización, especialmente en aplicaciones industriales. Se inicia con la creación de un banco de imágenes que sirve como base para un sistema de procesamiento de imágenes, el cual, junto con herramientas de inteligencia artificial, permite entrenar una red neuronal específica para la tarea de agarre.

La presente investigación profundiza en los métodos y tecnologías utilizados en la planificación de agarre y la manipulación robótica, destacando el uso de técnicas de aprendizaje profundo. El documento se encuentra organizado en capítulos que abarcan el contexto de la investigación, la motivación, el trabajo relacionado, los objetivos específicos y el desarrollo del sistema para la generación automática de regiones de agarre basado en visión por computadora y aprendizaje automático.

El marco del desarrollo de la presente investigación, se centró en el análisis físico de un número determinado de objetos dispuestos en escena y las características físicas y funcionales de un gripper de dos dedos empleado para ejecutar una tarea de agarre específica. A partir de un sistema de percepción visual bidimensional ajustado y la extracción de características geométricas de los objetos, se diseñó e implementó un sistema algorítmico capaz de establecer regiones de agarre a lo largo de los objetos empleados. Posteriormente, se estableció un número de parámetros de evaluación heurísticos con el objetivo de determinar la viabilidad de cada una de las regiones encontradas sobre cada objeto en relación a su espacio circundante.

Palabras clave: Agarre robótico, aprendizaje profundo, manipulador, visión de máquina.

Abstract

Design of a system that determines grasping regions of cylindrical objects in a semi-structured vision-based environment

This master's thesis focuses on the development of a system aimed at determining the grasping regions of cylindrical objects, specifically plastic bottles, in a semi-structured environment using computer vision. Despite the diversity of shapes, sizes, and colors that bottles present, an average size of 500 ml is assumed for the research. The project aims to address challenges in robotic manipulation and automation, especially in industrial applications.

It begins with the creation of an image bank that serves as the basis for an image processing system, which, along with artificial intelligence tools, allows the training of a specific neural network for the grasping task. This research delves into the methods and technologies used in grasp planning and robotic manipulation, highlighting the use of deep learning techniques. The document is organized into chapters that cover the context of the research, the motivation, related work, specific objectives, and the development of the system for the automatic generation of grasping regions based on computer vision and machine learning.

In the framework of the development of this research, the focus was on the physical analysis of a determined number of objects arranged in the scene and the physical and functional characteristics of a two-finger gripper used to perform a specific grasping task. Based on an adjusted two-dimensional visual perception system and the extraction of geometric characteristics of the objects, an algorithmic system was designed and implemented to establish grasping regions along the employed objects. Subsequently, a number of heuristic evaluation parameters were established to determine the feasibility of each of the regions found on each object in relation to its surrounding space.

Keywords: Deep learning, machine vision, manipulator, robotic grasping,

Contenido

Agradecimientos	III
Declaración de Obra Original	IV
Resumen	v
1 Planteamiento de la investigación	2
1.1 Contexto	2
1.2 Motivación	2
1.3 Organización del documento	3
1.4 Trabajos relacionados	4
1.4.1 Planificación de agarre	5
1.4.2 Agarre robótica	8
1.5 Justificación del problema	9
1.6 Objetivos	9
1.6.1 Objetivo general.	9
1.6.2 Objetivos específicos.	10
2 Generación del banco de datos	11
2.1 Introducción	11
2.2 Escenarios de adquisición de imágenes	11
2.2.1 Escenario para capturas fotográficas I	12
2.2.2 Escenario para capturas fotográficas II	13
2.3 Implementación	14
2.3.1 Captura individual de las botellas	16
2.3.2 Captura conjunta de las botellas	17
3 Sistemas de percepción (visión por computador)	18
3.1 Introducción	18
3.2 Procesamiento de Imágenes Digitales	19
3.2.1 Imágenes digitales	19
3.2.2 Modelo de color de escala de grises	21
3.2.3 Filtrado de imágenes	22
3.2.4 Modelo de color HSV (Hue-Saturation-Value)	32
3.3 Segmentación de imágenes digitales	35
3.3.1 Análisis de histogramas	36
3.3.2 Umbralización o thresholding	38

3.3.3	Transformaciones morfológicas	42
3.3.3.1	Dilatación (Dilate)	42
3.3.3.2	Erosión (Erode)	42
3.3.3.3	Apertura (Opening)	43
3.3.3.4	Cierre (Closing)	43
3.3.4	Detección de bordes y contornos	43
4	Sistema automático de reconocimiento de imágenes	46
4.1	Introducción	46
4.2	Segmentación de imágenes	46
4.3	Segmentación agnóstica	47
4.4	Implementación	48
5	Generación Automática de Regiones de Agarre (GARA)	51
5.1	Introducción	51
5.2	Generación y análisis del rectángulo delimitador	51
5.3	Generación de las regiones de agarre	54
5.3.1	Implementación	57
5.4	Parámetros de decisión	58
5.4.1	Parámetro I: área del objeto	59
5.4.2	Parámetro II: ubicación espacial del objeto	60
5.4.3	Parámetro III: evaluación espacial de las regiones de agarre generadas	61
5.4.4	Parámetro IV: evaluación de las regiones de agarre en el objeto	64
6	Validación	65
6.1	Introducción	65
6.2	Configuración y ejecución del contenedor docker ROS-Noetic	66
6.3	Integración del robot ABB IRB 140	66
6.3.1	Configuración del TCP	67
6.3.2	Transformación de coordenadas	67
6.3.2.1	Coordenadas de traslación de la cámara respecto a la base del robot	67
6.3.3	Matrices de Transformación	68
6.3.3.1	Matriz de Traslación	68
6.3.3.2	Matriz de Rotación	69
6.3.3.3	Parámetros Denavit-Hartenberg (DH)	69
6.3.3.4	Calibración de cámara	70
6.4	Implementación	72
7	Conclusiones	76
	Bibliografía	79

1 Planteamiento de la investigación

1.1. Contexto

La problemática del agarre y la manipulación en robótica parte de la premisa de contar con un modelo ideal del objeto, así como una estimación relativa del mismo en el espacio y la configuración de la mano del manipulador.

En la actualidad, diversos estudios abordan el desafío de la captura de objetos, muchos de los cuales emplean algoritmos de simplificación geométrica. Estos algoritmos facilitan el análisis, procesamiento de datos y la generación de hipótesis de agarre, basándose en el análisis de formas más primitivas.

Este documento presenta la tesis de maestría titulada "Diseño de un Sistema para Determinar Regiones de Agarre de Objetos Cilíndricos en un Entorno Semi-Estructurado Basado en Visión". La propuesta se centra en la captura de objetos cilíndricos en un entorno semi-estructurado, utilizando botellas de plástico PET como ejemplo. El enfoque simula un proceso de reciclaje, considerando las características físicas de los objetos y el entorno circundante.

Es crucial resaltar que esta investigación se sitúa en el contexto de la robótica aplicada, abordando un desafío significativo en la automatización industrial. El desarrollo de un sistema capaz de identificar regiones de agarre en objetos cilíndricos en un entorno variable y poco estructurado representa un avance considerable en la capacidad de los robots para interactuar eficaz y eficientemente con el mundo real. Esta aproximación tiene como objetivo mejorar la forma en que se enfrenta la problemática de la manipulación de objetos en diversas aplicaciones industriales.

1.2. Motivación

Un rectángulo de demarcación, posicionado sobre una región específica de un objeto, puede considerarse como una representación hipotética de una región de agarre. Esta metodología de aproximación ha sido ampliamente utilizada en los últimos años debido a su uso de pocos parámetros y su alta intuición. Sin embargo, aún presenta numerosas deficiencias que resultan en inconsistencias al seleccionar el rectángulo de aproximación más pertinente.

En el ámbito actual, los desafíos en la robótica de agarre plantean una serie de interrogantes: ¿Cómo realizar el agarre? ¿Dónde realizar el agarre? Aunque se ha abordado en la literatura exis-

tente, persisten limitaciones que podrían ser complementadas para mejorar la comprensión de estos aspectos.

1.3. Organización del documento

El presente documento, se encuentra organizado en VII capítulos temáticos que contemplan el desarrollo del estudio alrededor de un sistema de generación automática de regiones de agarre basado en visión por computadora y aprendizaje computacional.

En el capítulo 1 (**Planteamiento de la investigación**), se presentan un contexto, la motivación, la organización del documento, el estado de conocimiento, los objetivos propuestos y las contribuciones alcanzadas a lo largo del desarrollo del presente trabajo.

En el capítulo 2 (**Generación de la base de datos**), se detalla el proceso llevado a cabo para adquirir las imágenes digitales, comenzando con la creación y configuración de dos escenarios para la captura de fotografías. Se describen los criterios considerados de los escenarios propuestos para obtención de imágenes limpias, procurando minimizar la presencia de sombras y capturando los objetos en diversas poses.

En el capítulo 3 (**Sistemas de percepción - visión por computador**), se inicia con un análisis del procesamiento de imágenes digitales (secc. 3.2) adquiridas previamente, utilizando transformaciones en los modelos de color y aplicando técnicas de filtrado tanto en el dominio espacial como en el dominio frecuencial. Luego, se describe la etapa de segmentación (secc. 3.3), que se basa en la similitud y emplea algoritmos para obtener umbrales a partir del análisis de histogramas en el modelo de color HSV, lo que resulta en una imagen binarizada. Finalmente, se lleva a cabo la detección de bordes y contornos de los objetos de interés.

En el capítulo 4, (**Sistema automático de reconocimiento de imágenes**), se desarrolló un estudio sobre un sistema de reconocimiento automático de objetos en escenas multiclase, denominado segmentación agnóstica. En este sistema, cada conjunto de objetos con características similares pertenece a una clase específica. La implementación de este sistema permite obtener una segmentación robusta de los objetos en escenarios complejos, devolviendo los límites de demarcación o contornos de cada objeto capturado en la escena.

En el capítulo 5 (**Generación Automática de Regiones de Agarre - GARA**), se presenta una metodología para la generación automática de regiones de agarre (GARA) mediante el análisis geométrico de los objetos estudiados. Esta metodología comienza con la simplificación geométrica utilizando formas primitivas y considerando propiedades del efector final de dos dedos empleado. Luego, se describe un árbol de decisión heurístico desarrollado para seleccionar opciones viables de agarre entre el conjunto de regiones de agarre identificadas.

En el capítulo 6 (**Validación**), se lleva a cabo un proceso exhaustivo de validación del sistema propuesto empleando del robot industrial ABB IRB 140 en el laboratorio LABSIR de la Universidad Nacional de Colombia. En este capítulo, se detalla el cálculo de las transformaciones de coordenadas del sistema servo-visual utilizado y se describe la implementación de la librería de comunicación ROS-ABB denominada "LABSIR Bin Picking ROS Simulation packages". Además, se realizan simulaciones destinadas a determinar la adecuación de las trayectorias generadas mediante la ejecución de un algoritmo RRT en la librería de simulación MoveIt, basadas en las propiedades obtenidas de la posición en las regiones de agarre calculadas.

En el capítulo 7 (**Conclusiones**), se presenta el marco conclusivo derivado de los análisis y resultados obtenidos a lo largo del presente documento. Se destacan las ventajas y desventajas encontradas en cada capítulo, resumiendo las contribuciones principales del trabajo. Además, se plantean posibles direcciones para futuras investigaciones, basadas en las preguntas o ideas que han quedado pendientes de explorar o analizar en profundidad. Este capítulo cierra el documento proporcionando una visión global de los hallazgos y señalando el camino hacia posibles desarrollos y avances en el campo abordado.

1.4. Trabajos relacionados

El rápido desarrollo de la investigación robótica ha permitido a los robots trabajar de forma autónoma en entornos estructurados y no estructurados; no obstante, la incertidumbre presente en ambientes no estructurados, hoy en día aún representa un gran desafío para los manipuladores debido a la complejidad de computo [10, 57]. Bohg et. al. [6] afirman que el agarre en robótica implica encontrar un conjunto de puntos de contacto adecuados para las regiones táctiles del efector final y la configuración del conjunto de articulaciones de brazo y muñeca que permitan la realización de una tarea específica.

Para realizar la captura de un objeto de interés, es necesario implementar una configuración de agarre factible, que permita generar trayectorias idóneas que cumplan con las restricciones y conduzcan al *gripper* a realizar una captura óptima del objeto específico [10]. La principal dificultad presente en la captura de objetos, se debe a los cambios cinemáticos y la dinámica presente en el entorno [10].

La implementación de tareas **reach-to-grasp (RTG)** (véase Figura 1-1), se encuentra conformada de tres aspectos principales: **reconocimiento de objetos**, **planificación de agarre** (secc. 1.4.1) y **generación de trayectorias** [10].

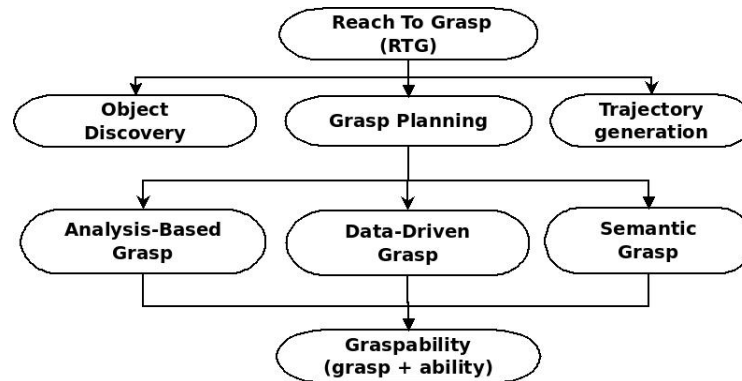


Figura 1-1: Llegar a agarrar (Reach-To Grasp(RTG))

1.4.1. Planificación de agarre

La planificación de agarre, es generalmente concebida como un *problema de optimización restringida*. En la búsqueda de ofrecer posibles soluciones al presente problema, se han propuesto diversidad de métodos que en su mayoría emplean optimización numérica [10, 30] o algoritmos de búsqueda [10, 11]. Sin embargo, estos métodos presentan principalmente dos inconvenientes [10]: primero, se basan en modelos de objetos ideales, de buena comprensión y agarre, por lo tanto, no son aptos en entornos no estructurados donde hay existencia de incertidumbres significativas: elementos desconocidos, obstrucciones de visibilidad, entre otros [10]. El segundo inconveniente parte del uso de una métrica de calidad para evaluar el desempeño de cada agarre, que resulta de difícil construcción y de bajo rendimiento en entornos no estructurados [10].

El proceso de captura robótica de objetos, es hoy en día uno de los principales temas de investigación. Varios autores dividen los métodos de planificación de agarre existentes en dos categorías: **métodos basados en análisis** y **métodos basados en datos** [6, 10, 14]. Eizicovits et. al. [14] afirman que cada uno de los métodos de planificación de agarre conocidos deben ser estudiadas desde los dos (2) entornos disponibles (véase Figura 1-2): **estructurado** y **no estructurado o dinámico**. [6, 14] agrupan los enfoques dependiendo del conocimiento a priori del objeto de interés en: **objetos conocidos, familiares** y **desconocidos**. Los **entornos estructurados** (véase Figura 1-2) se presentan en un ambiente altamente conocido donde existe muy poca incertidumbre en la escena estudiada [6, 14]. La información de los objetos puede adquirida por análisis morfológico debido a que presentan escasa o ninguna oclusión, lo que se ve representado en la obtención de modelos ideales de estos y al cálculo simplificado de las variables pertinentes para la ejecución de una captura asertiva, permitiendo evaluar el rendimiento de agarres a partir de una métrica de calidad [6, 10, 14].

Los **entornos no estructurados o dinámicos** (véase Figura 1-2) describen espacios de acción con un porcentaje considerable de incertidumbre, la información de los objetos puede no ser adquirida de una forma clara y precisa debido a un posible grado de oclusión y/o factores presentes en el medio causantes de interpretaciones erróneas de la escena [6, 14]. El objeto de interés puede

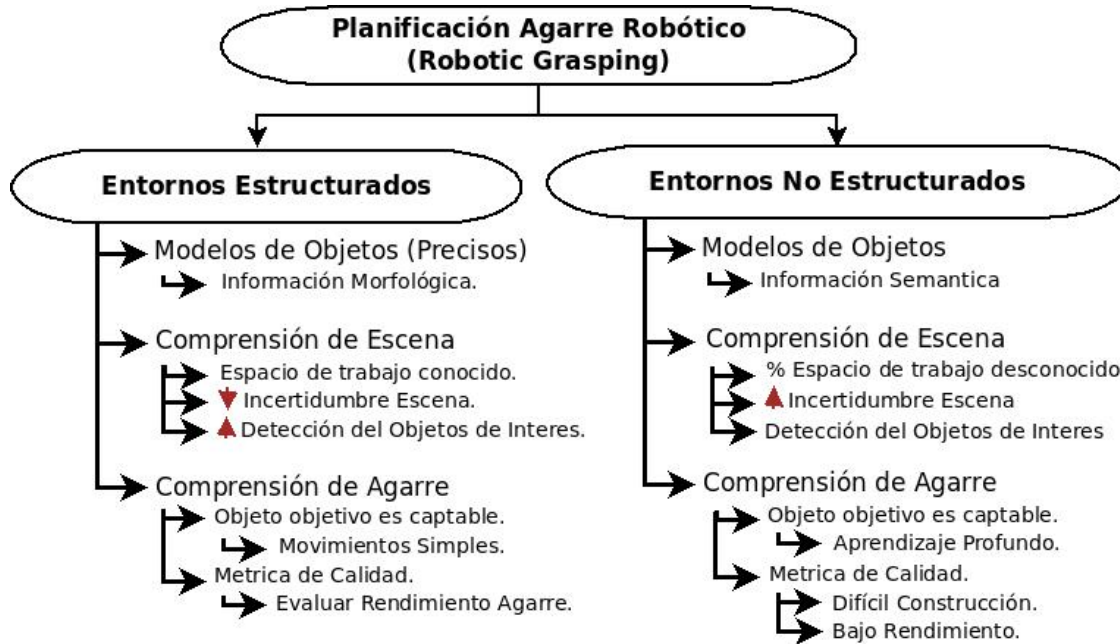


Figura 1-2: Entornos de estudio en la planificación de agarre

presentar posiciones variables debido a la dinámica del entorno, comprometiendo la exactitud de los cálculos realizados para la ejecución de un agarre; haciendo necesario la implementación de métodos robustos en la detección y la planificación de agarre [6, 10].

Siguiendo el criterio de agrupación propuesto por [6] y [14], se centra en el conocimiento a priori del objeto consultado. Los **objetos conocidos** constituyen este grupo, compuesto por elementos de interés previamente aprendidos, para los cuales ya existe una base de datos de experiencias generada fuera de línea. En este contexto, se pueden emplear configuraciones de agarre precalculadas, como se ha abordado en investigaciones previas [19, 27]. En estos casos, durante el tiempo de ejecución, solo es necesario calcular la ruta de aproximación, según lo indicado por [14]. Este enfoque permite optimizar el rendimiento al utilizar información previa sobre los objetos de interés, simplificando la planificación y ejecución de las tareas relacionadas con el agarre.

Los **objetos familiares**, parten del conocimiento de modelos de objetos y agarres similares anteriormente aprendidos; el desafío de detección y captura radica principalmente en la representación del objeto y la transferencia de experiencia previa [6, 14]. Bohg et. al. [6] introducen una sub-clasificación de los objetos familiares dependiendo de su grado de similitud: *similitud baja*¹ y *similitud alta*².

Los **objetos desconocidos** se centran principalmente en la estructura o características locales o globales de datos sensoriales para generar y clasificar candidatos de agarre [6, 14]; no obstante, en

¹**Similitud Baja:** el objeto se define en términos de su forma, color o textura [6].

²**Similitud Alta:** el objeto se define según su categoría [6].

muchas ocasiones la información captada por los sistemas sensoriales es escasa, ruidosa e/o incompleta [14]. Sin embargo, Eizicovits et. al. [14] afirman que la implementación de entradas sensoriales contribuye a reducir la complejidad y el tiempo de planificación de agarre en los sistemas, a la vez que incrementa la solidez del sistema

Los *métodos basados en análisis* [6, 11, 12, 14, 29, 48, 50, 57] desarrollan estrategias para la generación de agarres de cierre forzado utilizando herramientas multifuncionales que poseen habilidad, equilibrio, estabilidad y, en algunos casos, comportamiento dinámico [6]. Estos métodos abordan el problema del agarre formulándolo como una optimización restringida sobre criterios relacionados con una o varias de estas propiedades esenciales [6]. Dichos criterios se fundamentan en formulaciones geométricas, dinámicas y/o cinemáticas.

Estos enfoques parten de suposiciones sobre los modelos geométricos y físicos ideales de los objetos [6, 10], utilizando modelos de contacto simplificado, fricción de Coulomb y el modelado de cuerpos rígidos para simplificar el análisis del agarre [6]. Sin embargo, estas aproximaciones los hacen vulnerables al ruido, las incertidumbres [14], generando inconsistencias y ambigüedades, especialmente en el análisis dinámico [6]. Además, suelen tener un alto costo computacional, lo que los hace menos adecuados para la ejecución en tiempo real [6].

Roa et. al. [48] realizan una descripción de las propiedades evaluadas por medio de los criterios de calidad en los métodos basados en análisis. Se considera que un agarre tiene cierta *habilidad o destreza*, si el efector final tiene la capacidad de realizar movimientos con el objeto de forma compatible con la tarea exigida [48]. Un agarre presenta *resistencia a las perturbaciones*, cuando se garantiza la inmovilidad del objeto en todo sentido, ya sea por la posición o la fuerza aplicada por la herramienta; en otras palabras, por la forma de cierre o la fuerza de cierre ejercida por el efector final [48].

El *equilibrio* del agarre parte del concepto físico de equilibrio estático, donde la suma de las fuerzas (fuerza neta o resultante) y torques aplicados sobre el objeto es igual a cero; y por lo tanto, el sistema se encuentra en reposo o su centro de masa se mueve a velocidad constante [48]. De manera similar, la *estabilidad* del agarre parte del concepto físico de equilibrio dinámico, donde cualquier error en la posición del objeto causado por una perturbación desaparece tiempo después de que desaparece la perturbación [48]. En muchas ocasiones, el cierre del agarre es erróneamente equiparado con la estabilidad [6].

Los *métodos basados en datos* evolucionaron a partir de análisis simulados y fueron inicialmente evaluados con métricas clásicas basadas en formulaciones analíticas; sin embargo, estas métricas no se consideran aceptables para la configuración de un agarre óptimo [6]. Estos enfoques otorgan mayor importancia a la representación del objeto y al procesamiento perceptivo, incluyendo tareas como reconocimiento y clasificación de objetos, extracción de características, métricas de similitud y estimación de posturas [6].

No obstante, estos métodos enfrentan desafíos, ya que requieren una amplia base de datos pa-

ra el aprendizaje del sistema y carecen de una métrica de calidad adecuada, lo que dificulta la construcción de una función que garantice la calidad del agarre. Esto lleva a una dependencia significativa de los resultados obtenidos mediante la implementación práctica [6, 14]. Según [6], los métodos basados en datos pueden basarse en enfoques heurísticos o de aprendizaje, utilizando datos de entrenamiento etiquetados, demostración humana o datos de prueba y error.

A diferencia de los enfoques analíticos, los métodos basados en datos clasifican y evalúan los puntos de contacto de agarre, los modelos aprendidos de los puntos de contacto conocidos, las configuraciones de agarre basadas en características visibles y las predefinidas en su experiencia para varios objetos, con el objetivo de actualizar continuamente su árbol de decisión [6, 14]. Estos métodos tienen menor parametrización y por lo tanto, se adaptan a las incertidumbres del entorno.

Investigaciones recientes, a conllevado a desarrollar métodos de planificación de agarre mediante la combinación de enfoques basados en análisis y en datos con el propósito de reducir las limitaciones presentadas a nivel práctico [14, 30].

1.4.2. Agarre robótica

La manipulación eficiente de elementos requiere el uso de pinzas específicas capaces de sujetarlos firmemente sin causarles daño. Estos elementos pueden ser tanto rígidos como flexibles. La captura y manipulación de objetos rígidos implica el estudio cuidadoso de las fuerzas empleadas en el proceso de agarre, ya que algunos de ellos son estructuralmente frágiles, y una aplicación excesiva de fuerza podría resultar en su destrucción.

Cuando se trata de objetos flexibles y semiflexibles, el desafío es menor en términos de fuerza requerida. Sin embargo, es importante tener en cuenta que algunos objetos pueden deformarse estructuralmente sin sufrir daño completo en su material. Para abordar este problema, seguimos un modelo de rigidez suave y variable.

Conocer el parámetro de rigidez de los objetos que se pretenden manipular es esencial a la hora de seleccionar el tipo de gripper a utilizar, así como las características de sus dedos, con el fin de lograr la realización efectiva de las tareas de agarre [2]. Este conocimiento permite adaptar la herramienta de agarre de manera óptima, considerando la resistencia y la estructura del objeto objetivo.

Los grippers con dedos rígidos son idóneos para manipular objetos pesados, pero presentan dificultades al enfrentarse a objetos irregulares o delicados [37], en nuestro caso de estudio, las botellas tipo PET. Sin embargo, investigaciones recientes han explorado grippers que pueden ajustar su rigidez para adaptarse a diferentes objetos, permitiendo una mayor versatilidad en las capacidades de agarre [54].

La robótica blanda surge como una alternativa para manipular objetos suaves o delicados, utilizando materiales suaves o interactuando con objetos de características desconocidas [41, 47]. A

pesar de su eficacia en eventos no controlados, los grippers blandos a menudo carecen de fuerza y precisión, lo que dificulta la manipulación de objetos grandes o aquellos que requieren agarres ideales [25]. En el caso de grippers rígidos o híbridos, se emplea comúnmente un mecanismo de subactuación para lograr un agarre adaptable y envolvente [9, 43].

1.5. Justificación del problema

La manipulación de objetos es un área fundamental en la robótica, pero sigue siendo un desafío abierto debido a la falta de destreza con la que los robots manejan objetos complejos [28]. Identificar la configuración óptima de agarre implica considerar varios factores, como la tarea específica que el robot debe realizar, las características del objeto, el conocimiento previo disponible y el tipo de mano utilizada [6].

La integración de la visión en la manipulación robótica es crucial para mejorar la capacidad de los robots de asistir en diversas tareas humanas, desde la automatización industrial hasta los servicios domésticos [6]. La detección de agarre, que implica que el robot interactúe con su entorno utilizando sensores de visión para identificar objetos y determinar cómo agarrarlos, desempeña un papel fundamental en este proceso [7].

La síntesis de agarre, que se centra en encontrar los puntos adecuados en el objeto para un agarre seguro y estable, es el corazón del problema de manipulación robótica. Estos puntos de agarre deben garantizar la estabilidad del objeto frente a las fuerzas externas y cumplir con los criterios relevantes para la tarea de agarre.

Uno de los problemas recurrentes en robótica es la caracterización de las regiones de agarre alrededor de la estructura física de cada objeto específico. Esta tarea, aunque fundamental, no es sencilla y requiere el análisis de múltiples formas y características. En respuesta a este desafío, se han realizado diversos estudios que buscan generar estas regiones de agarre de forma semiautomática o a través del entrenamiento de redes inteligentes.

Sin embargo, estas regiones generadas a menudo carecen de información significativa sobre las dimensiones aproximadas de los objetos. Además, suelen presentar solapamientos entre ellas, lo que dificulta el proceso de selección de una región de aproximación que conlleve a un agarre efectivo.

1.6. Objetivos

1.6.1. Objetivo general.

Desarrollar un sistema de aprendizaje que determine regiones de agarre de botellas basado en visión en un entorno semi-estructurado que simula una línea de reciclaje.

1.6.2. Objetivos específicos.

- Establecer mediante revisión bibliográfica, los métodos más relevantes empleados en la generación de regiones de agarre.
- Generar una base de datos de imágenes de botellas plásticas de una línea de reciclaje simulada bajo entorno semi-estructurado.
- Reconocer mediante algoritmos de visión de máquina los objetos registrados en la base de datos con orientación a la detección semi-automática de regiones de agarre.
- Determinar a través de métodos matemáticos y análisis geométrico, las regiones y la dirección de aproximación para realizar el agarre.
- Comprobar el desempeño del sistema propuesto para el agarre de botellas en la línea de reciclaje simulada.

2 Generación del banco de datos

2.1. Introducción

La primera etapa que debe afrontar un Sistema de Visión Artificial (SVA), es la etapa de adquisición de imágenes. Esta etapa, implica la obtención más adecuada de las imágenes, ya que supone un proceso de reconocimiento exitoso; así mismo, como la representación más verídica de la escena, que permita conducir a la caracterización más cercana a la realidad.

En este capítulo, se expone el procedimiento llevado a cabo para la generación de una base de datos de imágenes digitales. Inicialmente, haciendo uso de uno de los métodos de adquisición de imágenes diseñados en la sección 2.2, se construye una base de datos realizando capturas individuales y conjuntas de botellas plásticas, en la sección 2.3.

2.2. Escenarios de adquisición de imágenes

Existen múltiples obstáculos que influyen directamente en el proceso de captura y adquisición de imágenes digitales: el posicionamiento y configuración de *la cámara*; *el entorno*, las características (geométricas y físicas) y el posicionamiento espacial de *los objetos* en la escena; *la iluminación*, *el fondo o background*, *el ruido eléctrico-óptico externo*, etc.

- **La cámara**

A la hora de diseñar un sistema de visión artificial, el primer elemento a escoger y caracterizar es el tipo de la cámara fotográfica a utilizar. Esta elección dependerá de ciertos criterios a considerar, tales como: *cámara de color o monocromática*, *la resolución*, *la relación calidad-ruido*, *el tipo de lente*, entre otros.

- **La iluminación**

En fotografía, la luz desempeña un papel primordial en la adquisición de imágenes en una escena determinada. Existen dos tipos de luz: la luz natural y la luz artificial. La *luz natural*, provista por el sol, no es controlable y presenta continuas fluctuaciones en el transcurso del día. Por otra parte, la *luz artificial*, creada por los seres humanos, no posee fluctuaciones significativas, presentando un comportamiento predecible y controlable.

La luz, posee múltiples características como: la temperatura de color, la dirección, la distancia, el tamaño y la difusión. La temperatura de color, representa la tonalidad cromática

que irradia un cuerpo a determinada temperatura.

La difusión de la luz, es un fenómeno de dispersión ocasionado por la reflexión y la refracción de la luz en un elemento determinado; ó, por la transmisión lumínica a través de un medio traslúcido y no transparente; produciendo una distribución espacial aleatoria de los rayos de luz. Ésta, es afectada por el tamaño y la distancia de la fuente de luz respecto al objeto iluminado; siendo considerada suave o difusa. La luz difusa, presenta un distribución homogénea de los rayos lumínicos, generando pocas sombras y contrastes. Por el contrario, la luz dura, presenta sombras marcadas y mucho contraste en la imagen generada.

■ El fondo o background

En el ámbito de los estudios fotográficos, se utilizan diversos tipos de fondos, conocidos como *backgrounds*. En particular, al adquirir imágenes de objetos en entornos controlados, es común emplear fondos monocromáticos, siendo el blanco y el negro las opciones más frecuentes. Estos fondos suelen estar compuestos principalmente de materiales como papel (papel cartón), tela, vinilo o PVC.

Los *fondos de color blanco* son preferidos por su capacidad para proporcionar una escena limpia y luminosa. Esta elección destaca al objeto principal de la imagen, creando un fondo neutro que resalta las características del objeto de interés. Por otro lado, los *fondos de color negro* ofrecen una apariencia más fuerte y opaca. Esto permite el uso de luces más intensas, laterales o incluso luces coloreadas, creando un contraste visual llamativo. La opacidad del fondo negro también contribuye a generar un ambiente más dramático en la imagen.

2.2.1. Escenario para capturas fotográficas I

Inicialmente, se planteó un escenario de captura de imágenes digitales utilizando una cámara de acción **Xiaomi Yi Lite (2017)**, como se muestra en la Figura 2-1a. Esta cámara posee una resolución de 16 MP, opera en modo de toma secuencial, con un valor máximo de ISO (ISO_{max}) de 800, un ajuste de exposición (EV) de +1.0, y tiene activada la corrección de distorsión.

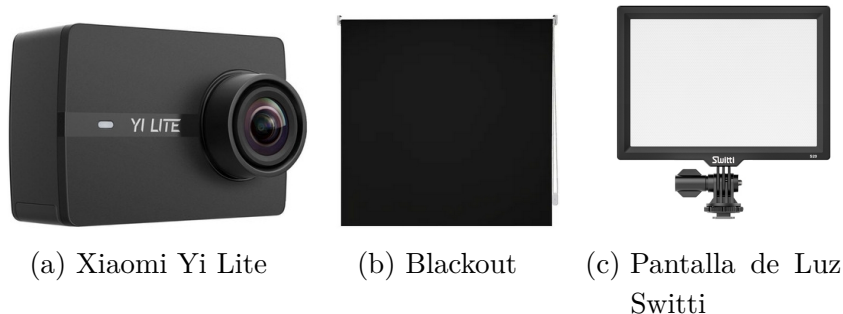


Figura 2-1: Elementos empleados escenario fotográfico I.

El fondo, representado por el Blackout en la Figura 2-1b, se configuró con dimensiones de 120 cm x 120 cm y presenta un color negro mate compuesto en un 90 % de fibra de PVC. Además, la iluminación de la escena se llevó a cabo mediante un sistema de dos (2) pantallas reguladas de luz Switti, como se muestra en la Figura 2-1c. Estas pantallas tienen la capacidad de operar en un rango de temperaturas de color entre 3200 K y 5600 K, permitiendo ajustar el brillo en una escala del 0 % al 100 % con variaciones de ± 1 %.

La metodología aplicada se consideró inapropiada debido a la baja y focalizada luminosidad proporcionada por las pantallas de luz utilizadas. En consecuencia, se tomó la decisión de incrementar inicialmente el valor de exposición (EV) a +2,0, seguido por el ajuste de la sensibilidad máxima (ISO_{max}) a 800. Estos cambios se realizaron con el propósito de compensar la deficiencia de luz experimentada. Sin embargo, es importante destacar que este ajuste resultó en un aumento notable en el ruido percibido y en una pérdida de la resolución de la imagen capturada.

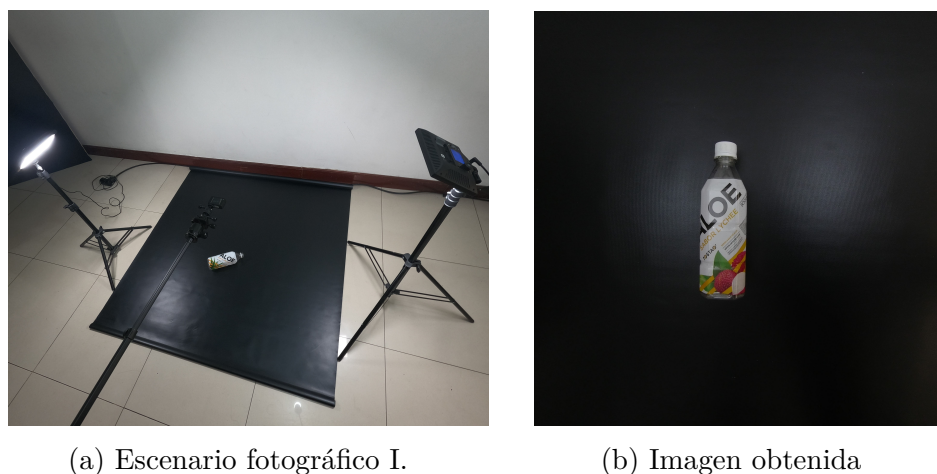


Figura 2-2: Implementación del escenario para capturas fotográficas I.

Adicionalmente, se observaron sombras de luces considerables sobre el Blackout utilizado, como se ilustra en la Figura 2-2b. Estas sombras, al ser registradas, generaron distorsiones en la textura y el contorno de los elementos durante la etapa de post-procesamiento, lo que impidió realizar una identificación adecuada de algunos objetos.

2.2.2. Escenario para capturas fotográficas II

El escenario de captura fotográficas II, se construyó empleando la cámara de acción **GoPro Hero 9 Black (2020)** (vease Figura 2-3a) con resolución de 20 MP, configurada en modo de toma secuencial, ISO_{max} de 200, valor de exposición (EV) de +0,5 y lente lineal, empleado para optimizar el proceso de adquisición de imágenes.

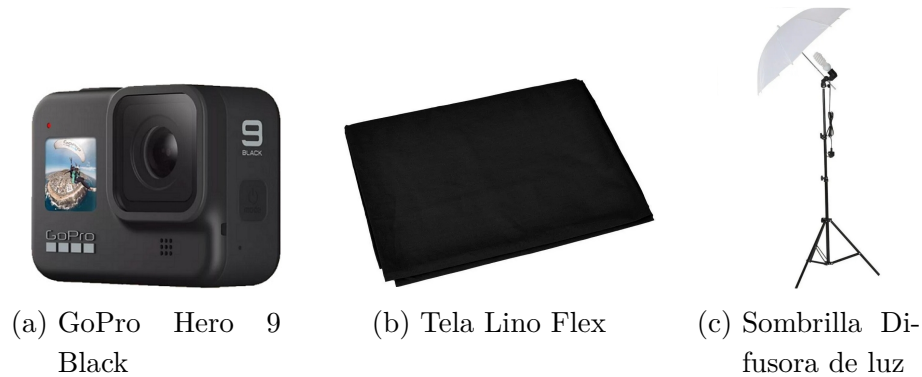


Figura 2-3: Elementos empleados escenario fotográfico II.

El fondo ambiente se logró mediante el uso de una doble capa de la tela de *lino flex*, como se presenta en la Figura 2-3b. Estas telas, de color negro mate y con dimensiones de 150 cm x 150 cm, se mantuvieron tensas y soportadas mediante marcos de madera del mismo tamaño.

La escena construida, ilustrada en la Figura 2-4a, se iluminó con un sistema compuesto por dos sombrillas de iluminación equipadas con bombillas LED de 15 W a 6500 K, tal como se puede apreciar en la Figura 2-3c. Este sistema de iluminación garantizó una dispersión uniforme de la luz, evitando la proyección de sombras de luces sobre la escena, las cuales podrían haberse originado debido a una luminosidad focalizada, como se muestra en la Figura 2-4b. Es relevante señalar que el material textil utilizado desempeñó un papel crucial al contribuir significativamente a la absorción de la luz empleada en la escena.



(a) Escenario Fotográfico del Método II.



(b) Imagen Obtenida

Figura 2-4: Implementación del escenario fotográfico II.

2.3. Implementación

La presente base de datos de imágenes digitales, se construye con la finalidad de obtener el material de estudio necesario para el análisis y la generación automática de regiones de agarre de objetos

cilíndricos en un entorno semi-estructurado. Con el propósito de dar cumplimiento a tal objetivo, se limitó la obtención de imágenes a un grupo reducido de botellas de plástico PET de uso comercial en Colombia (véase Figura 2-5), empleadas como medio contenedor de bebidas gasificadas, azucaradas e hidratantes.



Figura 2-5: Botellas plásticas empleadas en la generación de la base de datos.

Para la fase inicial de construcción de la base de datos, se llevó a cabo la captura individual de las 10 botellas de plástico PET seleccionadas, como se describe en la sección 2.3.1. Este proceso se realizó utilizando los elementos y el escenario fotográfico detallados en el Método II, descrito en la sección 2.2. El objetivo principal de esta etapa fue caracterizar algunas de las propiedades físicas presentes en las botellas y recopilar información relevante que sería utilizada en las siguientes etapas del proceso, específicamente en el reconocimiento y la generación automática de regiones de agarre.

Después de la captura individual de las botellas, se llevaron a cabo capturas conjuntas de los elementos disponibles, como se describe en la sección 2.3.2. El objetivo de estas capturas conjuntas fue simular el flujo de elementos presentes en un entorno de reciclaje de botellas. Este enfoque permitió obtener datos más realistas sobre cómo las botellas interactúan y se presentan en un escenario donde varios elementos coexisten. Estas capturas conjuntas proporcionaron información valiosa para la evaluación de situaciones prácticas y escenarios más complejos, lo que contribuyó a mejorar la robustez y la aplicabilidad de las estrategias de reconocimiento y generación de regiones de agarre.

2.3.1. Captura individual de las botellas

La captura de las fotografías individuales de cada botella se llevó a cabo con la cámara de acción GoPro Hero 9 Black (2020). La configuración de la cámara consistió en programarla para capturar imágenes cada 5 segundos. Además, la cámara estaba montada en el cabezal de trípode motorizado panorámico Movo Photo MTP-11, como se muestra en la Figura 2-6.



Figura 2-6: Movo Photo MTP-11

El cabezal de trípode motorizado panorámico Movo Photo MTP-11 se configuró para realizar recorridos de 360° durante períodos de 5 minutos, capturando tomas fotográficas cada 6° de recorrido. En total, se obtuvieron 300 imágenes por cada botella, correspondientes a:

- 240 tomas fotográficas, 60 capturas rotadas de la botella con etiqueta respecto a la cámara (Fig. 2-8a), por cada ángulo de rotación (0° (Fig. 2-7a), 90° (Fig. 2-7b), 180° (Fig. 2-7c), 270° (Fig. 2-7d)) del objeto.

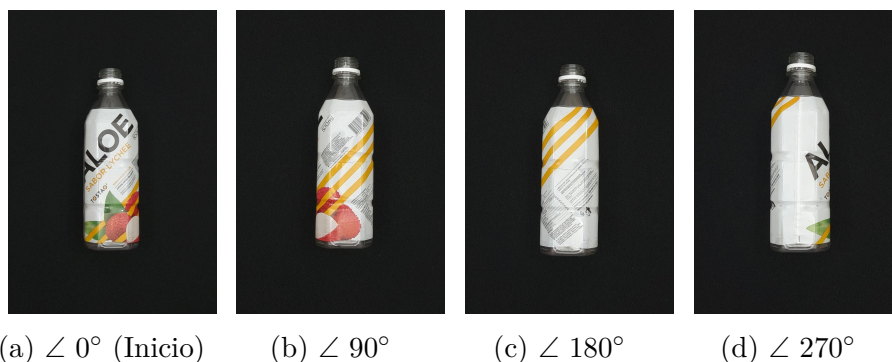


Figura 2-7: Ángulos de Rotación de las Botellas

- 60 capturas fotográficas rotadas de la botella sin etiqueta con relación a la cámara (Fig. 2-8b).

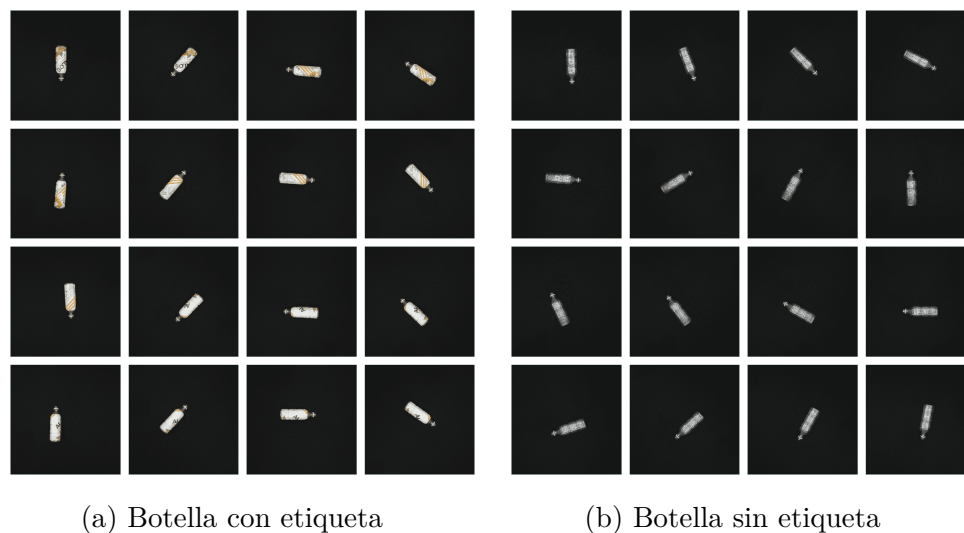


Figura 2-8: Tomas fotográficas realizadas a las botellas de forma individual

2.3.2. Captura conjunta de las botellas

La toma de imágenes de todo el conjunto de objetos se realizó utilizando la cámara *GoPro Hero 9 Black* de forma estática. Se capturaron imágenes de las 10 diferentes botellas plásticas disponibles en diversas poses y grupos de manera aleatoria (véase Figura 2-9). Estas incluían simultáneamente botellas con y sin etiquetado de revestimiento.

El proceso de toma fotográfica, se optimizó la captura realizándola en lapsos de tiempo de 20 a 30 segundos durante periodos de 1 hora, lo que resultó en un total de 600 imágenes. Es importante destacar que las botellas se posicionaron de manera aleatoria, variando en cantidad, posición, distribución, colores y condiciones especiales, como la falta de etiquetado de revestimiento en algunos envases.

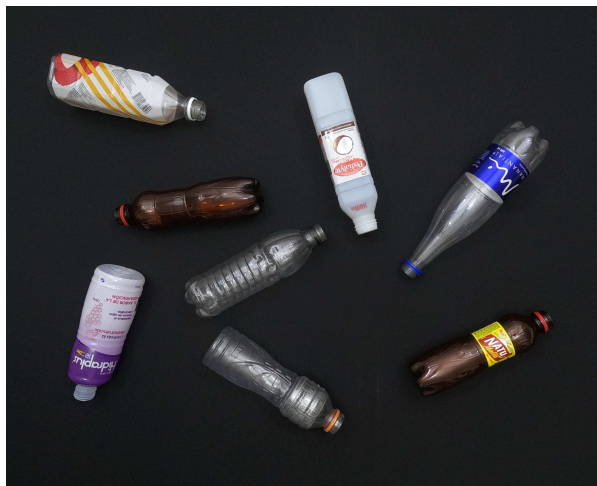


Figura 2-9: Toma fotográfica realizada a las botellas de forma conjunta

3 Sistemas de percepción (visión por computador)

3.1. Introducción

El problema de visión por computador tratado a continuación, considera la complejidad presente de un entorno semi-estructurado en la medida en la que supone niveles de luminancia variables y uniformes, evitando ciertos reflejos y sombras en los contornos de los objetos; permitiendo la distinción factible de los mismos, inmersos en un fondo homogéneo.

En la Figura 3-1, se expone el esquema empleado para desarrollar del proceso de percepción visual del presente capítulo; a partir de los datos previamente adquiridos en la sección 2.3.

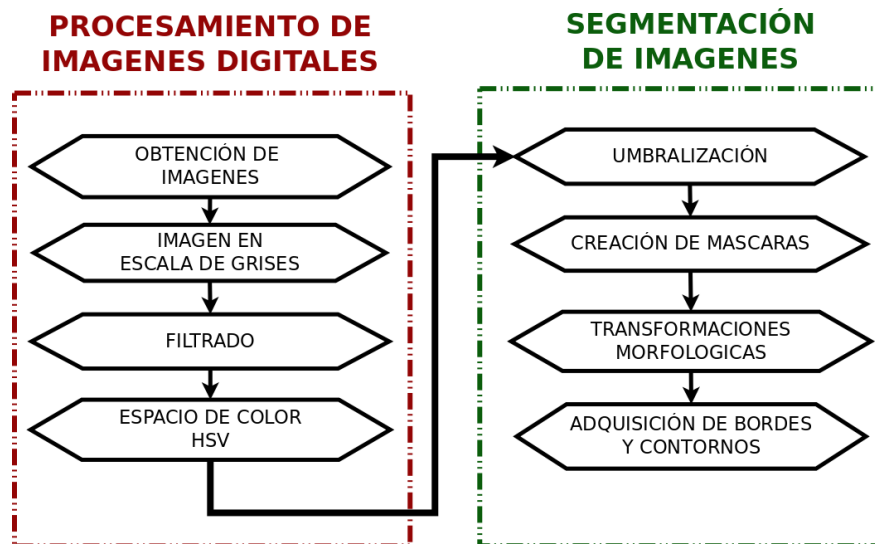


Figura 3-1: Proceso de percepción visual (sistema de visión artificial).

Inicialmente, se desarrolló el procesamiento de imágenes digitales en la sección 3.2, implementando técnicas y funciones con el objetivo de mejorar y adaptar el contenido relevante. Seguidamente, se llevó a cabo el proceso de segmentación de imágenes en la sección 3.3.

3.2. Procesamiento de Imágenes Digitales

En la Figura 3-2, se ilustra el proceso inicial del procesamiento de imágenes digitales. Primeramente, se desarrolló el análisis de fenómenos físicos y ópticos observados en las imágenes los objetos de la escena (sección 3.2.1). Este análisis se centra en aspectos como la reflexión de la luz y la transparencia u opacidad de los objetos. Posteriormente, se aplican de manera estructurada técnicas y funciones para mejorar y ajustar las imágenes. Este proceso incluye el análisis del color, abordado en las secciones 3.2.2 y 3.2.4, así como procesos de filtrado detallados en la sección 3.2.3. El objetivo de estas técnicas es preparar los elementos contenidos en la información visual para la siguiente etapa del proceso: la segmentación (sección 3.3).



Figura 3-2: Procesamiento de imágenes digitales.

3.2.1. Imágenes digitales

Una imagen digital, es una representación óptica discretizada y finita en luminosidad y coordenadas espaciales de un fenómeno físico, cuyos elementos son denominados píxeles; describiéndose, como la función $f(x, y)$ de las intensidades de luz en cada punto x, y dentro del espacio bidimensional de la imagen.

Paralelamente, una imagen digital, puede ser representada por una matriz de N filas por M columnas, descrita por una función de distribución $f(n, m)$ (Ec. 3-1) para todo $n \in [0, N - 1]$ y $m \in [0, M - 1]$, donde $N, M \in \mathbb{Z}^+$.

Normalmente, suelen ser empleadas escalas de intensidades 2^N bits; siendo el octeto de bits (8-bits) o *Byte*, el más comúnmente utilizado, definiendo el rango dinámico de las intensidades entre los

valores de 0 a 255.

$$f(n, m) = \begin{pmatrix} f_{11} & f_{12} & f_{13} & f_{14} & \cdots & f_{1m} \\ f_{21} & f_{22} & f_{23} & f_{24} & \cdots & f_{2m} \\ f_{31} & f_{32} & f_{33} & f_{34} & \cdots & f_{3m} \\ f_{41} & f_{42} & f_{43} & f_{44} & \cdots & f_{4m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ f_{n1} & f_{n2} & f_{n3} & f_{n4} & \cdots & f_{nm} \end{pmatrix} \quad (3-1)$$

Una imagen digital, puede ser obtenida de dos formas diferentes: A partir de la interacción de un dispositivo en capacidad de captar el mundo físico, o a través de una base de datos de imágenes preexistentes. Partiendo de lo expresado anteriormente, en la presente sección son analizadas algunas características percibidas en los datos obtenidos en la sección 2.3, adquiridos a través de la captura fotográfica de un grupo reducido de botellas plásticas de uso cotidiano, inmersas en un entorno semi-estructurado de luminosidad controlada.

Análisis de las imágenes

Las *botellas de plástico PET¹* se destacan por ser envases de alta resistencia y transparencia, disponibles en diversas formas, colores y tamaños. A pesar de estas características, las propiedades físicas de transparencia y reflexión lumínica pueden verse directamente influenciadas por la presencia de colores en el material y/o las propiedades del etiquetado de revestimiento. Esto hace imperativa la necesidad de caracterizar estas botellas para comprender y gestionar adecuadamente sus propiedades visuales.

Inicialmente, se singularizaron las tres posibles tonalidades de color fronteras observadas en las botellas utilizadas, como se ilustra en la Figura 3-3. La tonalidad de las botellas puede plantear ciertos desafíos en el proceso de detección debido al bajo contraste entre algunos objetos y el fondo circundante. Estas tres tonalidades específicas son las siguientes: blancas o claras (véase Figura 3-3a), translúcidas (véase Figura 3-3b) y negras u oscuras (véase Figura 3-3c).

Las botellas de tonalidades blancas o claras, como se muestra en la Figura 3-3a, no presentan desafíos significativos en su proceso de caracterización debido al alto contraste que tienen con el fondo utilizado. Sin embargo, las botellas de tonalidades negras u oscuras, ilustradas en la Figura 3-3c, representan un gran desafío debido al bajo contraste existente con el fondo. La similitud o proximidad en las tonalidades de color puede dificultar la percepción del contorno del objeto de interés, lo que complica el proceso de detección y reconocimiento.

¹PET: Tereftalato de polietileno



Figura 3-3: Tonalidades características límites de las Botellas PET empleadas.

Las botellas translúcidas, como se evidencia en la Figura 3-3b, presentan un desafío particular debido a su carencia de color, lo que las vuelve permeables a la proyección parcial del color del fondo en su superficie. Además, las propiedades reflexivas del material generan reflejos de luz que pueden entorpecer el proceso de caracterización, convirtiéndose en un desafío durante el proceso de reconocimiento. La tonalidad de estas botellas es especialmente susceptible a las pequeñas variaciones lumínicas presentes en la escena.

3.2.2. Modelo de color de escala de grises

Realizar la identificación de objetos o elementos en imágenes digitales no resulta ser una tarea sencilla. Para ello, es necesario determinar características o patrones que los identifiquen. Una de las características empleadas comúnmente es el color. No obstante, el color puede ser representado de múltiples formas, las cuales son conocidas como *espacios* o *modelos de color*.

Un espacio de color, es una distribución establecida de un grupo finito de colores que componen una imagen o vídeo. Un modelo de color, es un modelo matemático descriptivo de una gama de colores mediante un conjunto finito de números enteros positivos (\mathbb{Z}^+), empleando 3 o 4 valores cromáticos.

Las imágenes obtenidas en la sección 2.3, se adquirieron en el modo de color RGB. El modelo de colores RGB (Red-Green-Blue), es un modelo de representación que emplea valores individuales de croma *Rojo*, *Verde* y *Azul*. En este modelo, un color específico se obtiene a partir la suma ponderada de sus componentes cromáticas, a través de un modelo conocido como *síntesis aditiva*.

El modelo en escala de grises, es un modelo de representación que define los colores en la imagen empleando una única componente, la *luminosidad*. Los niveles de luminosidad, son descritos matemáticamente mediante una tupla de 8 – *bits*, adquiriendo valores en un rango de 0 (Negro) a 255 (Blanco); representando cada color, por un valor progresivo en el rango de las tonalidades de gris².

²**Gris Cromático:** gris creado a partir de los colores primarios del modelo de color CMY (Cyan-Magenta-Yellow).

Las imágenes en escala de grises, contienen menos información que en otros modelos de color. Sin embargo, son comunes en el procesamiento de imágenes debido a requieren menos espacio disponible, permiten una mejor caracterización y detección de bordes y contornos; y adicionalmente, facilitan la realización de procesos con cálculos complejos.

Existen dos métodos utilizados comúnmente en la conversión de imágenes RGB en Escala de Grises, el método promedio y el método ponderado.

- El **método promedio**, calcula el valor promedio de las componentes cromáticas R, G y B como el valor total en la escala de grises (Ec. 3-2); desconociendo las diferencias de sensibilidad del ojo humano³ en presencia de cada valor cromático en el modelo de color RGB, al asignar el mismo valor contributivo de cada una de las componentes al valor total de la luminancia calculada.

$$RGB[A] \text{ to Gray} : \text{Grayscale} = \frac{R}{3} + \frac{G}{3} + \frac{B}{3} \quad (3-2)$$

- El **método ponderado** o de **luminosidad (Y)**, es el método más implementado actualmente en el procesamiento de imágenes digitales y en el presente estudio. Este, a diferencia del *método promedio*, parte del análisis de la sensibilidad del ojo humano en presencia de los colores primarios del modelo de color RGB; determinando el valor ponderado de contribución de cada color al valor total de luminancia percibida. Este método de cálculo de la luminancia se encuentra codificado en la "Recommendation ITU-R BT.601-7 Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios"⁴ de la Unión Internacional de Telecomunicaciones, descrito en la siguiente ecuación:

$$RGB[A] \text{ to Gray}[Y]: Y \leftarrow 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \quad (3-3)$$

De forma inversa, la obtención de una imagen en el modelo de color *RGB*, partiendo de su modelo de color en escala de grises; estará determinada por la asignación del valor de luminancia ponderado obtenido de la imagen, a cada uno de los componentes cromáticos del espacio de color *RGB*, tal y como se describe en la siguiente ecuación:

$$\text{Gray}[Y] \text{ to RGB}[A]: R \leftarrow Y, G \leftarrow Y, B \leftarrow Y, A \leftarrow \text{máx}(\text{ChannelRange}) \quad (3-4)$$

3.2.3. Filtrado de imágenes

Toda imagen adquirida por medios ópticos o electrónicos, sufre cierta degradación expresada en forma de ruido y/o perdida de definición debido a las imprecisiones en el enfoque, perturbaciones aleatorias y al ruido intrínseco del dispositivo de captura. Los mecanismos empleados para mitigar estos efectos, incluyen técnicas u operaciones de *realce* y *restauración de imágenes*; permitiendo eliminar el ruido, ajustar el brillo, el color, los bordes y/o el contraste de una imagen de acuerdo

³Los ojos humanos son más sensibles al verde, posteriormente al rojo y finalmente, menos sensibles al azul.

⁴https://www.itu.int/dms_pubrec/itu-r/rec/bt/R-REC-BT.601-7-201103-I!!PDF-E.pdf

a las necesidades planteadas. Éstas técnicas se dividen en dos categorías: basados en el dominio espacial y basados en el dominio frecuencial.

Los métodos basados en el **dominio espacial**, hacen mención a la modificación directa del espectro de frecuencias espaciales mediante operaciones aritméticas entre cada uno de los píxeles de la imagen con un número determinado de sus píxeles vecinos. Este conjunto de operaciones, se encuentran descritas de forma generalizada por la expresión 3-5, donde $g(x, y)$ representa la imagen resultante de la operación de transformación T sobre un vecindario (x, y) de la imagen $f(x, y)$ [17, 53].

$$g(x, y) = T[f(x, y)] \quad (3-5)$$

Entre los filtros espaciales más usados se encuentran: el filtro de media o de suavizado, el filtro de media ponderado, el filtro de moda, el filtro gaussiano, el filtro laplaciano, el filtro de Sobel, entre otros.

Los métodos basados en el **dominio de la frecuencia**, modifican indirectamente la luminancia de cada píxel mediante el uso de píxeles vecinos como factores de ponderación. Estos métodos, se basan en las transformadas Fourier; donde, la imagen procesada $G(u, v)$ (Ec. 3-6) es el resultado de la Transformada Discreta de Fourier Inversa (IDFT) aplicada al producto de la función de transferencia del filtro $H(u, v)$ y la Transformada Discreta de Fourier (DFT) $F(u, v)$ de la imagen de entrada $f(x, y)$.

$$G(u, v) = H(u, v)F(u, v) \quad (3-6)$$

Uno de los problemas tratados en este estudio, tiene sus orígenes en las diferentes tonalidades de los objetos empleados; siendo esta, la causa principal de las similitudes entre algunos tonos cromáticos de los objetos captados y el fondo, obstaculizando la adecuada segmentación de los mismos. Adicionalmente, este problema es acrecentado por la implementación de las técnicas de filtrado empleadas para el realce y la restauración de imágenes digitales; que en cierta medida, son las causantes de la degradación en la nitidez debido a la *difuminación*, efecto que produce un suavizado en los gradientes de las mismas.

La aplicación de técnicas de filtrado en el dominio espacial y frecuencial de forma global y unilateral en imágenes digitales, en ocasiones no resulta ser eficiente; por tanto, se hace indispensable disponer de métodos combinados que actúen de forma local y en capacidad de extraer la información pertinente. Una de las técnicas de filtrado empleadas que cumple con estas condiciones, es el *Filtrado de Gabor*; filtro lineal que se actúa tanto en el dominio espacial, como frecuencial, empleado en la extracción de características de texturas en imágenes digitales.

El Filtro de Gabor

El Filtro de Gabor (Ec. 3-7), es un filtro lineal en el cual, la respuesta al impulso es el producto de una función sinusoidal compleja (Ec. 3-8) y una función gaussiana bidimensional (Ec.

3-9); conocidas como portadora y envolvente, respectivamente. Este, es ampliamente empleado en la extracción de información de texturas locales; operando como un filtro de paso de banda para una distribución específica de frecuencias dentro de una región localizada, determinada por las orientaciones espaciales entorno a un punto o área estudiada; logrando una resolución óptima tanto en el dominio espacial, como frecuencial [17].

$$g(x, y) = s(x, y) \cdot t(x, y) \quad (3-7)$$

$$s(x, y) = \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (3-8)$$

$$t(x, y) = -\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2} \quad (3-9)$$

La expresión matemática de la ecuación 3-10, describe la función de Gabor bidimensional; donde, su parte real (Ec. 3-11) e imaginaria (Ec. 3-12), son ortogonales y pueden ser empleadas de forma individual [17].

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (3-10)$$

$$Re\{g(x, y; \lambda, \theta, \psi, \sigma, \gamma)\} = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right) \quad (3-11)$$

$$Im\{g(x, y; \lambda, \theta, \psi, \sigma, \gamma)\} = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{x'}{\lambda} + \psi\right) \quad (3-12)$$

La función de Gabor bidimensional (Ec. 3-10), se encuentra localizada tanto en el dominio espacial (x, y) expresada en coordenadas polares (excentricidad y meridiano)(Ec. 3-13 y Ec. 3-14), como en el dominio de la frecuencia (θ, λ) ; adicionalmente, se encuentra conformada por una serie de parámetros que describen su forma: σ , representa la desviación estándar de la envolvente gaussiana y esta relacionada con el ancho de banda radial b ; γ , el factor de forma; y ψ define el desfase sinusoidal [?, 17].

$$x' = x \cos \theta + y \sin \theta \quad (3-13)$$

$$y' = -x \sin \theta + y \cos \theta \quad (3-14)$$

■ El kernel gaussiano de Gabor (\mathbf{K})

$\mathbf{K}_{size} = (k_{i,j})$, es denominado el tamaño del kernel (\mathbf{K}) gaussiano de Gabor. Al igual que otros núcleos de convolución:

1. K_{size} debe ser impar, de tal forma que tenga un centro, el cual es denominado radio.
2. La suma total de sus elementos debe ser igual a 1 si se quiere garantizar que el brillo de la imagen permanezca inalterado. Por consiguiente:
 - Si, $\sum_{i,j} k_{i,j} = 0$, la imagen resultante será muy oscura, sin llegar a la total ausencia de luminosidad.
 - Si, $\sum_{i,j} k_{i,j} < 1$, la imagen resultante será más oscura que la original.

- Si, $\sum_{i,j} k_{i,j} = 1$, brillo de la imagen permanezca inalterado.
 - Si, $\sum_{i,j} k_{i,j} > 1$, la imagen filtrada será más brillante que la imagen original.
3. Los valores internos del kernel, son truncados entre 0 y 255, con la finalidad de evitar valores superiores a 255 y valores negativos en las imágenes filtradas.

En la medida en que se incrementa el tamaño del kernel gaussiano de Gabor, el área espacial del filtro se estrecha. Las componentes matriciales del kernel se incrementan y, por consiguiente, la $\sum_{i,j} k_{i,j}$ aumenta, siendo mayor a 1 (> 1), generando una imagen resultante con mayor brillo que la imagen original [17, 58].

Implementación:

La variación en el tamaño del kernel gaussiano en los filtros de Gabor tiene un impacto significativo en la respuesta del filtro y, por consiguiente, en el brillo de la imagen resultante. A medida que el tamaño del kernel aumenta, se observa un estrechamiento en el área espacial del filtro, con un correspondiente aumento en las componentes matriciales y la suma total de elementos del kernel. Este incremento en la suma impulsa un aumento en el brillo de la imagen filtrada, en comparación con la imagen original, como se ejemplifica claramente en la Figura 3-1.

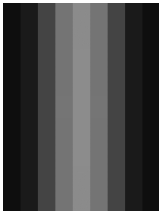
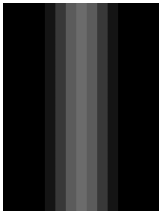
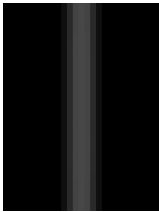
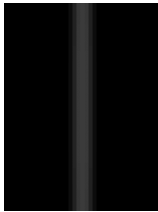


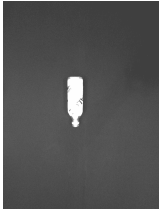

K	3	5	9	13
Filtros				
Imagen Filtrada				
$\sum_{i,j} k_{i,j}$	4.12	5.76	10.15	14.54

Tabla 3-1: Tamaño del kernel del filtro de Gabor

La elección adecuada del tamaño del kernel gaussiano de Gabor durante el proceso de filtrado determinará la intensidad lumínica de la imagen generada. Durante esta etapa, fue crucial encontrar un equilibrio para evitar imágenes con poco brillo que podrían carecer de detalles perceptibles y, al mismo tiempo, prevenir imágenes con un exceso de brillo que podrían dificultar la interpretación visual. La representación visual proporcionada en la figura citada destaca de manera elocuente cómo el tamaño del kernel gaussiano influye en el brillo de la

imagen resultante, subrayando la importancia de esta consideración en el procesamiento de imágenes mediante filtros de Gabor.

■ Orientación de la envolvente de Gabor (θ)

La orientación de la envolvente de Gabor (θ), determina la dirección de la normal en relación al paralelo de los canales en la función de Gabor. Este valor, puede adquirir valores reales en un rango entre 0° y 360° ; donde, 0° corresponde a una orientación vertical, 90° a una orientación horizontal y valores intermedios a orientaciones diagonales de la función de Gabor [?, 17].

Implementación:

La elección estratégica de θ influye significativamente en la capacidad de detección de bordes y texturas. Al ajustar θ según la orientación esperada de los objetos en una escena, se optimiza la respuesta de la función de Gabor, mejorando así la detección de características específicas. Este enfoque también contribuye a la adaptabilidad del modelo frente a variaciones en la iluminación, ya que puede alinearse con la dirección dominante de la luz en la escena, como se aprecia en la Tabla 3-2.

(θ)	0	$2\pi/8$	$4\pi/8$	$6\pi/8$
Filtros				
Imagen Filtrada				

Tabla 3-2: Ángulo de orientación (θ) en los filtros de Gabor.

En la práctica, la orientación de la envolvente de Gabor impacta en la eficiencia de la representación visual de objetos. En aplicaciones como la detección de contornos, el reconocimiento de patrones o el análisis de texturas, la elección cuidadosa de θ contribuye a resaltar características de interés. En resumen, θ en las funciones de Gabor ofrece un mecanismo clave para adaptar el modelo a las características particulares de una escena, mejorando la capacidad de representación y detección de objetos.

Con el objetivo de abordar las variadas orientaciones de objetos en una escena, se optó por emplear ocho valores específicos para el parámetro θ en las funciones de Gabor. Estos valores, 0 , $\frac{\pi}{8}$, $\frac{\pi}{4}$, $\frac{3\pi}{8}$, $\frac{\pi}{2}$, $\frac{5\pi}{8}$, $\frac{3\pi}{4}$, y $\frac{7\pi}{8}$, fueron seleccionados considerando la aleatoriedad inherente de

los objetos presentes. Esta estrategia busca abarcar diversas orientaciones y adaptarse a las características específicas de los objetos en cada escena capturada.

■ Desviación estándar de la envolvente gaussiana de Gabor (σ)

La desviación estándar de la envolvente gaussiana de la función de Gabor σ , determina el tamaño del campo receptivo, controlando el área de cobertura de la imagen gaussiana en el dominio espacial. No obstante, su incidencia está determinada por la longitud de onda λ del núcleo del filtro de Gabor, y el ancho de banda de frecuencia espacial de media respuesta b ; estos tres parámetros, se interrelacionan matemáticamente como se describe a continuación:

$$\sigma = \left(\frac{1}{\pi} \sqrt{\frac{\ln 2}{2}} \cdot \frac{2^b + 1}{2^b - 1} \right) \cdot \lambda \quad \therefore \text{para } b = 1 \Rightarrow \sigma = 0,56 \cdot \lambda \quad (3-15)$$

El ancho de banda de frecuencia espacial de media respuesta de la función de Gabor b , representa el diferencial entre las altas y bajas frecuencias. Este, determina el ancho de cada banda en la dirección radial de cada canal, así como el solapamiento entre canales contiguos; controlando el tamaño total de la envolvente de Gabor.

1. **Desviación estándar (σ):** La desviación estándar controla la dispersión de la envolvente gaussiana de la función de Gabor en el dominio espacial.
 - A medida que σ aumenta, la envolvente gaussiana se vuelve más ancha, cubriendo un área más extensa en la imagen.
 - Un valor mayor de σ implica un campo receptivo más grande y una influencia más amplia en la imagen.
2. **Longitud de onda (λ):** está asociada con la periodicidad de la función de Gabor.
 - Cuando λ es pequeña, la función de Gabor se repite con mayor frecuencia en la imagen, lo que puede resultar en un campo receptivo más pequeño.
 - Una λ más grande implica una función de Gabor más extendida en la imagen.
3. **Ancho de banda de frecuencia espacial (b):** está relacionado con la selectividad de la función de Gabor.
 - Un valor bajo de b indica una mayor selectividad en la frecuencia espacial, lo que puede resultar en un campo receptivo más específico.
 - Un valor alto de b significa una menor selectividad, generando una función de Gabor que afecta a una gama más amplia de frecuencias espaciales.

Implementación:

La interrelación matemática entre estos parámetros puede ser compleja; sin embargo, en

términos generales, es posible ajustar estos valores para adaptar el filtro de Gabor a las características específicas de la escena y el objeto en análisis. Por ejemplo, con el objeto de resaltar los bordes finos, es conveniente seleccionar un σ de menor magnitud e incrementar el valor λ empleado.

Como se puede apreciar en la Tabla 3-3, un incremento en el valor de σ , produce un aumento en el ancho de banda de cada canal de la envolvente de Gabor, generando solapamiento entre canales adyacentes y un efecto de difuminado entre estos. No obstante, el tamaño total de la envolvente de Gabor permanece inalterado. Este parámetro, causa la pérdida de luminosidad en la imagen resultante.

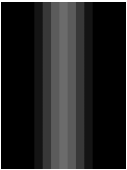
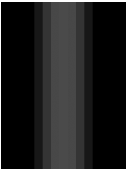
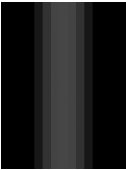
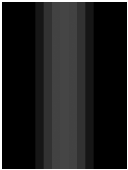




σ	1	$2\sqrt{2}$	$3\sqrt{2}$	$4\sqrt{2}$
Filtros				
Imagen Filtrada				

Tabla 3-3: Desviación estándar de la envolvente gaussiana del filtro de Gabor (σ)

En resumen, la elección de estos parámetros en la función de Gabor, afecta directamente al campo receptivo del filtro y, por ende, a la manera en que el filtro responde a las características específicas de los objetos en la escena, como bordes, texturas, etc. Como se observa en la Tabla 3-3, y debido a la pérdida parcial o total de información de los objetos puestos en la escena para valores iguales o superiores a $3\sqrt{2}$, se optó por emplear los valores inferiores de desviación estándar de la envolvente gaussiana de Gabor (1 y $2\sqrt{2}$).

■ Longitud de onda de la función de Gabor (λ)

La longitud de onda de la función de Gabor (λ), representa el factor sinusoidal de la función de Gabor. El incremento o decremento de este parámetro, aumenta o disminuye respectivamente el efecto dilatorio en el espectro de los filtros de Gabor. Su valor, es consignado en números reales iguales o mayores que 2 e inferiores a la quinta parte del tamaño en píxeles de la imagen de entrada, con el objeto de evitar obtener efectos indeseados en los bordes de la imagen.

Implementación:

En la practica, la longitud de onda (λ) de onda de la función de Gabor, influye significativa-

mente en la detección de objetos en una imagen. Una longitud de onda más corta aumenta la sensibilidad a detalles finos, mientras que una longitud de onda más larga incrementa la sensibilidad a características grandes y suaves.

Adicionalmente, Una longitud de onda específica puede resaltar algunas orientaciones (θ) particulares, siendo útil para la detección de bordes y texturas alineadas en determinada dirección. Es crucial evitar longitudes de onda excesivamente grandes en relación con el tamaño de la imagen, que puedan provocar efectos no deseados en los bordes de la imagen.

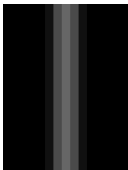
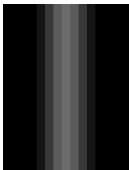
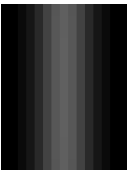
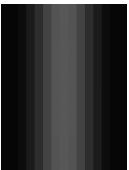


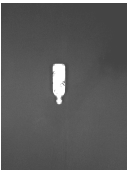
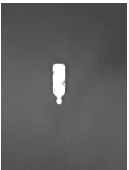
λ	3,0	5,0	8,0	10,0
Filtros				
Imagen Filtrada				

Tabla 3-4: Efectos de la variación de la longitud de onda λ en las filtros de Gabor.

En la Tabla 3-4, se aprecia el efecto dilatorio, el incremento en el número de canales y el aumento en la envolvente de Gabor, producto del incremento del factor sinusoidal de la función de Gabor. Ésta es percibida en la imagen filtrada como un incremento en el brillo, conllevando a cierta pérdida en la nitidez y el contraste.

■ Relación de aspecto espacial de la función de Gabor (γ)

La relación de aspecto espacial de la función de Gabor (γ), determina la elipticidad de la función de Gabor de la siguiente forma:

- Si, $\gamma < 1$, el soporte del filtro se alarga en la orientación de los canales paralelos de la función de Gabor.
- Si, $\gamma = 1$, el soporte del filtro es elipsoidal.
- Si, $\gamma > 1$, el soporte del filtro es circular.

γ	0,05	0,35	0,5	1,0	2,0
Filtros					
Imagen Filtrada					

Figura 3-4: Relación de aspecto espacial γ en la función de Gabor.

En la Figura 3-4, se aprecia el efecto de la elipticidad producto de la variación de la relación de aspecto γ ; donde, el incremento o decremente en el factor γ , incrementa o disminuye la altura espacial del filtro de Gabor respectivamente; centralizando o des-centralizando la acción del filtro.

- $\Psi(\psi)$ representa el desplazamiento de fase entre el centro de la red sinusoidal y el campo receptivo.

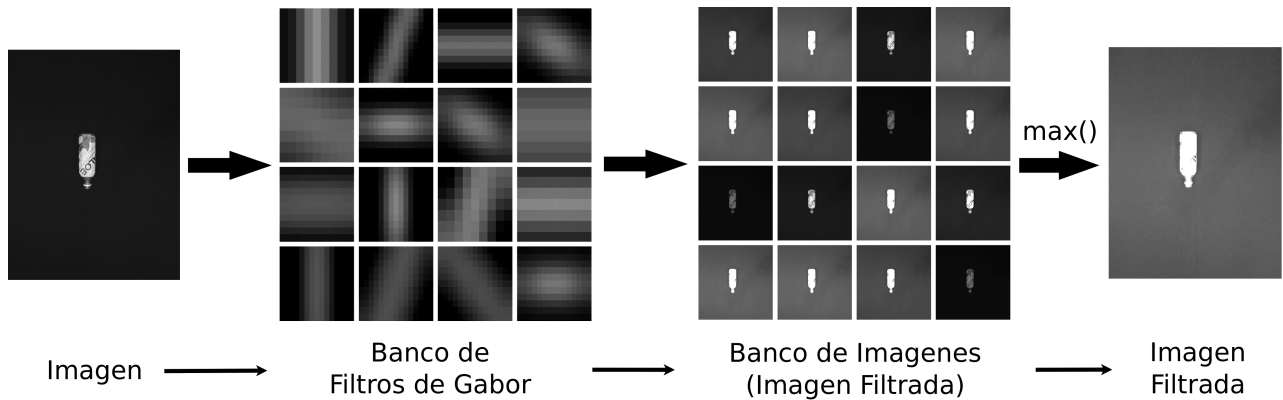


Figura 3-5: Clasificación de texturas multicanal.

Los filtros Gabor proporcionan la máxima resolución espacial en la caracterización, manteniendo el filtro en la banda más estrecha posible para la discriminación de las características espectralmente vecinas de las diferentes texturas.

Un banco de filtros de Gabor permitirá representar una determinada textura con un vector de características cuyos elementos son las variaciones obtenidas a cada respuesta al filtro. Este enfoque es conocido como *clasificación de texturas multicanal* (véase Figura 3-5).

La Clasificación de Texturas Multicanal es una combinación de múltiples convoluciones a partir de la variación de los parámetros de la función de Gabor. Ésta, se realiza empleando las normas L1 (Ec. 3-16), L2 (Ec. 3-17) o L-infinito (Ec. 3-18).

- Norma L1, calcula la suma ponderada de los valores en píxeles de cada filtro obtenido.

$$\|A\|_1 = \sum_{i=1}^n |a_i| \quad (3-16)$$

- Norma L2, hace referencia a la distancia euclidiana, calculando la raíz cuadrada en píxeles de la suma de los valores cuadrados de cada filtro obtenido.

$$\|A\|_2 = \left[\sum_{i=1}^n |a_i|^2 \right]^{\frac{1}{2}} \quad (3-17)$$

- Norma L-infinito, calcula el máximo de los valores absolutos en píxeles de los filtros obtenidos.

$$\|A\|_p = \left[\sum_{i=1}^n |a_i|^p \right]^{\frac{1}{p}} \quad (3-18)$$

Implementación:

A partir de los parámetros consignados en la Tabla 3-5, los conceptos estudiados y los análisis realizados en la actual sección, se construyó un banco de 192 filtros de Gabor (véase Figura 3-5).

Parámetro	Valor
K_{Size}	3, 5
$Sigma(\sigma)$	1, $2\sqrt{2}$
$Theta(\theta)$	0, $\pi/8$, $\pi/4$, $3\pi/8$, $\pi/2$, $5\pi/8$, $3\pi/4$, $7\pi/8$.
$Lambda(\lambda)$	3.0, 5.0, 7.0
$Gamma(\gamma)$	0.05, 0.5
$Psi(\psi)$	0

Tabla 3-5: Parámetros empleados en la construcción de los filtros de Gabor

Los filtros construidos, fueron normalizados empleando la Norma L2 Frobenius (Ec. 3-19), también denominada, Norma de Hilbert–Schmidt. Esta, fue aplicada a cada imagen a de N filas por M columnas, a la entrada del sistema; originando un banco de imágenes filtradas, correspondientes al banco de vectores de características como respuesta a cada uno de los filtros aplicados.

$$\|A\|_F = \left[\sum_{i=1}^N \sum_{j=1}^M |a_{i,j}|^2 \right]^{\frac{1}{2}} \quad (3-19)$$

Seguidamente, las imágenes filtradas, fueron comparadas una a una, píxel a píxel, obteniendo el valor máximo para cada píxel encontrado en el grupo de 192 imágenes filtradas; dando como resultado, una única imagen, producto de la comparación total de los valores para cada píxel (véase Figura 3-6).

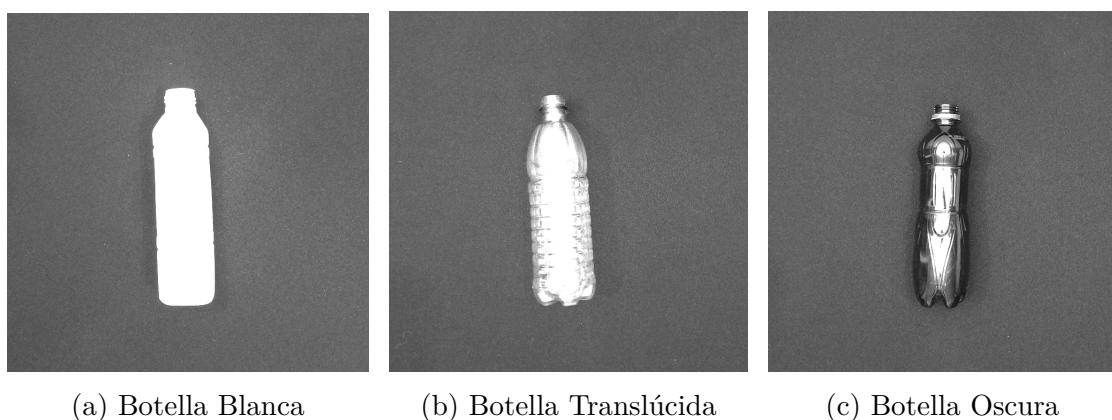


Figura 3-6: Imágenes filtradas aplicando la clasificación de texturas multicanal del filtrado de Gabor

3.2.4. Modelo de color HSV (Hue-Saturation-Value)

El modelo de color HSV (Hue, Saturation, Value), fue creado con el objeto de representar los colores de manera similar a la forma en la que los seres humanos perciben y describen el color. Este, es un modelo de color derivado del modelo de color RGB y similar al modelo de color HSL (Hue, Saturation, Lightness) (véase Figura 3-11). En los espacios de color HSV y HSL, la descripción del *brillo* (V) y la *luminosidad* (L) difieren en el rango asociado a cada tono de color específico.

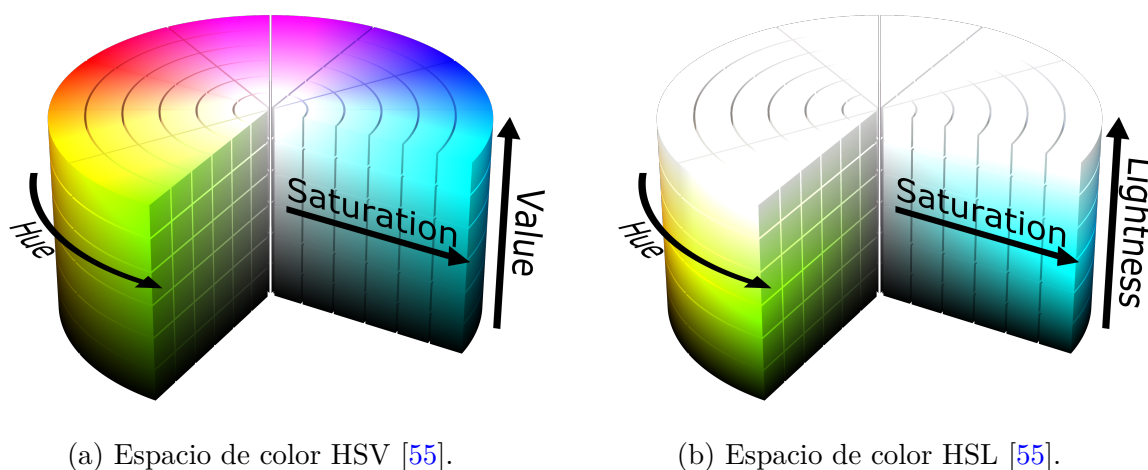


Figura 3-7: Espacios de color HSV-HSL. a) Espacio de color HSV. b) Espacio de color HSL.

En el modelo de color HSL, la *luminosidad* (L) se extiende en el rango entero desde el negro, hasta

el blanco, recorriendo el tono de color seleccionado. Por otra parte, en el modelo de color HSV, la componente de *brillo* (V), sólo posee la mitad de ese recorrido, extendiéndose desde el color negro, al tono seleccionado.

Los seres humanos, perciben y describen el color mediante los atributos de *matiz* (H), *saturación* (S) y *brillo* (V); expuestos a continuación y relacionados matemáticamente con el espacio de color RGB .

- La **matiz o tono** (**H**) (Ec. 3-20), describe la pureza de cada color en la gama cromática percibida.

$$H = \begin{cases} 0, & \text{si } C_{max} = C_{min} \\ \left(60^\circ \times \frac{G' - B'}{\Delta} + 360^\circ\right) \bmod 360^\circ, & \text{si } C_{max} = R' \\ 60^\circ \times \frac{B' - R'}{\Delta} + 120^\circ, & \text{si } C_{max} = G' \\ 60^\circ \times \frac{R' - G'}{\Delta} + 240^\circ, & \text{si } C_{max} = B' \end{cases} \quad (3-20)$$

Su valor, es representado como la componente angular en un rango de $[0^\circ, 360^\circ]$; siendo el ángulo 0° , el color rojo, 120° el color verde y 240° el azul (véase Figura 3-8).



Figura 3-8: Matiz o tono (H) en el modelo de color HSV [55].

- La **saturación** (**S**) (Ec. 3-21) se describe como la pureza relativa o el nivel de interacción o mezcla entre la matiz de un color puro determinado con la luz blanca.

$$S = \begin{cases} 0, & \text{sí } C_{max} = 0 \\ \frac{\Delta}{C_{max}} = 1 - \frac{C_{min}}{C_{max}}, & \text{sí } C_{max} \neq 0 \end{cases} \quad (3-21)$$

La saturación es descrita de forma radial. Su valor máximo se encuentra en la periferia del cilindro de color HSV (véase Figura 3-7a) y varía un rango de $[0\%, 100\%]$ (véase Figura 3-9).



Figura 3-9: Niveles de saturación (S) en el modelo de color HSV [55].

- El **brillo(V)** (Ec. 3-22), es un aspecto subjetivo al nivel de luminosidad que se encuentra asociada al color.

$$V = C_{max} \quad (3-22)$$

El nivel brillo, es representado en el eje vertical o altura del cilindro HSV (véase Figura 3-7a). Este, varía un rango de [0 %, 100 %] (véase Figura 3-10) y esta relacionado a la noción cromática de intensidad.



Figura 3-10: Niveles de brillo (V) en el modelo de color HSV [55].

Siendo, R', G', B' (Ec. 3-23), los valores normalizados de las componentes cromáticas del modelo de color R, G, B en el rango [0, 1].

$$R' = \frac{R}{255} \quad G' = \frac{G}{255} \quad B' = \frac{B}{255} \quad (3-23)$$

C_{max} (Ec. 3-24) y C_{min} (Ec. 3-25), corresponden a los valores máximos y mínimos de R', G', B' . Y, Δ (Ec. 3-26), a la diferencia existente entre estos.

$$C_{max} = \max(R', G', B') \quad (3-24)$$

$$C_{min} = \min(R', G', B') \quad (3-25)$$

$$\Delta = C_{max} - C_{min} \quad (3-26)$$

Implementación:

Las imágenes filtradas (véase Figura 3-6), obtenidas en la sección 3.2.3 a partir de la implementación del método de *clasificación de texturas multicanal*, empleando el filtrado de Gabor, son transformadas al modelo de color HSV aplicando una transformación en los espacios de color (Ec. 3-27).

$$A_{HSV} = \text{Gray}[Y_A] \text{ to HSV}[A] \quad (3-27)$$

Inicialmente, se realizó la conversión del modelo de color de escala de grises al modelo de color RGB (Gray[Y_A] to RGB[A]), empleando la ecuación 3-4, descrita en la sección 3.2.2. Seguidamente, empleando los modelos matemáticos expuestos en la presente sección, se obtiene la representación de las imágenes, en el modelo de color HSV (véase Figura 3-11).

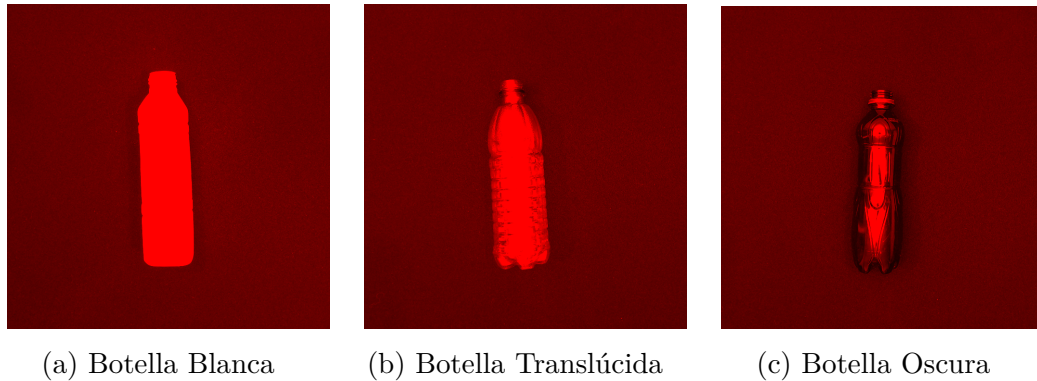


Figura 3-11: Imagen filtrada en el espacio de color HSV (Hue-Saturation-Value)

En las imágenes de la Figura 3-11, se aprecia una tonalidad de rojo predominante, correspondiente al valor angular 0° de la componente de *matiz* H . Sin embargo, esta matiz presenta niveles de grises considerables, congruente con bajos niveles de *saturación* S ; así, como niveles variables en la componente de *brillo* V ; resaltando el objeto de interés.

En la Figura 3-11c, el contorno del objeto presenta bajos niveles de brillo, siendo este un factor decisivo al momento de escoger los rangos de umbralización para el procesado de las imágenes que incluyan la diversidad característica de los objetos evaluados.

3.3. Segmentación de imágenes digitales

La segmentación de imágenes, es el proceso a través del cual, una imagen es dividida o fraccionada en zonas disjuntas e individualizadas mediante el etiquetado y la agrupación de píxeles que compartan algunas características visuales; posibilitando la distinción de los diversos objetos y el fondo. La segmentación, es empleada con el objeto de simplificar, localizar y/o delimitar las diferentes regiones de interés en una imagen específica [26, 53].

Los algoritmos de segmentación, se basan generalmente en la *discontinuidad* o *similitud* de los niveles de intensidad de los píxeles vecinos.

- Los algoritmos **basados en la discontinuidad**, dividen la imagen en función de los cambios bruscos en los niveles de intensidad. Entre estos encontramos: basados en la detección de puntos aislados, basados en la detección de líneas y basados en la detección de bordes.
- Los algoritmos **basados en la similitud**, dividen la imagen a partir de la búsqueda de zonas con valores similares de acuerdo a los criterios definidos. Entre estos encontramos: basados en

umbrales, basados en el crecimiento de regiones, basados en la división y fusión de regiones, basados en las técnicas de agrupamiento y basados en las técnicas de matching.

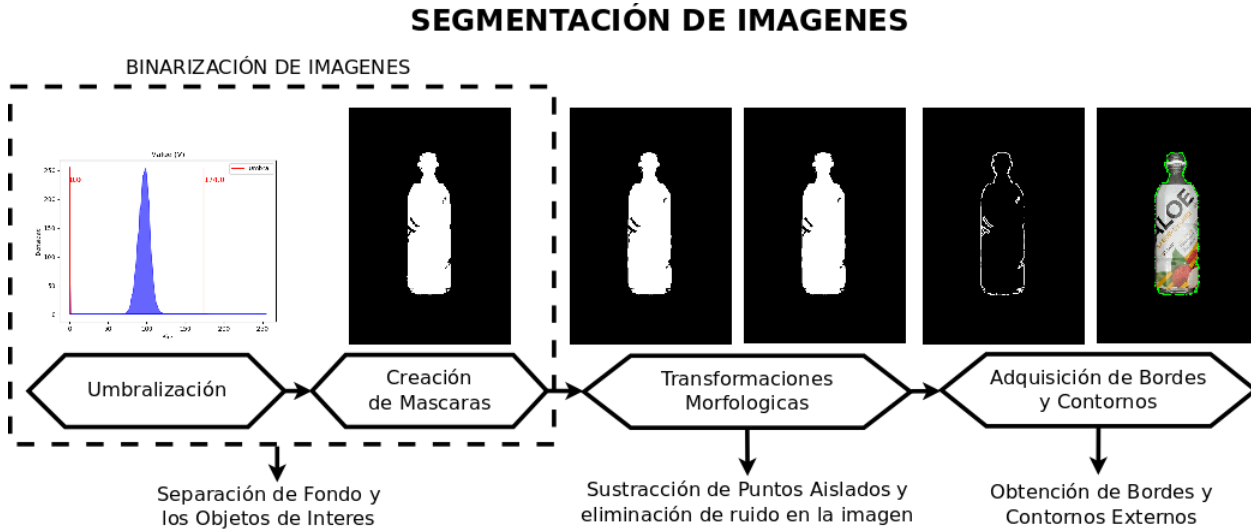


Figura 3-12: Proceso de segmentación de imágenes.

En esta sección, se lleva a cabo un proceso de segmentación de imágenes (véase Figura 3-12) utilizando algoritmos basados en la similitud. Inicialmente, en la sección 3.3.1, se realiza el análisis de los histogramas de las imágenes procesadas en la sección 3.2. Seguidamente, en la sección 3.3.2, se lleva a cabo el proceso de umbralización empleando el método de Otsu. Posteriormente, en la sección 3.3.3, las imágenes se someten a transformaciones morfológicas para eliminar puntos aislados y resultados indeseados que pudieron surgir durante las etapas anteriores. Finalmente, en la sección 3.3.4, se extraen los bordes y contornos de los objetos, proporcionando información sobre las fronteras de los elementos presentes en la escena. Este conjunto de técnicas permite una segmentación efectiva de la imagen, resaltando las regiones de interés y facilitando el análisis detallado de los objetos en la escena.

3.3.1. Análisis de histogramas

Dada la representación matricial de una imagen digital de N filas por M columnas (secc. 3.2.1). El **histograma** de una imagen $h(i)$ (Ec. 3-28), es un vector que describe la distribución frecuencial de valores de píxel de una imagen $f(m, n)$ en banda determinada δ . En este orden de ideas, para una imagen de $N - bits$ [38, 44], se tiene que:

$$h(i) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \delta(f(m, n) - i) \quad \forall i \in [0, 2^\alpha - 1] \quad (3-28)$$

Partiendo de lo anteriormente expuesto, y del estudio del espacio de color HSV realizado en la sección 3.2.4. Se realiza la obtención y análisis de los histogramas en cada una de los canales en el modelo de color HSV , para cada una de las tonalidades fronterizas estudiadas (véase Figura

3-13).

Implementación:

Los histogramas de los canales H , S , y V de las botellas translúcida y oscura, describen un comportamiento similar al observado en la Figura 3-13 para la botella de color blanco; donde, la distribución frecuencial de los valores de intensidad en los canales de *matiz* (H) (véase Figura 3-13a) y *saturación* (S) (véase Figura 3-13b), es cero; denotando la carencia total blancos en las imágenes (véase Figura 3-11) y manifestando una única tonalidad predominante, el rojo.

El histograma en el canal de *brillo* (V), presenta una distribución unimodal, abarcando toda la información presente del canal V (véase Figura 3-13c), en un rango aproximado entre los 60 bins y los 135 bins; incluyendo el máximo global del histograma que usualmente corresponde al fondo de la imagen.

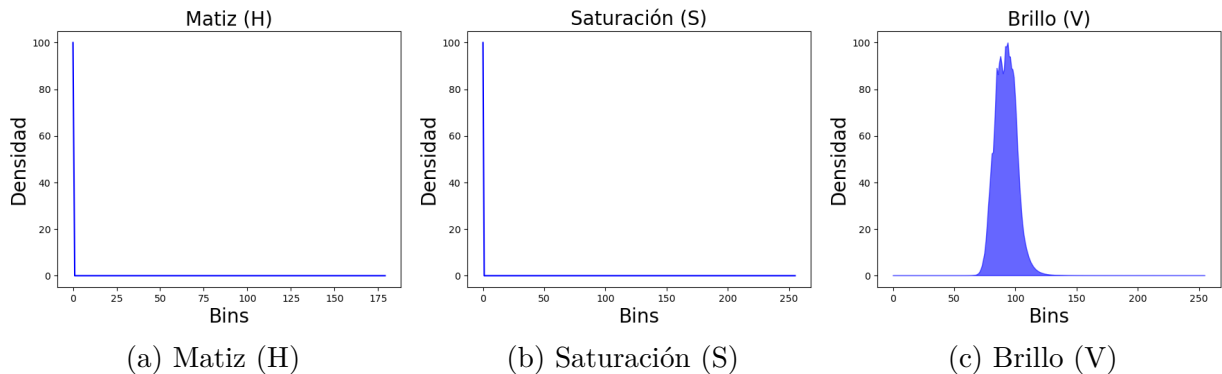


Figura 3-13: Histogramas de los canales de matiz (H), saturación (S) y brillo (V) de la botella blanca

La carencia de información observada en los canales de *matiz* H (véase Figura 3-13a) y *saturación* S (véase Figura 3-13b); contraria a la apreciable en el canal de *brillo* V (véase Figura 3-13c), responde al resultado de haber transformado la imagen de entrada RGB en una imagen en escala de grises.

En una distribución unimodal, los objetos aparecen poco contrastados respecto al fondo. No obstante, las botellas blanca (véase Figura 3-11a) y translúcida (véase Figura 3-11b), se aprecian claramente contrastadas, presentando niveles similares en la distribución de brillo. Por otra parte, en los bordes de la botella oscura (véase Figura 3-11c), se presentan bajos niveles de brillo, cercanos a los niveles que se perciben en el fondo de la imagen.

3.3.2. Umbralización o thresholding

La **umbralización** $f_U(x, y)$ (Ec. 3-29), consiste en suprimir todos los valores superiores o inferiores (igual a cero "0") de una imagen $f(x, y)$, en relación con un valor conocido como umbral T [31, 51].

$$f_U(x, y) = f(x, y) * (f(x, y) \geq T) \quad \therefore T \rightarrow Umbral \quad (3-29)$$

La **binarización**, es una variante de la umbralización que consiste en comparar cada píxel de una imagen, con un valor umbral (Ec. 3-30); clasificándolos de acuerdo a las relaciones de desigualdad matemática, asignando valores binarios 0 o 1.

$$f_B(x, y) = f(x, y) \geq T \quad \therefore T \rightarrow Umbral \quad (3-30)$$

Las técnicas de umbralización más conocidas, se fundamentan en el análisis y las aproximaciones estadísticas sobre el histograma de una imagen, empleando procedimientos paramétricos y no paramétricos. Sezgin y Sankur [51] clasificaron estas técnicas en métodos basados en la forma del histograma, la clusterización, la entropía, los atributos de la imagen, la información espacial y en características locales.

En la actualidad, existe una gran variedad de métodos empleados en la búsqueda del umbral óptimo. No obstante, la gran mayoría arroja resultados indeseados ante imágenes en ambientes poco o nada controlados, en presencia de ruido significativo y de iluminación variable. Por otra parte, el método de Otsu, es uno de los mejores métodos de umbralización empleados en imágenes obtenidas en entornos con elevada incertidumbre.

El método de Otsu

El método de Otsu [38], es un método de umbralización no supervisado basado en la clusterización; que realiza una búsqueda exhaustiva del umbral óptimo T mediante la maximización de la varianza interclase (between-class variance) [31, 38].

Sea n_i , el número de total de píxeles en el nivel de intensidad i . Sea N (Ec. 3-31), el número total de píxeles de una imagen $f(x, y)$; donde, p_i (Ec. 3-32) es considerada la distribución de probabilidad para cada nivel i [31, 38].

$$N = \sum_{i=0}^{L-1} n_i \quad (3-31)$$

$$p_i = \frac{n_i}{N} \quad (3-32)$$

Donde, el método de Otsu, inicialmente realiza una primera clasificación de los píxeles de la imagen $f(x, y)$ seleccionando un umbral t , usado para segmentar la imagen en dos clases: C_0 (Ec. 3-33) y C_1 (Ec. 3-33) [31, 38].

$$C_0 : \frac{p_0}{w_0(t)} \cdots \frac{p_t}{w_0(t)} \quad (3-33)$$

$$C_1 : \frac{p_{t+1}}{w_1(t)} \cdots \frac{p_{L-1}}{w_1(t)} \quad (3-34)$$

Seguidamente, se procede a recorrer el histograma de la imagen, incrementando progresivamente el valor del umbral t . Durante este proceso, se calculan las varianzas interclase de los grupos adyacentes a cada valor umbral determinado, siguiendo el método propuesto por Otsu [38]. La distribución inicial de probabilidad w_0 corresponde a los píxeles en el rango $[0, t]$, mientras que la segunda distribución de probabilidad está relacionada con los píxeles en el rango $[t + 1, L - 1]$. Aquí, w_0 (Ec. 3-36) y w_1 (Ec. 3-36) representan las distribuciones de probabilidad asociadas a las clases C_0 (Ec. 3-33) y C_1 (Ec. 3-33), respectivamente [31, 38]. Este proceso permite determinar el umbral óptimo que maximiza la varianza interclase y, por lo tanto, optimiza la separación entre las clases en la imagen.

$$w_0 = \sum_{i=0}^t p_i = w(t) \quad (3-35)$$

$$w_1 = \sum_{i=t+1}^{L-1} p_i = 1 - w(t) \quad (3-36)$$

A partir de estas relaciones matemáticas, en [38] se define la varianza entre clases σ_B^2 como:

$$\sigma_B^2 = w_0 w_1 (\mu_1 - \mu_0)^2 \quad (3-37)$$

$$\therefore \sigma_B^2(t) = \frac{[\mu_T w(t) - \mu(t)]^2}{w(t)[1 - w(t)]} \quad (3-38)$$

Siendo, μ_0 (Ec. 3-39) y μ_1 (Ec. 3-40), la media para las clases C_0 y C_1 ; y $\mu(t)$ y μ_T (Ec. 3-41), la media acumulativa en el rango $[0, t]$ y la media global respectivamente [38].

$$\mu_0 = \frac{1}{w_0} \sum_{i=0}^t i p_i = \frac{\mu(t)}{w(t)} \quad (3-39)$$

$$\mu_1 = \frac{1}{w_1} \sum_{i=t+1}^{L-1} i p_i = \frac{\mu_T - \mu(t)}{1 - w(t)} \quad (3-40)$$

$$\mu(t) = \sum_{i=0}^t i p_i \quad \mu_T = \sum_{i=0}^{L-1} i p_i \quad (3-41)$$

El objetivo principal del método de Otsu, es el cálculo exhaustivo de las diferentes varianzas interclase y la elección del valor umbral t^* óptimo; es decir, el valor umbral t obtenido de la mayor varianza interclase; siendo este, el umbral óptimo de la imagen $f(x, y)$:

$$\sigma_B^2(t^*) = \max_{0 \leq t \leq L} \{\sigma_B^2(t)\} \quad (3-42)$$

Implementación:

El cálculo del valor umbral se realizó utilizando los histogramas del canal de brillo (V) (véase Figura 3-14) mediante el método de Otsu. Este método comienza dividiendo la distribución de los

niveles de intensidad de los píxeles de la imagen $f(x, y)$, que se representan en el histograma en el rango $[0, L]$, en dos niveles: $[0, t]$ y $[t + 1, L - 1]$, mediante un umbral t . Este proceso segmenta la imagen en dos clases, C_0 y C_1 , asumiendo que representan el fondo de la imagen y el objeto de interés. Este enfoque de segmentación basado en la variabilidad y homogeneidad inter-categoría facilita la identificación del umbral óptimo que maximiza la separación entre las dos clases.

Como se evidencia en la Figura 3-14, el valor umbral obtenido para el histograma de la botella blanca (véase Figura 3-14a) es de mayor magnitud en comparación con el valor obtenido para la botella oscura (véase Figura 3-14c). Esto se debe a una mayor densidad y distribución de brillo en los píxeles de la botella blanca. Por otro lado, la botella translúcida (véase Figura 3-14b) presenta el valor umbral más bajo calculado, debido a su permisividad lumínica, permitiendo que la proyección parcial del color del fondo incida en sus píxeles y afecte su histograma. Estas variaciones en los valores umbrales son reflejo de las características intrínsecas de cada tipo de botella y resaltan la importancia de adaptar el proceso de segmentación a las particularidades de los objetos en la escena.

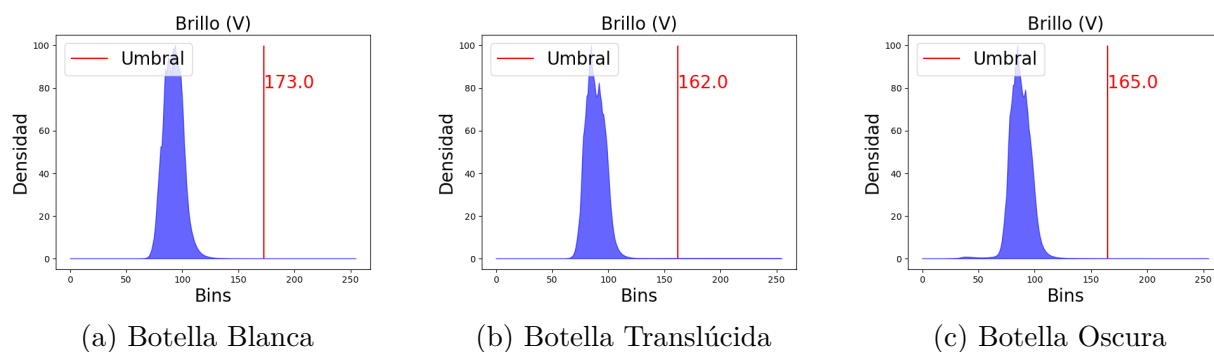


Figura 3-14: Histograma del canal de brillo (V) y valor umbral.

En la Figura 3-15, se presenta el resultado obtenido después de aplicar el proceso de umbralización a cada una de las botellas analizadas. Se destaca la pérdida de los contornos en la botella oscura (véase Figura 3-15c), debido a los bajos niveles de intensidad presentes en esa región, los cuales son cercanos a los percibidos en el fondo de la imagen. Este fenómeno ilustra los desafíos asociados con la segmentación de objetos de tonalidades oscuras cuando hay similitudes con el fondo en términos de intensidad, subrayando la importancia de emplear estrategias adicionales para mejorar la precisión en estos casos específicos.

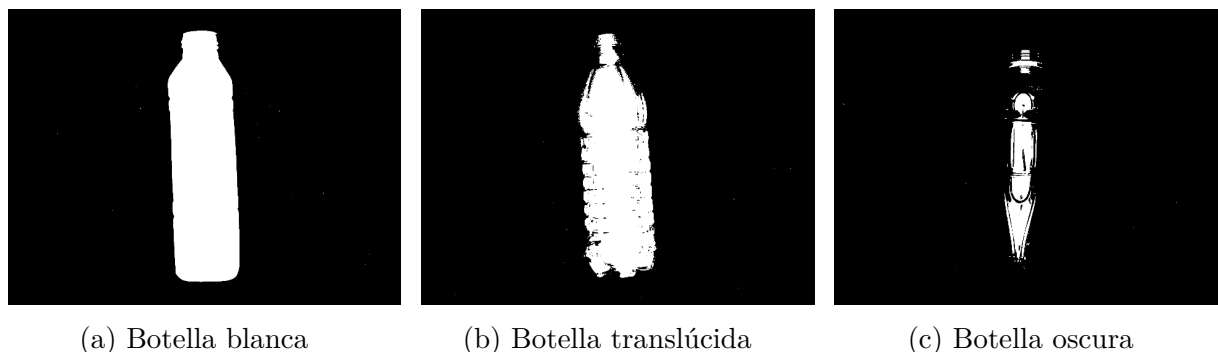


Figura 3-15: Imágenes resultantes después de aplicar del valor umbral en el proceso de segmentación.

A partir del valor umbral T obtenido para el canal de brillo mediante el método de Otsu (véase Figura 3-14), el análisis del histograma de *brillo* (V) realizado en la sección 3.3.1 y la pérdida de información en los contornos de la botella oscura después del proceso de umbralización (véase Figura 3-15c), se establece un margen de umbralización basado en las características congruentes observadas en los histogramas. Se construyen máscaras binarias, considerando que:

1. La mayor densidad de brillo observada en los histogramas se encuentra en el rango de $[60,135]$ niveles de intensidad (véase Figura 3-14). Para la segmentación de la imagen, se establece que el valor inferior del margen de brillo V_{bajo} será igual a 60.
2. El máximo global del histograma es un valor variable dependiente de la propiedad de reflexión lumínica y la tonalidad del objeto o de su etiquetado; el margen superior del nivel de brillo V_{alto} empleado en la binarización, será determinado por el calculo del valor umbral T .

Los resultados de la aplicación de las máscaras binarias construidas se presentan en la Figura 3-16.

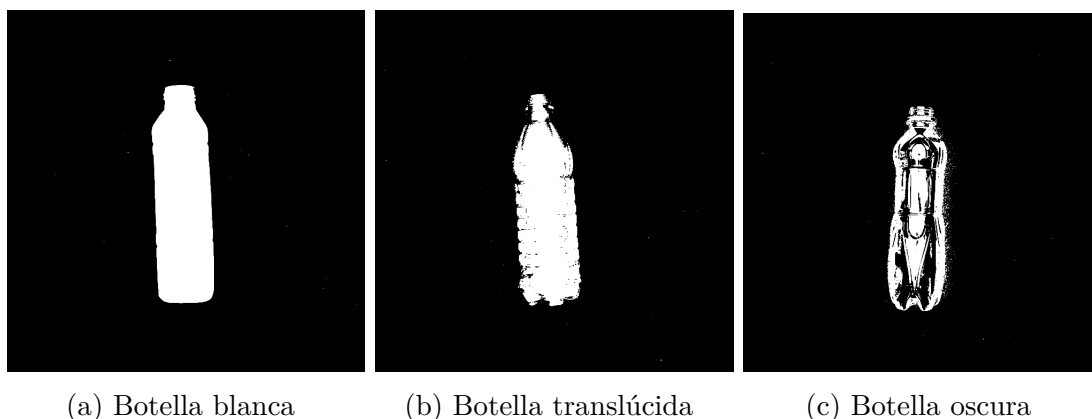


Figura 3-16: Creación de mascarar binarias

3.3.3. Transformaciones morfológicas

En matemáticas, la agrupación de elementos que comparten una serie de propiedades o similitudes, se le denomina conjuntos. Las transformaciones morfológicas en imágenes digitales, parten de la interacción de dos conjuntos, la imagen y el elemento estructurante, mediante operaciones empleadas con el fin de separar y modificar la forma y estructura de los objetos, permitiendo obtener contornos y descomponer formas complejas en estructuras primitivas; así como depurar entornos ruidosos.

En el ámbito de la morfología matemática, dos operadores fundamentales desempeñan un papel crucial: la dilatación y la erosión. La dilatación, detallada en la sección 3.3.3.1, consiste en expandir las regiones de píxeles más oscuros de una imagen, lo cual puede resultar útil para resaltar características específicas. Por otro lado, la erosión, explorada en la sección 3.3.3.2, implica contraer las regiones de píxeles oscuros, siendo valiosa para enfatizar detalles más finos en la imagen.

Estos dos operadores básicos proporcionan los cimientos para el desarrollo de operaciones morfológicas más avanzadas, como la apertura (consulte la sección 3.3.3.3) y el cierre (consulte la sección 3.3.3.4). La apertura implica la aplicación sucesiva de la erosión y la dilatación, y se utiliza para eliminar pequeños objetos o detalles, mientras que el cierre, que combina la dilatación seguida de la erosión, es eficaz para cerrar brechas en regiones de píxeles oscuros.

Estas operaciones morfológicas se emplean conjuntamente como filtros espaciales para reducir el ruido en una imagen, y su comprensión es esencial en el procesamiento de imágenes para la mejora de características específicas.

3.3.3.1. Dilatación (Dilate)

Dados dos conjuntos A y B de \mathbb{Z}^2 ; donde el conjunto A representa los elementos de la imagen $f(x, y)$ y B , el elemento estructurante. La dilatación, denotada por $A \oplus B$, se define como:

$$A \oplus B = \{z | (\hat{B}^5)_Z \cap A \neq \emptyset\} = \{c \in E^N | c = a + b \quad \forall a \in A \text{ y } b \in B\} \quad (3-43)$$

La dilatación, también denominada *crecimiento*, *llenado* o *expansión*, de una imagen A por el elemento estructurante B , consiste en aumentar gradualmente los límites de las regiones de píxeles de primer plano y reducir los vacíos internos. Esta operación es comúnmente empleada para incrementar el contorno de los objetos, con la finalidad de eliminar las discontinuidades presentadas en los bordes.

3.3.3.2. Erosión (Erode)

Dados dos conjuntos A y B de \mathbb{Z}^2 ; donde el conjunto A representa los elementos de la imagen $f(x, y)$ y B , el elemento estructurante. La erosión, denotada por $A \ominus B$, se define como:

$$A \ominus B = \{z | B_Z \subseteq A\} = \{x \in E^N | x = x + b \in A \quad \forall b \in B\} \quad (3-44)$$

La erosión $A \ominus B$, de una imagen A por el elemento estructurante B , consiste en reducir los límites de las regiones de los píxeles, incrementando los vacíos internos en esas áreas.

3.3.3.3. Apertura (Opening)

Dados dos conjuntos A y $B \in \mathbb{Z}^2$; La apertura del conjunto A por un elemento estructurante B , se define como:

$$A \circ B = (A \ominus B) \oplus B \quad (3-45)$$

Donde, la apertura $A \circ B$ es el resultado de aplicar sobre el conjunto A , una erosión, seguida de una dilatación. Esta operación es empleada en: la segmentación y simplificación de objetos, la extracción de formas en entornos ruidoso y la separación de objetos adyacentes.

3.3.3.4. Cierre (Closing)

El cierre del conjunto $A \in \mathbb{Z}^2$ por el elemento estructurante $B \in \mathbb{Z}^2$, está definida por:

$$A \bullet B = (A \oplus B) \ominus B \quad (3-46)$$

Donde, el cierre $A \bullet B$, primero se realiza una dilatación y posteriormente una erosión.

A partir de los conocimientos expuestos en la presente sección, se implementó una operación morfológica de *apertura*, seguida de un *cierre*, generando el denominado *filtro morfológico* f_m (Ec. 3-47).

$$f_m = \{[(A \ominus B) \oplus B] \oplus B\} \ominus B = (A \circ B) \bullet B \quad (3-47)$$

Esta operación de filtrado, permitió eliminar puntos aislados, reducir el ruido vecino a los bordes obtenidos.

3.3.4. Detección de bordes y contornos

La detección de bordes y contornos de objetos es una capacidad esencial de la visión humana, permitiendo la realización de tareas complejas como el emparejamiento de formas, la segmentación de imágenes y el reconocimiento de objetos [35].

En las últimas décadas, la investigación en detección de bordes ha sido extensa, siendo un paso clave en algoritmos de procesamiento de imágenes y visión artificial. Los bordes, áreas con cambios discernibles en la intensidad de la imagen, son esenciales para tareas como segmentación, descripción de características, reconocimiento de patrones, entre otras [3, 58].

La detección de bordes en imágenes digitales ha sido abordada mediante diversas aproximaciones, destacando especialmente las tecnologías clásicas y de aprendizaje profundo. Los operadores clásicos basados en gradientes, como Robert [49], Prewitt [32] y Sobel [40], que fueron inicialmente utilizados para esta tarea, mostraron limitaciones al no producir bordes nítidos y ser muy susceptibles al ruido. Otros operadores como Laplaciano, Gaussian-Laplacian y Canny [52, 59] también presentaron desventajas, como la identificación de bordes incorrectos y errores significativos en bordes curvos.

La técnica de bordes de Canny, aunque avanzada, enfrenta desafíos en la selección de umbrales, generando confusiones al determinar si un píxel es un borde o no. En general, se observa que la mayoría de los métodos clásicos de detección de bordes son sensibles al ruido y dependientes de áreas y momentos [?].

En la detección de bordes, la ubicación exacta de estos se ve desafiada por la incertidumbre proveniente de factores como el ruido en las imágenes y las variaciones en la intensidad de los píxeles, lo que dificulta la tarea de identificar dónde se encuentran los bordes en la imagen.

La dificultad en la localización de los bordes puede dar lugar a dos tipos de errores que impactan directamente en la efectividad de los algoritmos de detección. Por un lado, los falsos negativos ocurren cuando el algoritmo no logra identificar algunos de los bordes reales presentes en la imagen, lo que significa que detalles importantes de los objetos pueden pasar desapercibidos, afectando la calidad de la detección.

Por otro lado, los falsos positivos se presentan cuando el algoritmo señala puntos que no son bordes como si lo fueran. Este tipo de error puede surgir al interpretar incorrectamente variaciones en la intensidad de los píxeles como bordes. En conjunto, estos errores pueden comprometer la precisión y confiabilidad de los algoritmos de detección de bordes, subrayando la importancia de desarrollar métodos robustos capaces de enfrentar la inherente incertidumbre en la tarea de detección.

Implementación:

En esta etapa, se implementó el algoritmo de detección de bordes de Canny. Este algoritmo utiliza un filtro basado en la primera derivada de una gaussiana y es reconocido por su capacidad para identificar bordes en una imagen a través de múltiples etapas. El resultado de aplicar este algoritmo es una imagen que resalta de manera evidente los bordes presentes en la Figura 3-17.

Posteriormente, se realizó una operación de morfología de imágenes utilizando un elemento estructurante en forma de cruz aplicando un gradiente de 5x5. Esta operación específica es conocida como gradiente morfológico, diseñada para resaltar aún más los bordes de la imagen original.

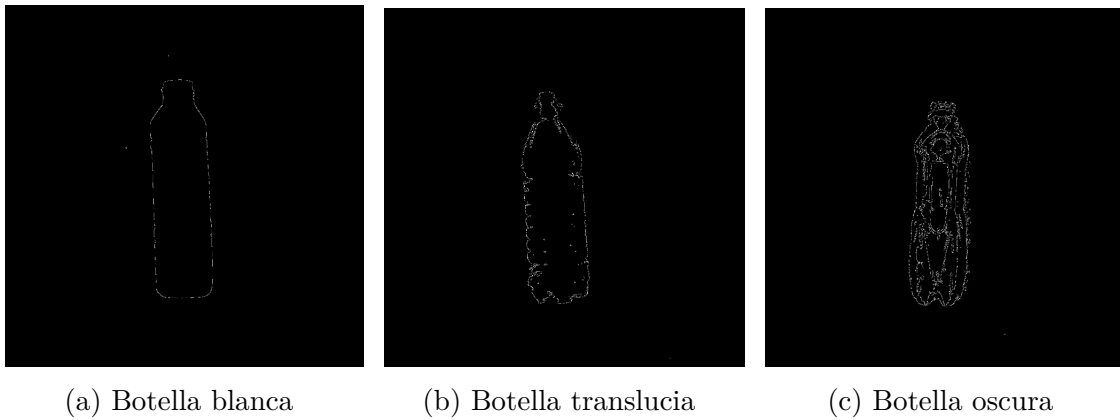


Figura 3-17: Bordes del Objeto.

En la Figura 3-17, se pueden observar los bordes obtenidos en la etapa actual de segmentación. En las figuras 3-17a y 3-17b, se muestra el contorno de la botella blanca y translúcida, respectivamente. Sin embargo, se nota que estos contornos están reducidos, y se ha perdido información debido a los procesos de filtrado y transformaciones morfológicas.

En la Figura 3-17c, después de aplicar los procesos de creación de máscaras, se logra obtener el contorno de la botella oscura en un fondo oscuro. Los contornos internos son el resultado de los reflejos luminosos focalizados del objeto.

4 Sistema automático de reconocimiento de imágenes

4.1. Introducción

En la fase inicial de la segmentación de imágenes, se aplicaron métodos como la umbralización, la creación de máscaras y la detección de bordes, los cuales fueron exhaustivamente explorados en la sección 3.3 y arrojaron resultados satisfactorios en términos generales. No obstante, al adentrarse en el análisis detallado de las características particulares de ciertos objetos, como se discutió en la sección 2.3, y al tener en cuenta las variaciones de luminosidad presentes en la escena, se hicieron evidentes ciertas limitaciones que afectaron el proceso.

En el presente capítulo se aborda la implementación de un sistema de detección, clasificación y segmentación multiclase, conocido como segmentación agnóstica. La segmentación es una tarea fundamental en visión por computadora, ya que consiste en categorizar las imágenes en partes diferenciadas y significativas. Este sistema automático de reconocimiento de imágenes digitales permite identificar de manera efectiva objetos y contornos en un escenario, basándose en un conjunto de características similares y previamente aprendidas.

4.2. Segmentación de imágenes

En los últimos años, la investigación en Redes Neuronales Convolucionales (CNN) ha experimentado avances notables, incluyendo métodos como R-CNN, Fast R-CNN y Faster R-CNN, diseñados para superar limitaciones previas [18].

M Loknath et al. proponen un algoritmo que emplea Fast R-CNN y RPN para la detección de objetos, destacando la eliminación del costo asociado a obtener una Región de Interés (ROI) mediante una red unificada [16]. Obtienen eficiencia con una red de detección que ofrece características de convolución de imagen completa mediante RPN, marcando así una mejora significativa en la eficiencia de detección de objetos.

Shaqing Ren y col. propusieron un algoritmo que utiliza Fast R-CNN y RPN para la detección de objetos. Lograron reducir el tiempo de ejecución de las redes de detección mediante avances como SPPnet [15]. La introducción de una red de detección que ofrece características de convolución de imagen completa mediante RPN mejoró la eficiencia.

Juan Du propuso "You Only Look Once"(YOLO) [1, 46]. YOLO rompe con la tradición de la familia CNN y mejora significativamente la detección de objetos de manera simple y eficiente. Supera a Faster R-CNN en velocidad. Posteriormente, se presentan las versiones YOLO V1 y V2, destacando su equilibrio óptimo entre velocidad y exactitud, así como su capacidad sólida para analizar la imagen completa [45].

Hoy en día, se presenta YOLOv4, una evolución que incorpora técnicas avanzadas como el aumento de datos Mosaic. Su red principal se fundamenta en CSPDarknet53, fusionando conexiones parciales de etapas cruzadas (CSP) con Darknet53 para reducir la complejidad computacional y mejorar la capacidad de aprendizaje. En comparación con sus predecesores, YOLOv4 representa un avance significativo al abordar problemas de múltiples escalas y mejorar la fusión de características [46, 60].

En relación con la detección de objetos no vistos, se abordan problemas como la segmentación de instancias y la clasificación adaptativa. Diversos enfoques, como SD Mask R-CNN, Gouda et al., UCN, UOIS, MSMFormer, UOAI, e INSTR, utilizan imágenes RGB y de profundidad, así como técnicas de clustering y arquitecturas de transformers para abordar estos desafíos [1, 8, 16, 45]. También se menciona FewSol, que se centra en la detección de objetos en escenarios de pocos ejemplos, combinando la segmentación de objetos no vistos y la clasificación de pocos ejemplos [61].

La investigación en redes neuronales convolucionales aborda avances para la detección de objetos, destacando métodos como Fast R-CNN, RPN, y YOLO, y explorando problemas específicos como la detección de objetos no vistos, la segmentación de instancias y la clasificación adaptativa [1, 8, 45, 46, 60].

4.3. Segmentación agnóstica

En ciertos casos de sistemas de agarre robótico, no es imperativo llevar a cabo un aprendizaje explícito de las categorías de objetos. Más bien, la tarea se centra en realizar la clasificación utilizando un conjunto de imágenes específicas de cada objeto [20, 21, 24]. Es esencial destacar que las clases de objetos en este contexto difieren de las clases convencionales utilizadas en detectores de objetos, ya que cada objeto es considerado como una clase única en sí misma.

Este enfoque resulta beneficioso en situaciones en las cuales la recopilación de un conjunto de datos es impracticable debido a la falta de experiencia técnica o restricciones de tiempo [39]. Asimismo, se emplea en contextos con una gran variedad de objetos, como en el ámbito logístico e industrial, donde un detector de objetos debe ser capaz de ajustarse a un conjunto flexible de clases [23].

Se resalta la distinción entre las redes de coincidencia de plantillas basadas en aprendizaje profundo y los detectores de objetos adaptativos de clase [21, 24]. Mientras que las primeras están diseñadas para detectar una única clase de objeto, los últimos deben tener la capacidad de detectar

simultáneamente diversas clases, siendo especialmente eficientes en entornos con múltiples objetos.

En [21] propone un detector de objetos adaptativo de clase para aplicaciones de agarre robótico, que consta de dos etapas fundamentales: la segmentación de objetos no vistos y la clasificación adaptativa. Este enfoque emplea de red siamesa profunda, para llevar a cabo la clasificación, entrenando la red únicamente con imágenes RGB sin incorporar objetos de prueba.

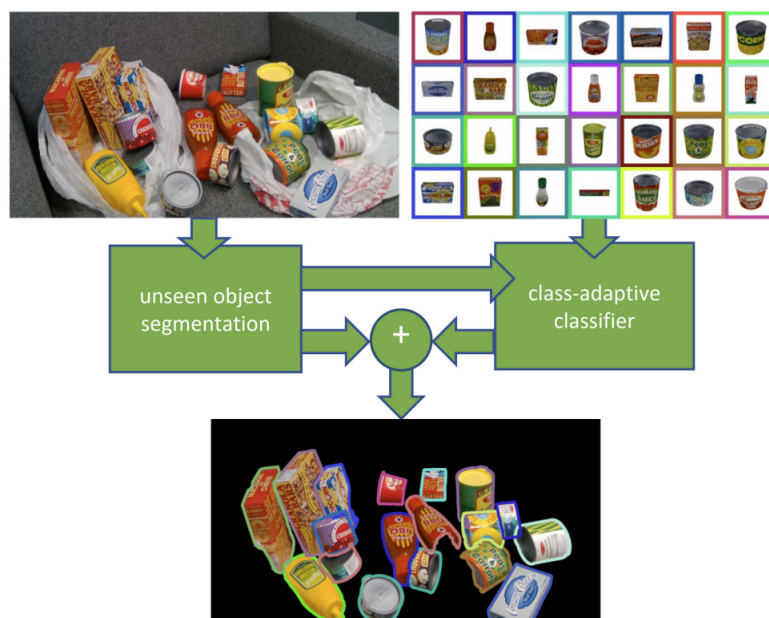


Figura 4-1: Sistema de segmentación agnóstica multi-clase [21]

Figura 4-1 se representa la metodología de dos etapas basada en la combinación de enfoques anteriores junto con la implementación de una red neural siamesa [21]. El enfoque de dos etapas presenta ventajas significativas en el contexto del agarre robótica, ya que aborda la detección de objetos no vistos. Entre los beneficios se incluyen la capacidad para detectar simultáneamente múltiples objetos, la separación del proceso de desarrollo de segmentación y la adaptabilidad a soluciones de problemas relacionados [21, 36, 39].

4.4. Implementación

En la fase de desarrollo de la etapa de segmentación de imágenes, se implementaron inicialmente métodos de umbralización, creación de máscaras y detección de bordes, como se describió en la sección 3.3. Estos métodos ofrecieron resultados satisfactorios en términos generales. Sin embargo, al profundizar en el análisis de las características específicas de ciertos objetos (secc. 2.3) y considerar las variaciones de luminosidad en la escena, se revelaron limitaciones que afectaban el proceso de segmentación.

Algunos desafíos significativos que surgieron incluyen la pérdida de información en los contornos de los objetos, la distorsión de los mismos debido a su proximidad espacial y la dificultad para que el sistema se adapte a la diversidad de variaciones luminosas en la escena. Estos aspectos comprometieron la robustez y eficacia del sistema propuesto. La mejora de estos aspectos se convirtió en una prioridad para garantizar una mejor segmentación en condiciones diversas.

Para abordar la problemática identificada, se optó por emplear el modelo de segmentación *Image Agnostic Segmentation* mediante la biblioteca Detectron 2. Este modelo destaca por ser un clasificador adaptativo de clases que utiliza una arquitectura de red neuronal siamesa, donde cada objeto específico representa su propia clase. En esencia, el modelo evalúa la similitud entre dos imágenes: la imagen de consulta y la imagen de la galería.

Detectron 2 es una biblioteca de código abierto diseñada específicamente para abordar tareas de detección de objetos en visión por computadora. Su arquitectura modular proporciona flexibilidad, permitiendo la combinación de diversos componentes y facilitando la creación y el entrenamiento de modelos personalizados. Detectron 2 es versátil y abarca diversas tareas, incluyendo segmentación semántica y detección de keypoints.

Detectron 2 hace uso de las capacidades avanzadas de PyTorch ofreciendo implementaciones de arquitecturas populares, entre las cuales se incluyen ResNet y FPN. Además, proporciona funciones específicas como segmentación panóptica, Densepose, Cascade R-CNN, cuadros delimitadores rotados, PointRend, DeepLab, entre otras.

Se empleó una muestra compuesta por 180 fotografías, abarcando un conjunto total de 1532 imágenes de botellas plásticas. Es esencial destacar que la evaluación de la efectividad se basó en el conteo manual de objetos identificados, actuando como punto de referencia en comparación con el conteo optimizado de contornos realizado por el sistema de detección.

Umbral de Confianza	No. Botellas Detectadas	% Detección
0.9	≈ 1352	88.25 %
0.8	≈ 1392	90.86 %
0.7	≈ 1471	96.54 %
0.6	≈ 1471	96.54 %

Tabla 4-1: Tasa de detección de botellas vs umbral de confianza.

Como se puede apreciar en la Tabla 4-1, el porcentaje de éxito en la detección alcanzó el 96.54 % al emplear un umbral de puntuación (*score thresh*) de 0.72. Al disminuir el valor de umbral, no se observó una variación significativa en el rendimiento del sistema. Sin embargo, al aumentar el umbral hasta 0.9, el porcentaje de reconocimiento descendió a un 88.25 %, afectando principalmente a los objetos que presentaban algún contacto en sus superficies.

El proceso de contraste entre el número de objetos identificados de forma manual y el conteo optimizado de contornos por el sistema revela las discrepancias y similitudes en las detecciones. La efectividad, expresada como un porcentaje, refleja la proporción de botellas correctamente identificadas en comparación con el total presente en la muestra.

5 Generación Automática de Regiones de Agarre (GARA)

5.1. Introducción

Una de las problemáticas más destacadas en la investigación actual en robótica surge de la falta de conocimiento acerca de las zonas a las que los manipuladores pueden acceder en relación con los objetos.

Un agarre exitoso se basa en la comprensión de los atributos de la pinza, la geometría del objeto y su representación, lo que permite la generación de múltiples soluciones de agarre a partir de regiones de agarre y parámetros de aproximación específicos. Sin embargo, en la práctica, estas suposiciones suelen resultar ineficaces debido a la presencia de ciertas incertidumbres.

Este capítulo se enfoca en el desarrollo de un sistema innovador para la generación automática de regiones de agarre para un gripper de 2 dedos. Este sistema simplifica estructuras geométricas complejas al aproximarse a formas geométricas primitivas, como se detalla en la sección 5.2. En la sección 5.3, se lleva a cabo un análisis detallado y se presenta la construcción de nuestro sistema de generación automática de regiones de agarre.

Para evaluar la pertinencia de las regiones de agarre generadas, se introduce en la sección 5.4 un método que considera varios factores cruciales. Entre ellos se encuentran la proximidad entre objetos, la ubicación espacial de la región generada en el objeto analizado y el área de contacto abarcada por dicha región.

Este enfoque integral y la consideración de factores específicos, como la geometría primitiva y la configuración del gripper, mejoran significativamente la generación de regiones de agarre y su pertinencia en aplicaciones prácticas.

5.2. Generación y análisis del rectángulo delimitador

Dado un conjunto finito de m -puntos en un espacio bidimensional, es posible encontrar un casco poligonal convexo y contenedor compuesto por n -puntos. En este contexto, es importante destacar que el número total de m -puntos contenidos en dicho casco puede potencialmente superar el número de los n -puntos que conforman el polígono envolvente [13]. Cabe resaltar que la presencia

del conjunto de puntos interiores no afecta la estructura del casco generado. Este casco poligonal representa una aproximación inicial a una forma geométrica primitiva capaz de contener el conjunto de puntos que describen un objeto observable en una imagen [13].

Una vez obtenido el casco poligonal del objeto a analizar, se procede a la construcción de n -rectángulos delimitadores de área mínima. Este proceso implica iterar sobre cada uno de los bordes del polígono convexo encontrado. La iteración se realiza proyectando los vértices del polígono en la línea del borde correspondiente [13]. La máxima distancia observada entre los vértices proyectados en la línea del borde del polígono determina el *ancho* del rectángulo delimitador. De manera análoga, la máxima distancia entre los vértices proyectados en la línea perpendicular al borde del polígono establece la dimensión *alto* del rectángulo delimitador [13].

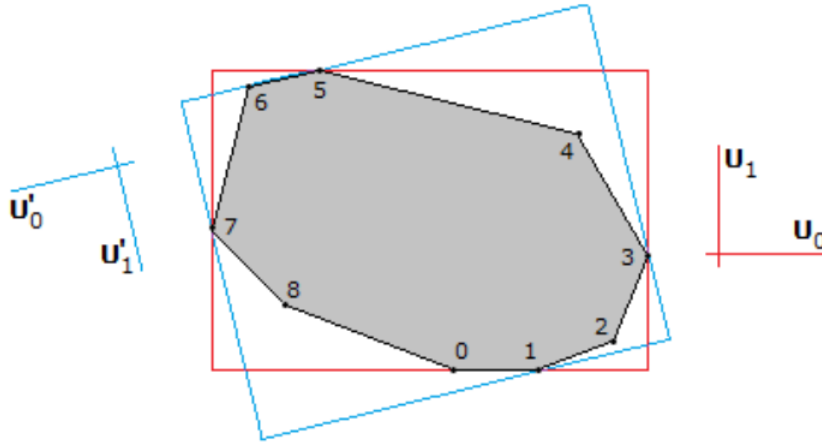


Figura 5-1: Configuración típica de un rectángulo delimitador para un casco poligonal de ocho (8) vértices [13].

La Figura 5-1 muestra un rectángulo en color rojo, orientado según el eje U_{01} , y apoyado en el borde $\langle 0, 1 \rangle$, cruzando los vértices $\langle 1, 3, 5, 7 \rangle$ [13]. Para determinar la rotación mínima y avanzar al siguiente rectángulo, se elige el borde $\langle 5, 6 \rangle$ como el próximo en ser procesado. Es relevante señalar que un vértice de un polígono puede tener asociados dos bordes de un rectángulo; en tal caso, ese vértice se considera una esquina del rectángulo [13].

El siguiente ángulo candidato de rotación, estará formado por los bordes: Inferior $\langle 1, 2 \rangle$, derecho $\langle 3, 4 \rangle$, superior $\langle 5, 6 \rangle$, izquierdo $\langle 7, 8 \rangle$; donde, la configuración conduciría igualmente a la intercepción de los vértices $\langle 1, 3, 5, 7 \rangle$ [13].

El rectángulo siguiente, visualizado en la Figura 5-1 y destacado en color azul, se compone de los vértices $\langle 6, 7, 1, 3 \rangle$. Este rectángulo está orientado en la dirección del eje U'_{01} y su nuevo borde inferior, formado por los vértices $\langle 5, 6 \rangle$, se elige como el punto de apoyo [13].

El análisis anteriormente descrito para calcular el rectángulo delimitador de área mínima de un

polígono convexo de n vértices se basa en el algoritmo de los *calibradores giratorios* [56]. Actualmente, se dispone de diversos algoritmos que se fundamentan en este método [42, 56], los cuales son ampliamente utilizados para abordar problemas de optimización en distintas aplicaciones.

Estos algoritmos se aplican en diversas áreas, como el cálculo del ancho de un conjunto de puntos [5, 22], la determinación de la distancia entre dos polígonos convexos [3, 56], el análisis del recorrido espacial de dos polígonos [4, 33], entre otros.

En la biblioteca computacional OpenCV, originalmente desarrollada por Intel desde 1999 y reconocida como Open Computer Vision (Visión Artificial Abierta), se incluye la función `cv2.boundingRect()`, comúnmente referida como el **rectángulo delimitador recto**¹. Esta función genera un rectángulo aproximado alrededor de una imagen binaria, destacando la región de interés (ROI). No obstante, el rectángulo delimitador resultante carece de información relevante sobre la orientación o inclinación del área demarcada.

En contraste, la función `cv2.minAreaRect()`, conocida como el **rectángulo rotado** y empleada en el análisis actual, permite obtener el área rectangular mínima alrededor de un contorno, de manera similar a la función `cv2.boundingRect()`. Además, considera la rotación del rectángulo y devuelve una estructura `Box2D` que incluye información sobre el centro (x, y) , el ancho (W), el alto (H) y el ángulo de rotación del rectángulo delimitador ($\angle\beta$). Cabe destacar que se requiere utilizar la función `cv2.boxPoints()` para obtener los puntos correspondientes a los cuatro vértices del rectángulo calculado.

El ángulo de rotación ($\angle\beta$) del rectángulo delimitador puede variar en el rango $[-90^\circ, 0^\circ)$, excluyendo el valor cero (0). Este ángulo no cuenta con un eje de referencia fijo y refleja la inclinación del segmento de recta formado entre los puntos P_0 y P_3 , como se ilustra en la Figura 5-2.

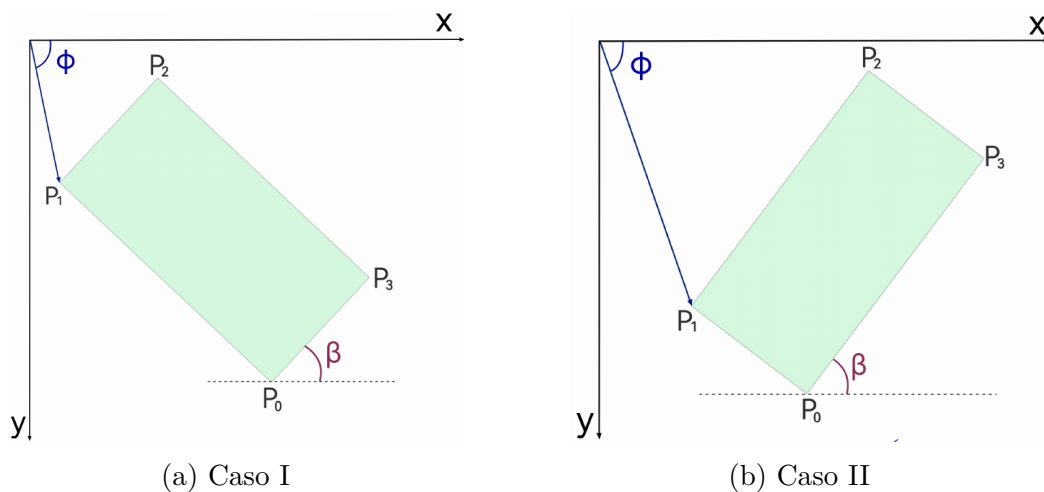


Figura 5-2: Relación entre el ángulo de rotación $\angle\beta$ y el segmento de recta $\overline{P_0P_3}$.

¹Función de OpenCV.

Partiendo del conocimiento de que cada uno de los objetos cilíndricos estudiados son representados por la aproximación de un polígono irregular aproximado en el espacio de la imagen, así como por la magnitud del ancho (w) y el alto (h) de los mismos. La orientación de estos, estará determinada por el discernimiento del lado ancho (w) y el lado alto (h) del polígono encontrado. De esta manera:

- Sí $\|\overline{P_3P_0}\| = \min(w,h)$, entonces $\|\overline{P_3P_0}\|$ corresponde al lado *ancho* del objeto (Fig. 5-2a).
- Sí $\|\overline{P_3P_0}\| = \max(w,h)$, entonces $\|\overline{P_3P_0}\|$ corresponde al lado *alto* del objeto (Fig. 5-2b).



Figura 5-3: Generación del rectángulo delimitador de mínima área.

5.3. Generación de las regiones de agarre

El Sistema de Generación Automática de Regiones de Agarre se fundamenta en la información obtenida a partir del rectángulo delimitador de área mínima, el cual está definido por el centro (x_c, y_c) , el ancho w_o , el alto h_o , y el ángulo de rotación $\angle\phi$ (véase la Figura 5-4a).

El distanciamiento entre los centros de las regiones de agarre se establece mediante un cálculo iterativo y progresivo de la distancia euclidiana $d \in \mathbb{R}^2$ con respecto al punto central del rectángulo delimitador de área mínima $P_C = (x_c, y_c)$. La distancia euclidiana entre dos puntos P_i y P_f , se determina mediante la siguiente ecuación:

$$d_E(P_i, P_f) = \sqrt{(y_f - y_i)^2 + (x_f - x_i)^2} \quad \therefore d_E \in \mathbb{R}^+ \quad (5-1)$$

Dado que el agarre de los objetos cilíndricos generalmente es realizado por el lado más angosto, el *ancho*; y, analizando el rectángulo delimitador $\square ABCD$ de la Figura 5-4a, se infiere que:

- El alto h_o (véase la Figura 5-6b), limita el número total de regiones de agarre consecutivas calculables con características específicas.

- El ancho w_o (véase la Figura 5-6b), determina el valor mínimo de apertura del gripper sobre el objeto que se desea obtener.

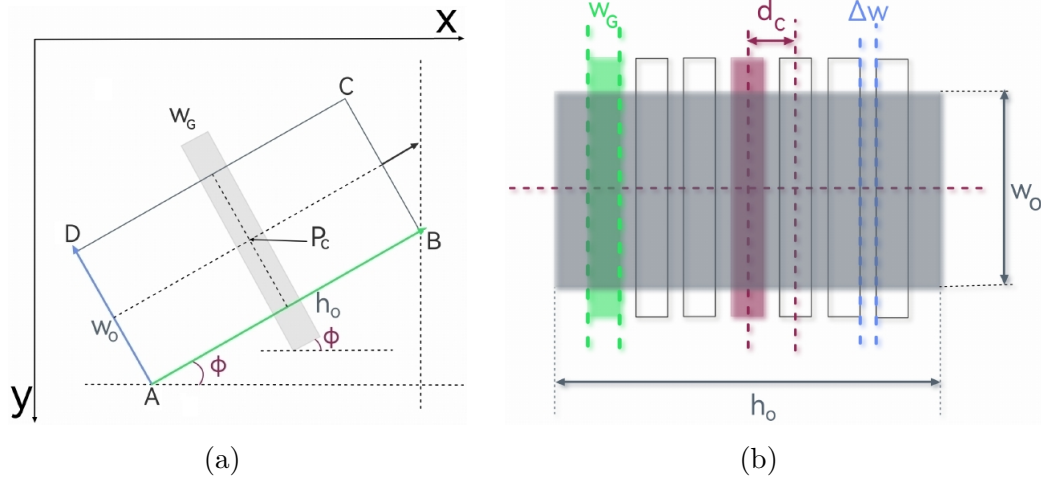


Figura 5-4: Generación de las regiones de agarre

Como se observa en la Figura 5-4b. La distancia entre los centros sucesivos de las regiones de agarre d_C (Ec. 5-2), se define como la suma del ancho dactilar del gripper w_G^2 (véase la Figura 5-4a y 5-4b) y la distancia de separación entre regiones de agarres Δ_w , distancia asignada con el objeto de evitar la superposición de las regiones contiguas agarre; donde, con fines prácticos, se define en la ecuación 5-3, como una porción porcentual de w_G , en el rango [15 %, 50 %].

$$d_C = w_G + \Delta_w \quad (5-2)$$

$$\therefore \Delta_w = \frac{\delta}{100} \cdot w_G \quad \forall \delta \in [15, 50] \quad (5-3)$$

La cantidad total de centros de regiones de agarre, representada por η_C (Ec. 5-4), se define como:

$$\eta_C = \mathbb{Z} \left\{ \frac{h_o}{d_C + \frac{5}{4}\Delta_w} \right\} + 1 \quad \therefore \eta_C \in \mathbb{Z}^+ \quad (5-4)$$

En la ecuación 5-4, la adición de $\frac{5}{4}\Delta_w$ en el divisor tiene como objetivo prevenir posibles desbordamientos en el área distal de contacto del gripper, especialmente en las regiones limítrofes de contacto con el objeto.

Para discernir la orientación del rectángulo delimitador de área mínima, teniendo en cuenta la relación entre su altura y ancho, se realiza un análisis basado en lo expuesto en la sección 5.2 y lo observado en las figuras 5-4a y 5-5. Se llega a la conclusión de que:

- Sí, $\|\overline{AB}\| = \min(w_o, h_o)$, entonces el segmento \overline{AB} corresponde al lado *ancho* del objeto. En este caso, el ángulo de inclinación de las regiones de agarre, $\angle RG$, es igual a $\angle \phi - 90^\circ$ (véase Figura 5-5a).

²El ancho dactilar del gripper $w_G = 0,15 * h_o$ para efectos de simulación.

- Sí, $\|\overline{AB}\| = \max(w_o, h_o)$, entonces el segmento \overline{AB} corresponde al lado *alto* del objeto. En este escenario, el ángulo de inclinación de las regiones de agarre, $\angle RG$, es igual a $\angle \phi$ (véase Figura 5-5b).

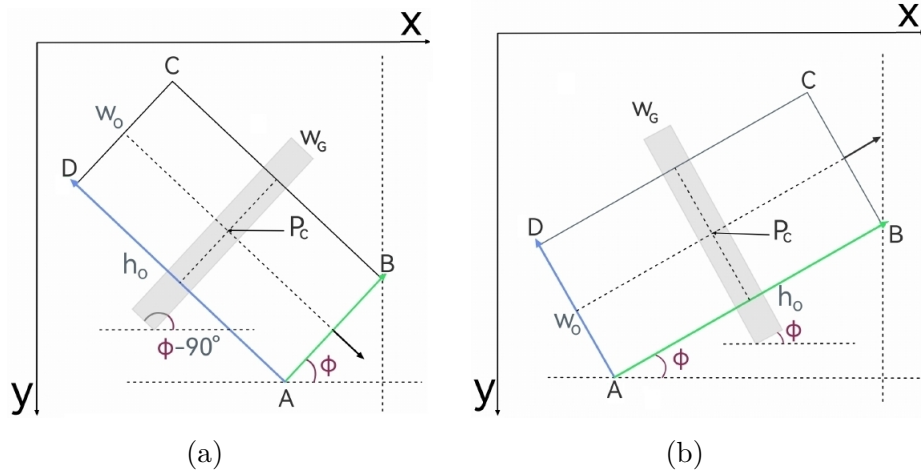


Figura 5-5: Relación entre el ancho w_o , el alto h_o y el ángulo de inclinación de las regiones de agarre $\angle RG$.

Partiendo del punto central $P_C = (x_c, y_c)$ (véase Figura 5-5a, 5-5b y 5-6b) y el ángulo de rotación $\angle \phi$; se establecen los centros de las regiones de agarre (véase Figuras 5-4b y 5-6c) a lo largo del rectángulo delimitador de área mínima, separados por una distancia d_C (Ec. 5-2). Para ello, inicialmente se obtiene el trazado de la recta $\in \mathbb{R}^2$ (Ec. 5-5) que pasa por el punto P_C y tiene una pendiente de inclinación m_{RG} ³, la cual está relacionada con la pendiente m_{AB} (Ec. 5-6) del segmento \overline{AB} (véase Figura 5-5).

$$Ax + By + C = 0 \rightarrow y = mx + b \quad (5-5)$$

$$m_{AB} = \tan \phi \quad \forall \quad \phi \neq 90^\circ \quad (5-6)$$

Donde m_{RG} representa la pendiente de inclinación de la región de agarre para todo $\phi \neq 90^\circ$, entonces:

- Sí, $\|\overline{AB}\| = w_o$ (véase Figura 5-5a), entonces, la pendiente $m_{RG} = -1/m_{AB}$.
- Sí, $\|\overline{AB}\| = h_o$ (véase Figura 5-5b), entonces, la pendiente $m_{RG} = m_{AB}$.

Dado un punto $P_C = (x_c, y_c)$, b estará determinada por la ecuación 5-7:

$$b = y_c - (m_{RG} \cdot x_c) \quad (5-7)$$

Dado un punto $P_C = (x_c, y_c)$ y un ángulo $\phi = 90^\circ$, donde:

³ m_{RG} es la pendiente de inclinación u orientación de la(s) región(es) de agarre obtenida(s) para un objeto determinado.

- Sí, $\|\overline{AB}\| = w_o$ (véase Figura 5-5a), entonces:

$$x = [x_c, x_c]. \quad (5-8)$$

$$y = [y_c, y_c] + [-d_C \cdot i, d_C \cdot i] \forall i \in [1, \eta_C]. \quad (5-9)$$

- Sí, $\|\overline{AB}\| = h_o$ (véase Figura 5-5b), entonces:

$$x = [x_c, x_c] + [-d_C \cdot i, d_C \cdot i] \forall i \in [1, \eta_C]. \quad (5-10)$$

$$y = [y_c, y_c]. \quad (5-11)$$

De esta manera, se establecen los centros de las regiones de agarre en concordancia con el ángulo de rotación del rectángulo delimitador, así como la relación entre su altura h_0 y ancho w_0 . Estos centros sirven de base para la proyección de las regiones de agarre correspondientes a un objeto específico.

5.3.1. Implementación

En el proceso de generación de regiones de agarre, la Figura 5-6a ilustra el contorno encontrado para un objeto específico después de aplicar procesos de filtrado de imágenes, segmentación y reconocimiento automático de objetos. En la Figura 5-6b, utilizando los métodos descritos en el capítulo 5, sección 5.2, se determina el rectángulo delimitador de área mínima, obteniendo el punto central del objeto $P_C = (x_c, y_c)$ y los cuatro puntos correspondientes de dicho rectángulo.



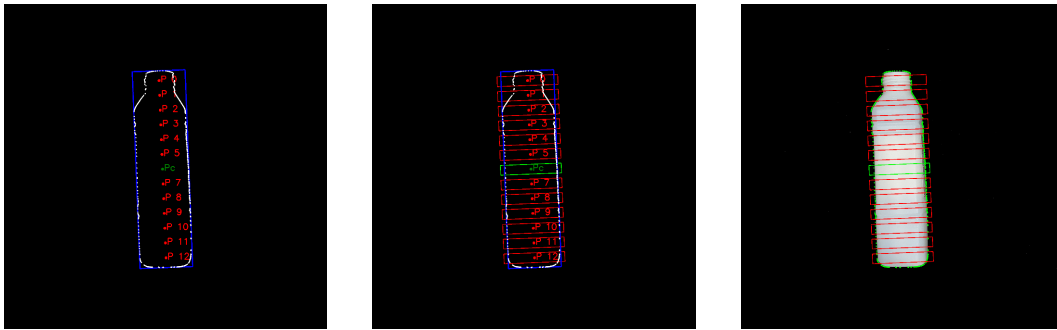
(a) Contorno de la botella (b) Centro del rectángulo delimitador de área mínima P_C (c) Centros de las regiones de agarre P_C

Figura 5-6: Generación de los centros de las regiones de agarre

En la Figura 5-6c, se establecen los centros de las regiones de agarre mediante los conceptos y la heurística detallados al final de la sección 5.3. Estos centros actúan como puntos de referencia para

proyectar las regiones de agarre sobre el objeto, teniendo en cuenta la orientación del rectángulo delimitador y la relación entre su altura y ancho. Este proceso se realiza de manera iterativa sobre cada objeto presente en la escena y permite abordar diversas situaciones de agarre.

En la Figura 5-7, se presentan tres imágenes que ilustran el análisis y cálculo de la generación automática de regiones de agarre. En la Figura 5-7a, similar a la Figura 5-6c, se muestran los centros calculados para la proyección de las regiones de agarre alrededor de un objeto específico. En la Figura 5-7b, se aprecian las regiones de agarre encontradas. Como se destaca, estas se encuentran separadas por una distancia Δ_w para evitar posibles solapamientos, y se extienden más allá del rectángulo delimitador y del contorno del objeto (véase Figura 5-7c).



(a) Contorno, rectángulo delimitador y puntos centrales de las regiones de agarre. (b) Regiones de agarre calculadas. (c) Regiones de agarre en el objeto de interés.

Figura 5-7: Generación automática de regiones de agarre

Estas regiones de agarre establecen una apertura mínima del gripper para facilitar la aproximación al objeto al momento de realizar el agarre. La disposición estratégica de las regiones y su ajuste en relación con la geometría del objeto permiten lograr un agarre efectivo y seguro en diversas situaciones.

5.4. Parámetros de decisión

En el ámbito del diseño de sistemas de agarre robótico, la toma de decisiones en el proceso de agarre juega un papel fundamental para garantizar la eficacia y eficiencia en la manipulación de objetos. En esta sección, se presentan los cuatro parámetros de decisión empleados para evaluar y determinar la pertinencia de cada una de las regiones de agarre generadas para cada uno de los objetos en el espacio estudiado. Se desarrolla un análisis tanto de forma individual (consulte las secciones 5.4.1, 5.4.4) como grupal (consulte las secciones 5.4.2, 5.4.3).

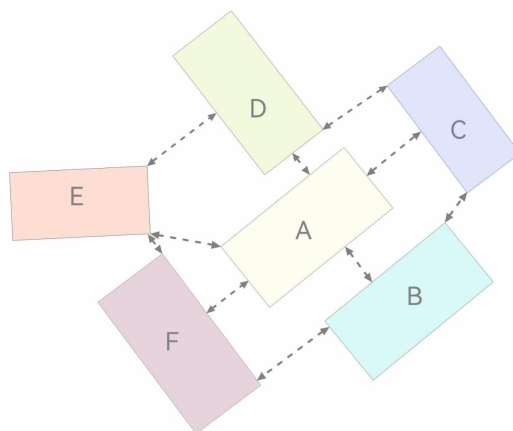


Figura 5-8: Interrelación de los objetos en el espacio.

En el ejemplo ilustrado de la Figura 5-8, se muestran seis (6) objetos identificados como *A*, *B*, *C*, *D*, *E* y *F*. Nuestro análisis se centra en comprender la relación entre estos objetos, con el objetivo principal de evaluar su influencia en las posibles tareas de agarre y manipulación.

A continuación, se realiza un análisis de la disposición espacial de estos objetos y la interferencia potencial entre ellos. Considerando el objeto de interés *A*, la presencia de los cinco (5) objetos restantes en sus proximidades podría afectar las tareas de agarre y manipulación. Por lo tanto, se examinan las regiones específicas donde estos objetos interfieren en el entorno cercano de *A*.

Este análisis nos permitirá comprender los desafíos y consideraciones importantes en el diseño de estrategias de manipulación, destacando zonas críticas y áreas libres de interferencias que pueden influir en el éxito de las operaciones de agarre en un escenario específico. La identificación de estas zonas críticas y la evaluación de la interferencia potencial entre objetos son fundamentales para optimizar la planificación de agarre y garantizar la eficiencia y seguridad en las tareas de manipulación.

5.4.1. Parámetro I: área del objeto

La primera consideración se centra en la dimensión física de los objetos, dando prioridad a los elementos según su área para establecer una base sólida en el diseño de estrategias de agarre. Maximizar la superficie de contacto con el sistema de agarre puede contribuir a una manipulación más efectiva.

Teniendo en cuenta el diseño o la estructura del gripper empleado, el tamaño físico de los objetos se convierte en un parámetro clave. Si el gripper es de grandes proporciones, manipular objetos en áreas reducidas podría representar un desafío. No obstante, el análisis de objetos de mayor tamaño requiere un estudio más estructurado de sus proporciones.

Es esencial considerar que la distancia longitudinal entre los puntos de contacto del gripper con el

objeto no debe exceder las capacidades de apertura del gripper, ya que esto haría que el objeto fuera inaccesible. En este contexto, se lleva a cabo un estudio y clasificación categórica de los objetos de mayor a menor área, eliminando aquellos cuyas dimensiones superan el rango de operación del gripper empleado.

5.4.2. Parámetro II: ubicación espacial del objeto

El segundo parámetro considera la influencia de la disposición espacial de los objetos circundantes al elemento de interés (véase Figura 5-9a). Al examinar la proximidad y el impacto de los objetos en ambos lados mayores del objeto de interés, se busca comprender las interacciones que podrían afectar la selección y la implementación de la estrategia de agarre.

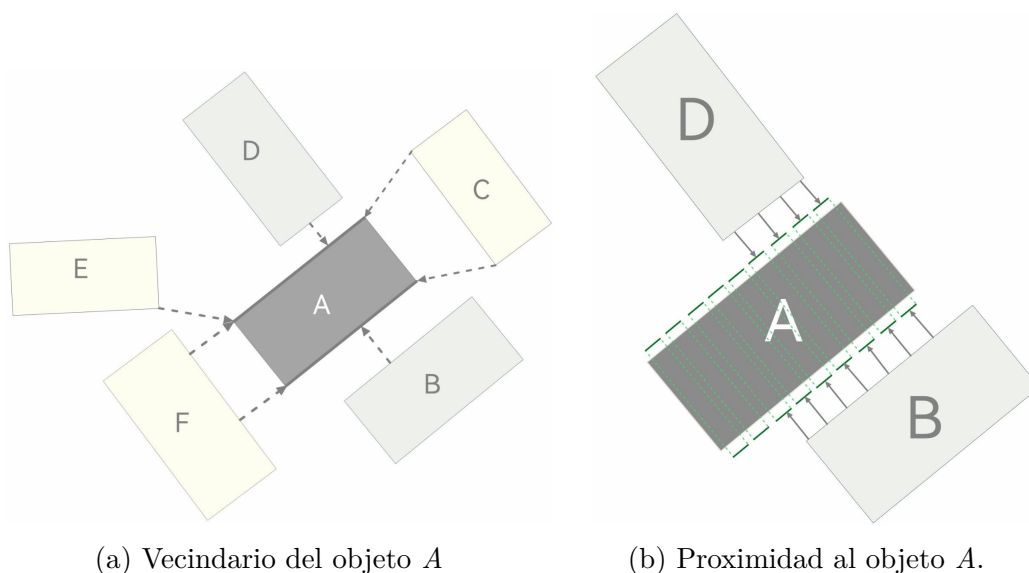


Figura 5-9: Objeto de interés A en el espacio.

El objeto de interés A , ilustrado en la Figura 5-9a, está situado en un entorno que incluye un número limitado de otros objetos (B , C , D , E y F). Dado que la acción de acceso y agarre de los objetos implica envolver el lado de menor dimensión, es decir, el ancho (w_0), los dedos del gripper entrarán en contacto con las áreas más extensas del objeto, que corresponden a la altura (h_o).

Identificar las áreas despejadas en una o todas las zonas adyacentes a cada uno de los lados de mayor magnitud (h_o) del objeto de interés permite encontrar regiones de agarre accesibles y libres de obstáculos. Esto contribuye a reducir eventualmente el posible riesgo de colisión del gripper con los objetos circundantes al objeto de deseado.

Como se muestra en la Figura 5-9a, los objetos B y D interfieren en las cercanías del lado más largo del objeto de interés A . A diferencia de los objetos C , E y F , los objetos B y D (véase Figura 5-9b) podrían representar un desafío para desarrollar las tareas de acceso y agarre del objeto A . Por esta razón, el análisis realizado para determinar una o varias regiones adecuadas para la aproximación,

agarre y manipulación del objeto deseado, se centra en conocer los posibles desafíos que los objetos B y D puedan representar para dicha tarea.

5.4.3. Parámetro III: evaluación espacial de las regiones de agarre generadas

En el tercer parámetro, se realiza una evaluación espacial de las regiones de agarre identificadas. Se examina la posibles interferencia de objetos cercanos en relación con cada lado de proximidad de las regiones de agarre generadas, lo que permite una adaptación a las condiciones del entorno.

Dado que los dedos del gripper son estructuras tridimensionales, el análisis de las regiones que entrarán en contacto con el objeto y el volumen de cada dedo toman relevancia a la hora de elegir una ubicación del gripper adecuada para ejecutar la tarea de agarre, evitando al máximo posibles colisiones.

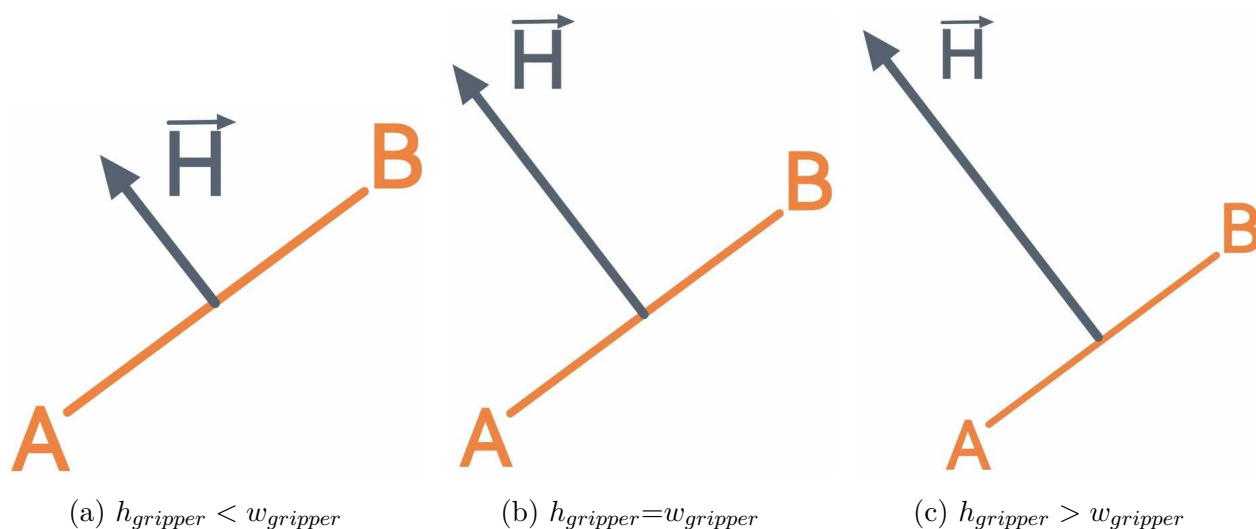
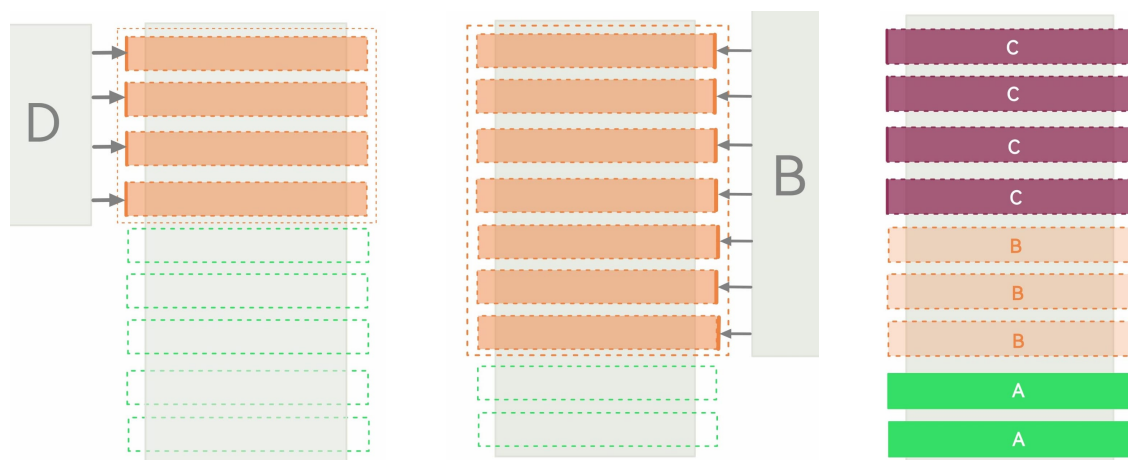


Figura 5-10: Evaluación espacial de las aproximación del gripper.

En la Figura 5-10, se ilustra la relación de acceso de un dedo del gripper con la región de contacto establecida para efectuar la tarea de agarre. En la figura, el segmento de recta \overline{AB} representa el ancho del área de contacto del gripper, y el vector \vec{H} representa la profundidad o grosor del mismo. De esta forma, el grosor del gripper puede ser de menor (véase Figura 5-10a), igual (véase Figura 5-10b) o mayor (véase Figura 5-10c) magnitud al ancho de la región de contacto generada entre el gripper y el objeto de interés. Cabe analizar que la profundidad del dedo del gripper adquiere relevancia a la hora de establecer criterios de accesibilidad al objeto deseado, dado que esta puede ser una de las causales por las cuales un objeto no puede ser accedido, ya que la estructura física del gripper empleado podría colisionar con las estructuras u objetos cercanos.



(a) Interferencia del objeto D en las regiones de agarre del objeto de interés A . (b) Interferencia del objeto B en las regiones de agarre del objeto de interés A . (c) Regiones de agarre en el objeto de interés A .

Figura 5-11: Interferencias sobre las regiones de agarre del objeto de interés A

La proximidad de los objetos B y D con respecto al objeto de interés A muestra un impacto significativo en varias de las regiones de agarre generadas. En el caso del objeto D , se observa una posible interferencia en las primeras cuatro (4) regiones de agarre en uno de los lados más largos del objeto A , específicamente el lado izquierdo visto desde la Figura 5-11a. En cuanto al objeto B , su posible interferencia es aún más notable, afectando siete (7) regiones a lo largo del lado derecho del objeto A , visto desde la Figura 5-11b.

Como resultado, en la Figura 5-11c, se demarcan tres regiones generadas del estudio de las posibles interferencias de objetos cercanos al objeto de interés. Las cuatro regiones superiores observadas en la Figura 5-11c, demarcadas con el carácter **C** y sombreadas de color rojo, representan las regiones de agarre que presentan interferencias en ambos lados. Partiendo de nuestro actual ejemplo, son las regiones interferidas tanto por el objeto B (véase Figura 5-11b) como por el objeto D (véase Figura 5-11a). Estas regiones de agarre generadas se han descartado completamente debido a la gran probabilidad de colisión existente.

Posteriormente, en la Figura 5-11c, se observan cuatro regiones sombreadas de color naranja, etiquetadas con el carácter **B**, las cuales presentan interferencia por algún objeto aledaño a uno de sus lados. En nuestro caso de uso, el objeto B (véase Figura 5-11b).

Finalmente, en la Figura 5-11c, se observan las últimas dos regiones, caracterizadas por la letra **A** y sombreadas de color verde, regiones de agarre libres de obstáculos cercanos.

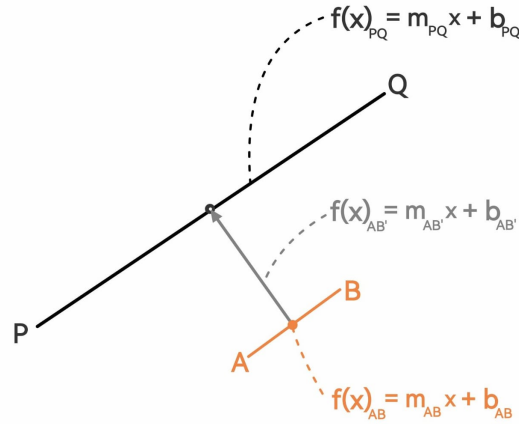


Figura 5-12: Relación de proximidad entre una región de agarre y un objeto

En la Figura 5-12, representa el análisis geométrico que se llevó a cabo en el cálculo de la relación de proximidad entre una región de agarre y un objeto específico. Esta evaluación se basa en la suposición de que el objeto será capturado de forma estática y que el gripper se acercará en dirección del eje Z.

El segmento \overline{AB} en la Figura 5-12, representa el ancho del dedo del gripper empleado, siendo igual al ancho w_0 de las regiones de agarre identificadas. En relación con este segmento de recta \overline{AB} , el centro estará determinado por la siguiente ecuación:

$$\overline{AB}_C = \frac{x_1 - x_0}{2}, \frac{y_1 - y_0}{2} \quad (5-12)$$

Partiendo del punto calculado, se determina la recta perpendicular al segmento \overline{AB} y orientada en dirección opuesta al objeto de interés. Esta recta está representada por la función $f(x)_{\overline{AB}'}$ en la Figura 5-12. El segmento de recta \overline{PQ} , trazado en la misma figura con fines de estudio, constituye uno de los lados del polígono delimitador de un objeto cercano.

A partir de la recta perpendicular $f(x)_{\overline{AB}'}$ con respecto al segmento \overline{AB} , se lleva a cabo el cálculo del punto de intersección entre dicha recta y el segmento \overline{PQ} . En el caso de que exista un punto de intersección, se realiza una evaluación para confirmar su pertenencia al segmento de recta \overline{PQ} . Una vez confirmada dicha pertenencia, se procede a determinar la distancia euclidiana entre el punto encontrado y el punto \overline{AB}_C .

Posteriormente, la distancia calculada se compara con un umbral predefinido, que representa una magnitud significativa del parámetro de profundidad o grosor del dedo del gripper utilizado. Si la distancia resultante es inferior a este umbral, se descarta la región de agarre, ya que, para propósitos prácticos, existe un riesgo inminente de colisión.

Este proceso se lleva a cabo de manera iterativa, examinando las regiones de agarre asociadas con cada objeto en relación con el espacio de los objetos. En caso de que una región de agarre entre en conflicto con algún objeto vecino, se clasifica como una región no accesible. Por el contrario, si

no hay conflictos con objetos vecinos, se clasifica como una región potencialmente accesible. Este enfoque iterativo y de evaluación continua permite ajustar dinámicamente la clasificación de las regiones de agarre en función de las interacciones con los objetos circundantes.

5.4.4. Parámetro IV: evaluación de las regiones de agarre en el objeto

El último parámetro considera la posición espacial de las regiones de agarre dentro del objeto de interés. Al evaluar la proximidad al centro de masa, se destaca la relevancia de las regiones más centrales en el objeto, reconociendo que esta proximidad puede influir significativamente en la calidad y estabilidad del agarre.

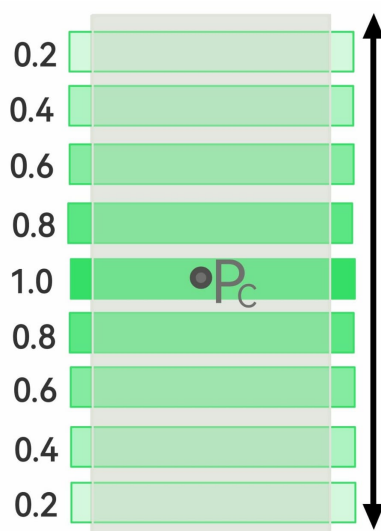


Figura 5-13: Ponderación de las regiones de agarre en el objeto de interés.

En la Figura 5-13, se ilustra, a modo de ejemplo, la distribución de pesos de las regiones de agarre encontradas. A medida que se acerca al centro del objeto, la ponderación dada es mayor. Este enfoque se basa en el análisis presentado, donde, en estructuras cilíndricas como las botellas, la región cercana al centro de masa tiende a ser la de mayor proporción volumétrica del cuerpo. Sin embargo, es importante destacar que esta distribución puede variar en casos específicos.

Utilizando este parámetro y teniendo conocimiento de los centros de las regiones de agarre, así como la distancia entre ellos y el número total de regiones de agarre encontradas por objeto, se establece una relación de valor para cada región de agarre identificada. No obstante, como se aborda en el análisis de la sección 5.4.3, no todas las regiones de agarre estarán siempre disponibles para el sistema de agarre debido a posibles interferencias de objetos cercanos. Por lo tanto, la región de accesibilidad de un objeto específico se determina mediante el análisis heurístico del parámetro tratado en la sección 5.4.3.

6 Validación

6.1. Introducción

En el contexto de la investigación y el desarrollo en robótica, la eficiencia y la reproducibilidad son aspectos cruciales. En la actualidad, la capacidad de configurar y desplegar un entorno de desarrollo de manera rápida y coherente se ha vuelto esencial para acelerar el proceso de experimentación y desarrollo de cualquier aplicación computacional. Es en este escenario, entran en juego dos tecnologías claves: Docker y ROS (Robot Operating System).

Docker es una plataforma de contenedores que permite empaquetar aplicaciones y sus dependencias en un entorno aislado y portátil. Este enfoque, revoluciona la forma en la que los desarrolladores pueden gestionar sus entornos de desarrollo y ejecución. Al emplear contenedores, es posible encapsular y distribuir entornos de trabajo de manera eficiente, garantizando que las mismas configuraciones y dependencias estén disponibles en cualquier lugar donde se ejecuten, independientemente del sistema operativo subyacente.

Por otro lado, ROS, el sistema operativo robótico ampliamente utilizado en la comunidad de robótica, proporciona una serie de herramientas y bibliotecas para el desarrollo de aplicaciones robóticas complejas. Sin embargo, configurar un entorno ROS tradicionalmente ha sido una tarea desafiante y propensa a errores, especialmente al trabajar en diferentes sistemas o proyectos. La combinación de ROS con Docker ofrece una solución elegante para estos desafíos, permitiendo la creación de contenedores ROS preconfigurados que pueden ser compartidos y utilizados de manera consistente.

En el marco de este proyecto, se decidió implementar la versión de ROS Noetic en un contenedor Docker debido a una restricción de vital importancia: ROS Noetic es compatible únicamente con la versión 20.04 de Ubuntu. La incompatibilidad con versiones posteriores de Ubuntu se debe en gran medida a las actualizaciones de las librerías y dependencias del sistema operativo, así como al rápido desarrollo de ROS2, que ha dejado gradualmente sin soporte a las versiones de ROS1 anteriormente utilizadas. Estas versiones son esenciales para la comunicación y configuración con ciertos robots industriales.

La configuración de este entorno se detalla en el archivo Dockerfile alojado en el repositorio de versiones <https://github.com/daochoam/ROS-Docker> en GitHub. Este archivo incluye todas las herramientas y dependencias requeridas para la ejecución y el mantenimiento de un entorno ROS Noetic. Esta configuración garantiza la compatibilidad y estabilidad necesarias para el desarrollo efectivo de aplicaciones robóticas en este entorno particular. Puedes acceder al Dockerfile haciendo

clic en el siguiente enlace: [Dockerfile](#).

6.2. Configuración y ejecución del contenedor docker ROS-Noetic

El Dockerfile propuesto, ejecuta la construcción e implementación de una imagen Docker local destinado a proporcionar un entorno de desarrollo cohesivo y funcional para proyectos basados en ROS Noetic.

El archivo, parte de la ejecución de una imagen base de Ubuntu 20.04 y procede a instalar una variedad de herramientas y utilidades esenciales de desarrollo, así como herramientas de visualización a través de la proyección de entornos visuales empleando los recursos del sistema anfitrión (host) a partir de los permisos proporcionados. Este, contiene los paquetes de Python necesarios para la construcción y ejecución de aplicaciones ROS Noetic y sus dependencias, integrando adicionalmente paquetes como Gazebo 11, MoveIt y RViz, garantizando la disponibilidad de componentes claves para el desarrollo y la visualización dinámica.

Adicionalmente se provee un script shell para la configuración y despliegue del contenedor ROS Noetic. Este permite configurar de forma rápida y dinámica la comunicación entre el contenedor y el sistema anfitrión a través del mapeo de puertos, así mismos establece un canal de mapeo directo entre el contenedor y los periféricos del sistema anfitrión, empleados en el presente proyecto para la comunicación entre ROS Noetic (Docker) - RoboStudio (Windows) y/o ROS Noetic (Docker) - Robot ABB IRB 140. Se establecen los permisos necesarios para la ejecución visual de Gazebo, MoveIt y RViz, así como el entorno de desarrollo Visual Studio Code (VSCode) y se integra una configuración Docker Volumen, como medio de intercambio y persistencia de la información, usualmente empleado en la configuración y edición de los Workspaces ROS.

6.3. Integración del robot ABB IRB 140

En etapas previas, se ha llevado a cabo un análisis geométrico y matemático detallado mediante el uso de algoritmos computacionales. Este enfoque se ha centrado en la extracción de características geométricas de los objetos estudiados, simplificando la tarea a través de formas geométricas primitivas. El desarrollo analítico ha buscado generar un conjunto de regiones aptas para el acceso y agarre mediante sistemas robóticos, considerando las condiciones del entorno de ejecución y la naturaleza de la herramienta de aprehensión.

Hasta el momento, el análisis se ha concentrado en el estudio geométrico para la generación de las regiones de agarre. A partir de estas regiones y tras un proceso heurístico de toma de decisiones, se elige al menos una de las regiones de agarre generadas y evaluadas, considerándola como la opción principal para llevar a cabo un agarre utilizando un manipulador real.

A continuación, se presenta la configuración e integración del robot ABB IRB 140 como manipulador empleado para la ejecución de la tarea de agarre específica.

6.3.1. Configuración del TCP

En el ámbito de la robótica, la sigla TCP (Tool Center Point) corresponde al punto central de la herramienta. Este punto se ubica en el extremo final del robot y actúa como referencia crucial para la orientación del mismo, facilitando la consecución de objetivos definidos en el espacio cartesiano.

En el contexto de este proyecto específico, se llevó a cabo la configuración del TCP mediante la utilización de tres (3) puntos de aproximación. El propósito de esta configuración fue calcular la nueva posición del TCP al integrar el gripper como efector final.

6.3.2. Transformación de coordenadas

6.3.2.1. Coordenadas de traslación de la cámara respecto a la base del robot

La posición de la cámara en relación con la base del robot fue establecida a través de la configuración del TCP (secc. 6.3.1). Este proceso consideró el tipo de gripper utilizado y se ejecutó utilizando el FlexPendant para garantizar una alineación precisa y óptima.

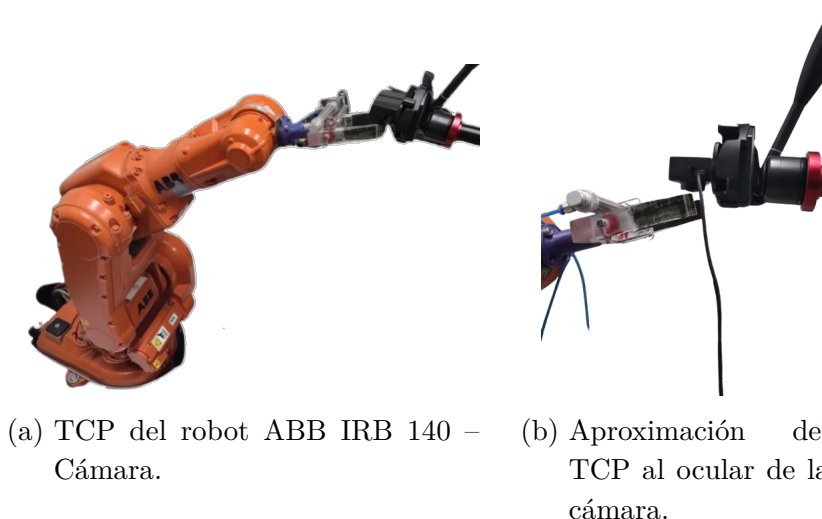


Figura 6-1: Obtención de los parámetros de traslación de la cámara respecto a la base del robot

En este proceso, se aproximó de forma manual, el TCP del conjunto robot ABB IRB 140 - gripper, con el objetivo de obtener las coordenadas de traslación del ocular de la cámara con respecto a la base del robot (véase Figura 6-4b). Este procedimiento se llevó a cabo utilizando las funcionalidades del FlexPendant (véase Figura 6-2a) del robot ABB IRB 140 para garantizar una configuración precisa y ajustada a las necesidades del sistema.



Figura 6-2: Parámetros de translación de la cámara obtenidos en el FlexPendant.

El procedimiento llevado a cabo permitió obtener el conjunto de magnitudes de translación del ocular de la cámara con respecto a la base del robot (véase Figura 6-2b). Los datos obtenidos se detallan en la Tabla 6-1.

Coordenada	Traslación (mm)
x	621.4
y	60.4
z	974.5

Tabla 6-1: Parámetros de translación del ocular de la cámara respecto a la base del robot ABB IRB 140

6.3.3. Matrices de Transformación

Un manipulador, puede ser considerado como una cadena cinemática formada por objetos rígidos (eslabones) unidos o enlazados por articulaciones.

6.3.3.1. Matriz de Traslación

La matriz de translación, representada por la Ec. 6-1, se utiliza para expresar el desplazamiento de un sistema de coordenadas a lo largo de un vector de translación. En el contexto de un sistema servo-visual, esta matriz podría aplicarse para describir el cambio de posición de la cámara con respecto a la base del robot. La forma general de la matriz de translación en el espacio tridimensional es de la forma:

$$\mathbf{T}(t_x, t_y, t_z) = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6-1)$$

Donde, t_x , t_y y t_z , representan las componentes del vector de traslación.

6.3.3.2. Matriz de Rotación

La matriz de rotación, descrita por la ecuación 6-2, representa la orientación de un sistema de coordenadas alrededor de uno o más ejes. En el contexto de un sistema servo-visual, esta matriz se utiliza para expresar los cambios de orientación de la cámara con respecto a la base del robot. La forma general de la matriz de rotación tridimensional es:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6-2)$$

La matriz de rotación, expresada por la ecuación 6-2, representa la orientación de un sistema de coordenadas alrededor de uno o más ejes. En el ámbito de un sistema servo-visual, esta matriz se emplea para describir los cambios de orientación de la cámara con respecto a la base del robot. La forma general de la matriz de rotación tridimensional es:

$$\mathbf{R}_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6-3) \quad \mathbf{R}_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6-4)$$

$$\mathbf{R}_z(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6-5)$$

6.3.3.3. Parámetros Denavit-Hartenberg (DH)

Los parámetros **Denavit-Hartenberg (D-H)** (Ec. 6-6) constituyen un conjunto de cuatro parámetros fundamentales utilizados en la descripción de la cinemática de manipuladores robóticos. Estos parámetros ofrecen una metodología estandarizada para representar las transformaciones geométricas y de orientación entre los eslabones sucesivos de un robot, facilitando así el análisis de la cadena cinemática.

$$\mathbf{DH} = \begin{bmatrix} \cos(\theta) & -\cos(\alpha) \sin(\theta) & \sin(\alpha) \sin(\theta) & a \cos(\theta) \\ \sin(\theta) & \cos(\alpha) \cos(\theta) & -\sin(\alpha) \cos(\theta) & a \sin(\theta) \\ 0 & \sin(\alpha) & \cos(\alpha) & d \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (6-6)$$

Los cuatro parámetros Denavit-Hartenberg son:

- **Longitud del eslabón (a_i) (mm o pulgadas):** Representa la distancia desde la intersección entre el eje z_{i-1} y el eje x_i hacia el origen del sistema i a lo largo del eje x_i (mm o pulgadas).
- **Ángulo de giro del eslabón (α):** Representa la distancia desde la intersección entre el eje z_{i-1} y el eje x_i hacia el origen del sistema i a lo largo del eje x_i .
- **Desplazamiento de la articulación (d_i) (mm o pulgadas):** Representa la distancia a lo largo del eje z_{i-1} desde el origen O_{i-1} hasta la intersección del eje x_i con el eje z_{i-1} .
- **Ángulo de la articulación (θ):** Es el ángulo formado por los ejes x_{i-1} y x_i medido en un plano perpendicular a z_{i-1} .

Implementación:

Inicialmente, se describieron los parámetros Denavit-Hartenberg (DH) del robot utilizado en el laboratorio LABSIR UNAL, el ABB IRB 140. La información detallada se presenta en la Tabla 6-2.

Link	θ_i	α_i	a (mm)	d_i (mm)
A_{01}	$\theta_1 + \frac{\pi}{2}$	$\frac{\pi}{2}$	70	352
A_{12}	$\theta_2 + \frac{\pi}{2}$	0	360	0
A_{23}	θ_3	$\frac{\pi}{2}$	0	0
A_{34}	θ_4	$-\frac{\pi}{2}$	0	380
A_{45}	θ_5	$\frac{\pi}{2}$	0	0
A_{56}	θ_6	0	0	65

Tabla 6-2: Eslabón del robot ABB IRB 140

Aplicando la ecuación Denavit-Hartenberg estudiada en la presente sección y utilizando los parámetros DH obtenidos del robot ABB IRB 140 consignados en la Tabla 6-6, se establece la relación de transformación desde el eslabón 6 hasta la base del robot, como se describe en la ecuación 6-7.

$$\mathbf{ABB}_{06} = \mathbf{ABB}_{01} * \mathbf{ABB}_{12} * \mathbf{ABB}_{23} * \mathbf{ABB}_{34} * \mathbf{ABB}_{45} * \mathbf{ABB}_{56} \quad (6-7)$$

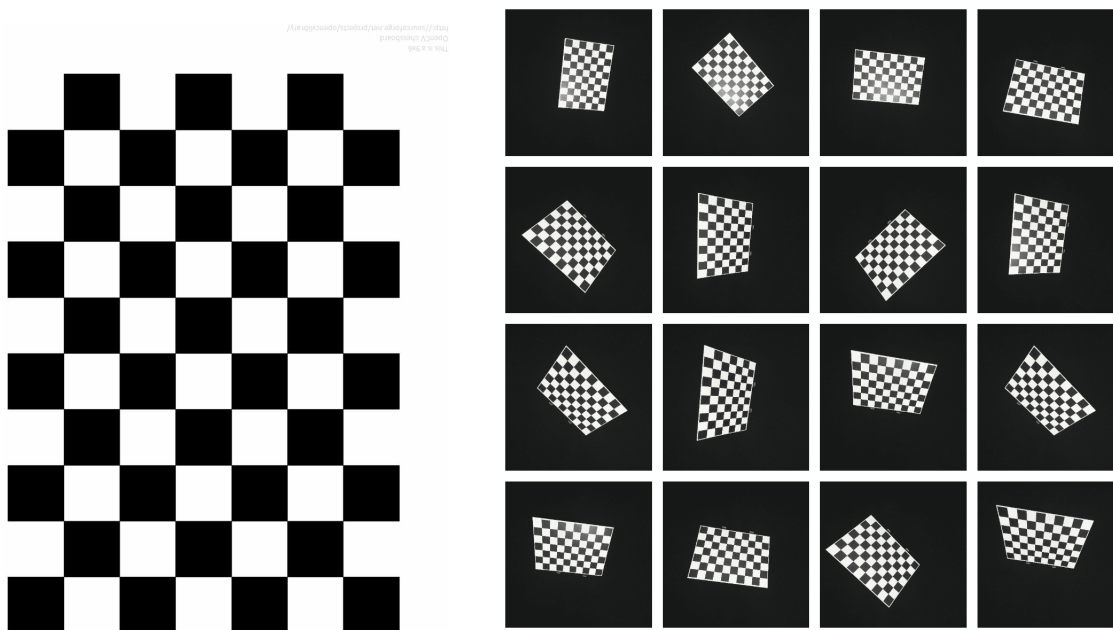
6.3.3.4. Calibración de cámara

La calibración de cámara implica determinar los parámetros intrínsecos, tales como la distancia focal y el centro óptico; y los parámetros extrínsecos de la cámara como la posición y orientación del cuadro de referencia de la cámara con respecto al mundo real. Los intrínsecos describen su funcionamiento interno, mientras que los extrínsecos se relacionan con su posición y orientación.

Algunas cámaras estenopeicas, introducen cierta distorsión significativa en las imágenes captadas que se traducen en una variación en la escala y/o la posición de los puntos en la imagen con respecto a la escena observada. Esta distorsión se constituye por una componente radial y otra tangencial.

- La **distorsión radial** es percibida como la alteración en la linealidad presente en la imagen debido a las variaciones en la refracción de las lentes de naturaleza esférica y a las imperfecciones de su curvatura, produciendo desviaciones laterales del haz de luz que las atraviesa; provocando así una disparidad entre del punto captado y el punto esperado como reflejo del punto real en el espacio observado [34].
- La **distorsión tangencial** es comprendida como la imperfección producida por la falta de alineación paralela entre la lente que capta la imagen y la escena, produciendo errores escalares en algunas áreas de la imagen, haciéndolas parecer más cercanas de lo esperado [34].

El proceso de calibración de cámaras, es imprescindible para la obtención de métricas fiables de la escena a partir de imágenes de la misma. La exactitud de la calibración determinará posteriormente la precisión de las medidas que se realicen a partir de las imágenes. Es por este motivo que es imprescindible realizar la calibración de la cámara con plenas garantías de que los parámetros obtenidos son los más parecidos a los reales.



(a) ChessBoard empleado en el proceso de calibración de cámara (b) Imágenes captadas para el proceso de calibración de cámara

Figura 6-3: Proceso de calibración de cámara

A continuación se presentan los parámetros obtenidos (Ec. 6-8) mediante el proceso de calibración mediante la realización de 33 tomas fotográficas (véase Figura 6-3b) utilizando el tablero de ajedrez mostrado en la Figura 6-3a.

$$\text{Camera_Matrix} = \begin{bmatrix} 1,692 \times 10^3 & 0,000 & 1,810 \times 10^3 \\ 0,000 & 1,692 \times 10^3 & 1,416 \times 10^3 \\ 0,000 & 0,000 & 1,000 \end{bmatrix} \quad (6-8)$$

6.4. Implementación

Con el objetivo de llevar a cabo la verificación final de los análisis realizados, que se desprenden de la caracterización de los objetos presentes en las escenas y concluyen con la generación de las regiones de agarre pertinentes para cada elemento específico, se propuso realizar un proceso de verificación experimental, abarcando tanto pruebas físicas como simuladas. Este proceso incluirá la integración del manipulador industrial ABB IRB 140.

Se configuró un escenario compuesto por un manipulador ABB IRB 140 y una cámara GoPro Hero 9 montada en un trípode a una distancia de $x=62.1 \text{ cm}$, $y=6.1 \text{ cm}$, $z=97.4 \text{ cm}$ de la base del manipulador, según se indica en la Tabla 6-1. Se estableció que, según el marco del manipulador y la pose de la cámara, el marco de referencia se encontraba rotado π y $\pi/2$ con respecto a los ejes X y Z, respectivamente, de la base del robot. Adicionalmente se empleó una caja como marco referencial contenedor de los objetos empleados para la realización de pruebas reales.

La comunicación y planificación de trayectorias del robot se realizaron mediante la implementación del repositorio "LABSIR Bin Picking ROS Simulation packages"¹. Está configurada para establecer una comunicación y planeación de trayectorias con el robot industrial utilizado en el laboratorio LABSIR de la Universidad Nacional de Colombia.

La librería incluye, como marco de ejecución, la biblioteca MoveIt, utilizada para la planificación y ejecución de trayectorias de agarre. Es importante señalar que esta librería ejecuta principalmente rutinas basadas en algoritmos Rapid Random Tree (RRT). Debido a esto, se hace necesario realizar múltiples ejecuciones simuladas para una misma tarea de agarre antes de publicar los pasos de aproximación obtenidos en el robot real. Esto se debe a que el algoritmo RRT incluido arroja trayectorias de movimiento de manera aleatoria sin realizar ninguna optimización sobre las trayectorias calculadas. En algunos casos determinados, esto puede llevar a calcular movimientos que alcancen los límites de operación del robot empleado.

En nuestro marco experimental, la cámara utilizada captura la escena bajo condiciones controladas de iluminación. Utilizando los parámetros obtenidos a través de la calibración de la cámara, se realiza una corrección de la distorsión generada por los parámetros intrínsecos de la misma. Luego, la imagen se filtra con los parámetros propuestos para el filtro de Gabor, como se describe en la sección 3.2.3, con el objetivo de resaltar las diferencias entre los objetos y el fondo de la escena.

¹Respalda y almacenada en el repositorio "https://github.com/GitSRealpe/abb_irb140_ros_sim"

La imagen obtenida en el proceso de filtrado de imágenes de Gabor, es empleada para alimentar el sistema de segmentación de imágenes multiclase preentrenado *Image Agnostic Segmentation*, expuesto en la sección 4.3. Este sistema produce como resultado el conjunto de puntos correspondientes a los contornos de cada uno de los objetos presentes en la escena.

Los contornos obtenidos para cada objeto se someten a evaluación geométrica y matemática mediante los algoritmos detallados en el sistema propuesto de generación automática de regiones de agarre, como se describe en el capítulo 5. Como resultado preliminar, se obtienen las regiones de agarre para cada objeto de manera individual.

Posteriormente, mediante un análisis heurístico, se establece un árbol de decisiones basado en cálculos y condiciones matemáticas de las regiones de agarre de cada objeto en relación con los objetos circundantes. El objetivo es parametrizar la toma de decisiones sobre la accesibilidad de los objetos.

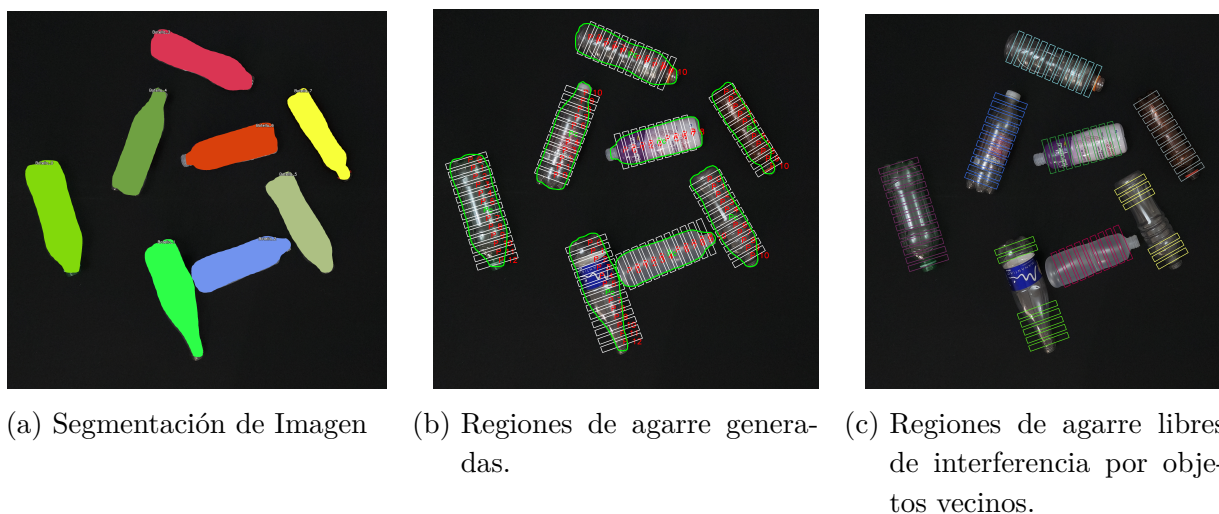


Figura 6-4: Proceso de segmentación, generación y evaluación de las regiones de agarre generadas.

El tiempo de procesamiento y cálculo de las regiones de agarre es elevado, lo que representa una limitación significativa del sistema actual. Este alto consumo de tiempo se debe a la complejidad de los análisis visuales y la necesidad de evaluar múltiples características específicas de los objetos, lo cual es especialmente problemático en situaciones con solapamientos. La eficiencia del sistema se ve comprometida, afectando su aplicabilidad en escenarios donde se requiere una respuesta rápida y/o en tiempo real. En la Figura 6-5 se puede observar el tiempo de procesamiento de las regiones de agarre en función del número de botellas en la escena.

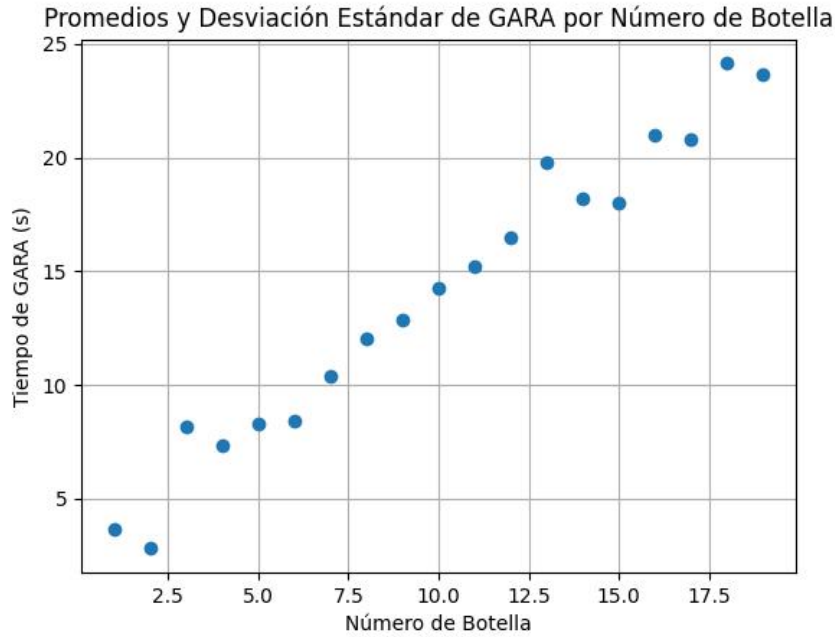


Figura 6-5: Tiempos promedios en etapa de Generación Automática de Regiones de Agarre vs número de botellas en la escena-

En la salida del sistema, se obtiene una región de agarre para cada objeto en un instante determinado, considerando su ubicación espacial, proximidad a objetos vecinos y el área de contacto calculada a la cual el gripper tendrá acceso. El sistema propuesto integra datos relevantes para la ejecución de la tarea de agarre. Estos datos incluyen el punto central o centro de masa del objeto de interés, empleado como punto de aproximación vertical del Punto Central de la Herramienta (TCP).

Adicionalmente, se calcula el ángulo de inclinación bidimensional, el cual se utiliza para sintetizar el ángulo de orientación del efector final. El ancho del objeto es empleado para determinar la longitud mínima de apertura de los dedos del gripper. Asimismo, se considera el ancho del objeto para obtener un valor aproximado la altura del objeto captado en la escena, teniendo en cuenta la disposición estudiada de las botellas en el escenario propuesto.

A partir de los parámetros Denavit-Hartenberg del manipulador ABB empleado, los parámetros de calibración de cámara, y la relación de traslación y rotación estudiadas del marco de referencia del robot y la cámara, se aplica la estructura de la ecuación ???. Esta ecuación permite aproximar el marco de referencia de la imagen captada al marco de referencia del espacio del robot, facilitando así la transformación de los ejes coordenados de los objetos en la imagen al eje coordenado de operación del robot empleado.

$$\text{Camera_ABB} = \text{IRB140}_{\text{DH}} \times \mathbf{R}_z(\theta) \times \mathbf{R}_x(\theta) \times \mathbf{T}(x, y, z) \times \text{Camera_Matrix} \quad (6-9)$$

Partiendo de la ecuación 6-9 y considerando un punto específico en el espacio de la imagen con parámetros de orientación determinados, se llevaron a cabo las simulaciones correspondientes mediante ROS-MoveIt. En este proceso, se posicionó el robot ABB en una pose específica dentro del simulador, se procedió a la publicación de las coordenadas calculadas en el espacio del robot y se ejecutaron las simulaciones utilizando algoritmos RRT del simulador. El propósito era obtener la trayectoria de ejecución óptima, como se ilustra en la Figura 6-6.

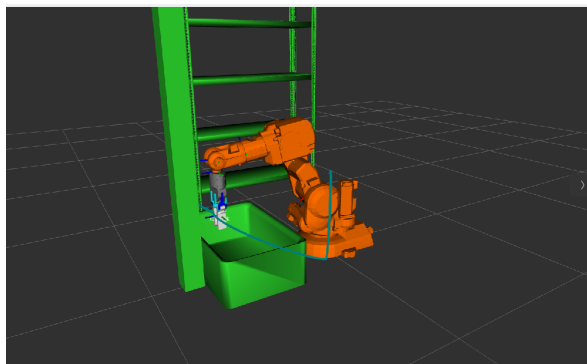


Figura 6-6: Simulación del sistema en el entorno ROS-MoveIt-RViz.

En la Figura 6-6, se presenta la simulación llevada a cabo en el entorno ROS-MoveIt. Se capturó una imagen del objeto en la escena real y se analizaron todas las etapas del sistema propuesto para obtener las coordenadas x e y . El parámetro de profundidad z se mantuvo constante para efectos de la presente simulación. La evaluación de la trayectoria del manipulador desde una ubicación específica en este entorno proporciona una representación visual y contribuye significativamente a la verificación y validación del rendimiento del sistema bajo condiciones simuladas.

7 Conclusiones

Con respecto a la **generación del banco de datos** (Capítulo 2), se implementaron dos escenarios para la captura y generación de una base de datos de imágenes digitales bidimensionales, utilizando materiales comúnmente empleados en escenarios de fotografía. Los dos escenarios fueron contrastados entre sí, exponiendo las ventajas y falencias percibidas en el conjunto de elementos utilizados.

El primer escenario fotográfico presentó problemas de reflexión lumínica en el fondo y dificultades en la focalización de las fuentes de luz, lo que aumentó el ruido en las imágenes. En respuesta, se diseñó un segundo escenario que mostró una mejora significativa: al emplear un material tenue y no reflectivo y utilizar sombrillas de luz para lograr una distribución uniforme, se mitigaron los efectos negativos de las fuentes lumínicas, mejorando así la calidad de las imágenes capturadas.

Finalmente, se realizó la captura del material necesario mediante dos metodologías: una para la captura individual de elementos y otra para la captura conjunta. La primera permitió obtener aproximadamente 2500 imágenes individuales, con varias rotaciones y tomas fotográficas con y sin etiqueta, facilitando la caracterización detallada de cada objeto. La segunda, con 600 imágenes de botellas en conjunto, simuló un escenario real de reciclaje. Esta amplia base de datos visual sirvió de soporte para el análisis de las características de las botellas en diversas condiciones.

En relación a los **sistemas de percepción (visión por computador)** (Capítulo 3), se abordaron diversas estrategias para el procesamiento de imágenes, destacando la importancia de ajustar parámetros clave en filtros de Gabor, como el tamaño del kernel gaussiano (γ), la orientación (θ) y la longitud de onda (λ). Además, se exploraron modelos de color, como HSV, para obtener una representación más completa de las imágenes.

La implementación de histogramas en los canales H, S y V proporcionó una valiosa visión de la distribución de intensidades y colores, facilitando la identificación de características específicas. La aplicación del método de Otsu para la umbralización demostró ser eficaz, adaptando umbrales óptimos a la variabilidad de intensidades en el canal de brillo (V).

El uso de máscaras binarias basadas en rangos específicos de brillo permitió resaltar áreas de interés, demostrando ser una estrategia efectiva para la segmentación. Sin embargo, se identificaron desafíos en la segmentación de objetos oscuros, donde la similitud de intensidades con el fondo resultó en la pérdida de contornos.

Como resultado, durante el presente estudio, se desarrolló un banco de 192 filtros de Gabor, los

cuales demostraron ser efectivos en el realce de tonalidades y texturas de las diversas botellas analizadas en la base de datos generada. Para la segmentación de los objetos, se estableció que el valor mínimo del histograma de brillo fuera de 60 bins, mientras que el valor máximo fue determinado mediante el método de umbralización de Otsu. Este enfoque permitió la creación de máscaras eficientes para la segmentación de los objetos.

Por otra parte, respecto al **sistema automático de reconocimiento de imágenes** (Capítulo 4), durante la fase de desarrollo de la etapa de segmentación de imágenes, se identificaron desafíos significativos en el proceso, incluyendo la pérdida de información en los contornos de los objetos, distorsión debido a la proximidad espacial y dificultades para adaptarse a variaciones lumínicas.

La evaluación del sistema en una muestra de 180 fotografías, que comprenden un total de 1532 imágenes de botellas plásticas, reveló un alto porcentaje de éxito del 96.54% con un umbral de puntuación (*score thresh*) de 0.72. Sin embargo, ajustar este umbral impactó en la efectividad, descendiendo al 88.25% al aumentar el umbral a 0.9, especialmente en objetos con contacto en sus superficies.

El contraste entre el conteo manual y el conteo optimizado destacó diferencias y similitudes, subrayando la importancia de la validación manual como referencia. Este análisis comparativo ofrece una evaluación completa del rendimiento del sistema, proporcionando información valiosa para futuras mejoras y ajustes en la detección de objetos en entornos específicos.

En lo relativo a la **generación automática de regiones de agarre (GARA)** (Capítulo 5), se centró en la generación automática de regiones de agarre para optimizar el proceso de manipulación de objetos mediante un sistema robótico. Se destacaron cuatro parámetros de decisión fundamentales para evaluar la pertinencia de cada región de agarre generada en el espacio estudiado.

El análisis de estos parámetros se lleva a cabo de manera iterativa y continua, permitiendo ajustar dinámicamente la clasificación de las regiones de agarre en función de las interacciones con los objetos circundantes. La metodología propuesta contribuye en la búsqueda de regiones de accesibilidad de los objetos inmersos en un espacio determinado.

Aunque los resultados actuales destacan la eficacia del sistema en la obtención de regiones de agarre, es crucial reconocer áreas de mejora. La dependencia de análisis visuales y características específicas de los objetos limita la capacidad de seleccionar candidatos óptimos de agarre, especialmente al enfrentarse a solapamientos. La respuesta del sistema, se ve afectada por los criterios de proximidad, la toma de decisiones, la cantidad de objetos en la escena y las variaciones del entorno.

Para mejorar la eficiencia, se sugiere la implementación de una red de aprendizaje automático. Esto permitiría al sistema adaptarse y aprender patrones más complejos, mejorando la capacidad de selección y evaluación de regiones de agarre de manera rápida y eficiente. Sin embargo, al adoptar el aprendizaje automático, es esencial garantizar la persistencia de la respuesta principal de

la propuesta, supervisando y controlando la red para mantener decisiones coherentes y seguras en diversos contextos.

En cuanto a **integración y validación** (Capítulo 6), el sistema desarrollado genera regiones de agarre de manera automática para objetos cilíndricos en un entorno semiestructurado utilizando visión por computadora. La combinación de técnicas de procesamiento de imágenes, segmentación y planificación de agarre ha permitido obtener resultados prometedores en la generación de regiones de agarre en condiciones controladas.

La implementación de la comunicación efectiva con el robot industrial ABB IRB 140, junto con la planificación de trayectorias en el entorno ROS-MoveIt, ha contribuido a la validez y aplicabilidad del sistema en un contexto de manipulación robótica. La realización de experimentos, tanto en entornos reales como simulados, desempeña un papel crucial para demostrar la eficiencia y robustez del sistema propuesto.

A pesar de los avances en la obtención de regiones de agarre, persisten limitaciones en la selección de candidatos debido a la dependencia de análisis visuales y características específicas, especialmente en situaciones de solapamiento. Para futuros trabajos, se sugiere diversificar los objetivos del sistema para abarcar una mayor variedad de formas y tamaños de objetos, extender el algoritmo al espacio tridimensional y explorar el uso del método GARA para entrenar redes neuronales en la identificación de regiones de agarre, aprovechando técnicas de aprendizaje profundo para mejorar la capacidad del sistema en diversos escenarios y con distintos objetos.

Bibliografía

- [1] ADARSH, Pranav ; RATHI, Pratibha ; KUMAR, Manoj: YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. En: *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)* (2020), 3, p. 687–694. ISBN 9781728151977
- [2] AMEND, John R. ; BROWN, Eric ; RODENBERG, Nicholas ; JAEGER, Heinrich M. ; LIPSON, Hod: A positive pressure universal gripper based on the jamming of granular material. En: *IEEE Transactions on Robotics* 28 (2012), 4, Nr. 2, p. 341–350. – ISSN 15523098
- [3] BAREQUET, Gill ; WOLFERS, Barbara: Optimizing a strip separating two polygons. En: *Graphical Models and Image Processing* 60 (1998), Nr. 3, p. 214–221. – ISSN 10773169
- [4] BHATTACHARYA, Binay ; TOUSSAINT, Godfried: Computing shortest transversals. En: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 510 LNCS (1991), p. 649–660. – ISBN 9783540542339
- [5] BHATTACHARYA, Binay K. ; TOUSSAINT, Godfried T.: Efficient algorithms for computing the maximum distance between two finite planar sets. En: *Journal of Algorithms* 4 (1983), 6, Nr. 2, p. 121–136. – ISSN 0196–6774
- [6] BOHG, J ; MORALES, A ; ASFOUR, T: Data-driven grasp synthesis—a survey. En: *IEEE Transactions on robotics* 30 (2014), p. 289–309
- [7] CEBOLLADA, S ; PAYÁ, L ; FLORES, M ; . . . , A Peidro Expert Systems w. ; 2021, undefined: A state-of-the-art review on mobile robotics tasks using artificial intelligence and visual data. En: *Expert Systems with Applications* 167
- [8] DAI, Jifeng ; HE, Kaiming ; SUN, Jian: Convolutional feature masking for joint object and stuff segmentation. En: *Computer Vision and Pattern Recognition* 07-12-June-2015 (2014), 10, p. 3992–4000. – ISBN 9781467369640
- [9] DEIMEL, Raphael ; BROCK, Oliver: A novel type of compliant and underactuated robotic hand for dexterous grasping. En: <https://doi.org/10.1177/0278364915592961> 35 (2015), 8, Nr. 1-3, p. 161–185. – ISSN 17413176
- [10] DENG, Z ; ZHENG, X ; ZHANG, L ; ZHANG, J: A learning framework for semantic reach-to-grasp tasks integrating machine learning and optimization. En: *Robotics and Autonomous Systems* 108 (2018), p. 140–152

-
- [11] DETRY, Renaud ; EK, Carl H. ; MADRY, Marianna ; PIATER, Justus ; KRAGIC, Danica: Generalizing grasps across partly similar objects. En: *2012 IEEE International Conference on Robotics and Automation*, IEEE, 5 2012. – ISBN 978–1–4673–1405–3, p. 3791–3797
- [12] DOMAE, Yukiyasu ; OKUDA, Haruhisa ; TAGUCHI, Yuichi ; SUMI, Kazuhiko ; HIRAI, Takashi: Fast graspability evaluation on single depth maps for bin picking with general grippers. En: *2014 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 5 2014. – ISBN 978–1–4799–3685–4, p. 1997–2004
- [13] EBERLY, David: Minimum-Area Rectangle Containing a Set of Points. En: *Geometric Tools, LLC* (2015)
- [14] EIZICOVITS, D ; BERMAN, S: Automatic graspability map generation based on shape-primitives for unknown and familiar objects. En: *Robotics and Autonomous Systems* (2018)
- [15] ERHAN, Dumitru ; SZEGEDY, Christian ; TOSHEV, Alexander ; ANGUELOV, Dragomir: Scalable Object Detection using Deep Neural Networks. En: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2013), 12, p. 2155–2162. – ISBN 9781479951178
- [16] FARABET, Clement ; COUPRIE, Camille ; NAJMAN, Laurent ; LECUN, Yann: Learning hierarchical features for scene labeling. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (2013), Nr. 8, p. 1915–1929. – ISSN 01628828
- [17] FOGEL, I ; SAGI, D: Biological Cybernetics Gabor Filters as Texture Discriminator. En: *Biol. Cybern* 61 (1989), p. 113
- [18] GIRSHICK, Ross ; DONAHUE, Jeff ; DARRELL, Trevor ; MALIK, Jitendra: Region-based Convolutional Networks for Accurate Object Detection and Segmentation. En: *IEEE transactions on pattern analysis and machine intelligence* 38 (2015), p. 142–158
- [19] GOLDFEDER, C ; CIOCARLIE, M ; DANG, H ; ALLEN, PK: The columbia grasp database. En: *IEEE international conference on robotics and automation* (2009), p. 1710–1716
- [20] GOUDA, Anas ; GHANEM, Abraham ; KAISER, Pascal ; HOMPEL, Michael T.: Object class-agnostic segmentation for practical CNN utilization in industry. En: *2021 6th International Conference on Mechanical Engineering and Robotics Research, ICMERR 2021* (2021), p. 97–105. ISBN 9781665406420
- [21] GOUDA, Anas ; ROIDL, Moritz: DoUnseen: Tuning-Free Class-Adaptive Object Detection of Unseen Objects for Robotic Grasping. En: *arXiv preprint arXiv:2404.06277* (2023)
- [22] HOULE, Michael E. ; TOUSSAINT, Godfried T.: Computing the Width of a Set. En: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10 (1988), Nr. 5, p. 761–765. – ISSN 01628828
- [23] HU, Xiao d. ; WANG, Xin q. ; MENG, Fan j. ; HUA, Xia ; YAN, Yu j. ; LI, Yu y. ; HUANG, Jing ; JIANG, Xun l.: Gabor-CNN for object detection based on small samples. En: *Defence Technology* 16 (2020), 12, Nr. 6, p. 1116–1129. – ISSN 22149147

- [24] HU, Zhongxu ; TAN, Runjia ; ZHOU, Yanxin ; WOON, Junyang ; LV, Chen: Template-Based Category-Agnostic Instance Detection for Robotic Manipulation. En: *IEEE Robotics and Automation Letters* 7 (2022), 10, Nr. 4, p. 12451–12458. – ISSN 23773766
- [25] HUGHES, Josie ; CULHA, Utku ; GIARDINA, Fabio ; GUENTHER, Fabian ; ROSENDO, Andre ; IIDA, Fumiya: Soft manipulators and grippers: A review. En: *Frontiers Robotics AI* 3 (2016), 11, Nr. NOV, p. 223168. – ISSN 22969144
- [26] ITO, Seiji ; YOSHIOKA, Michifumi ; OMATU, Sigeru ; KITA, Kouji ; KUGO, Kouichi: An image segmentation method using histograms and the human characteristics of HSI color space for a scene image. En: *Artificial Life and Robotics 2006 10:1* 10 (2006), 7, Nr. 1, p. 6–10. – ISSN 1614–7456
- [27] KASPER, Alexander ; XUE, Zhixing ; DILLMANN, Rüdiger: The KIT object models database: An object model database for object recognition, localization and manipulation in service robotics. En: *The International Journal of Robotics Research* 31 (2012), 7, Nr. 8, p. 927–934. – ISSN 0278–3649
- [28] KUMRA, Sulabh ; KANAN, Christopher: Robotic grasp detection using deep convolutional neural networks. En: *IEEE International Conference on Intelligent Robots and Systems 2017* (2017), 12, p. 769 – 776. – ISBN 9781538626825
- [29] LENZ, Ian ; LEE, Honglak ; SAXENA, Ashutosh: Deep learning for detecting robotic grasps. En: *The International Journal of Robotics Research* 34 (2015), 4, Nr. 4-5, p. 705–724. – ISSN 0278–3649
- [30] LI, Miao ; HANG, Kaiyu ; KRAGIC, Danica ; BILLARD, Aude: Dexterous grasping under shape uncertainty. En: *Robotics and Autonomous Systems* 75 (2016), 1, p. 352–364. – ISSN 0921–8890
- [31] LIAO, Ping-Sung ; CHEN, Tse-Sheng ; CHUNG, Pau-Choo: A Fast Algorithm for Multilevel Thresholding. En: *JOURNAL OF INFORMATION SCIENCE AND ENGINEERING* 17 (2001), p. 713–727
- [32] LIU, Chunfang ; SUN, Fuchun ; BAN, Xiaojuan: An effective method for grasp planning on objects with complex geometry combining human experience and analytical approach. En: *Science China Information Sciences* 59 (2016), 11, Nr. 11, p. 112212. – ISSN 1674–733X
- [33] LOZANO-PÉREZ, Tomás: Spatial Planning: A Configuration Space Approach. En: *IEEE Transactions on Computers* C-32 (1983), Nr. 2, p. 108–120. – ISSN 00189340
- [34] MCGLONE, J. C. ; MIKHAIL, Edward M. ; BETHEL, James S. *Manual of Photogrammetry, 5th Edition - Dialnet.* 1 2004
- [35] MEBATSION, H.K. ; PALIWAL, J.: Machine vision based automatic separation of touching convex shaped objects. En: *Computers in Industry* 63 (2012), 9, Nr. 7, p. 723–730. – ISSN 0166–3615

- [36] MERCIER, Jean P. ; GARON, Mathieu ; GIGUERE, Philippe ; LALONDE, Jean F.: Deep Template-based Object Instance Detection. En: *Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021* (2019), 11, p. 1506–1515. ISBN 9780738142661
- [37] MIZUSHIMA, Kaori ; OKU, Takumi ; SUZUKI, Yosuke ; TSUJI, Tokuo ; WATANABE, Tetsu-you: Multi-fingered robotic hand based on hybrid mechanism of tendon-driven and jamming transition. En: *2018 IEEE International Conference on Soft Robotics, RoboSoft 2018* (2018), 7, p. 376–381. ISBN 9781538645161
- [38] OTSU, Nobuyuki: THRESHOLD SELECTION METHOD FROM GRAY-LEVEL HISTOGRAMS. En: *IEEE Trans Syst Man Cybern SMC-9* (1979), Nr. 1, p. 62–66. – ISSN 00189472
- [39] PARK, Kiru ; PATTEN, Timothy ; PRANKL, Johann ; VINCZE, Markus: Multi-task template matching for object detection, segmentation and pose estimation using depth images. En: *Proceedings - IEEE International Conference on Robotics and Automation 2019-May* (2019), 5, p. 7207–7213. – ISBN 9781538660263
- [40] PATIL, Rajesh V. ; REDDY, Y. P.: An Autonomous Technique for Multi Class Weld Imperfections Detection and Classification by Support Vector Machine. En: *Journal of Nondestructive Evaluation* 40 (2021), 9, Nr. 3. – ISSN 15734862
- [41] PFEIFER, Rolf ; LUNGARELLA, Max ; IIDA, Fumiya: The challenges ahead for bio-inspired 'soft' robotics. En: *Communications of the ACM* 55 (2012), 11, Nr. 11, p. 76–87. – ISSN 00010782
- [42] PIRZADEH, Hormoz. ; TOUSSAINT, Godfried T.: *Computational geometry with the rotating calipers*, McGill University, Tesis de Grado, 1999
- [43] PRATTICCHIZZO, Domenico ; MALVEZZI, Monica ; GABICCINI, Marco ; BICCHI, Antonio: On the manipulability ellipsoids of underactuated robotic hands with compliance. En: *Robotics and Autonomous Systems* 60 (2012), 3, Nr. 3, p. 337–346. – ISSN 0921–8890
- [44] RAMLI, S ; MUSTAFA, M M. ; WAHAB, D A. ; HUSSAIN, A: Plastic bottle shape classification using partial erosion-based approach. En: *2010 6th International Colloquium on Signal Processing & its Applications, IEEE*, 5 2010. – ISBN 978–1–4244–7121–8, p. 1–4
- [45] REDMON, Joseph ; ANGELOVA, Anelia: Real-time grasp detection using convolutional neural networks. En: *2015 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 5 2015. – ISBN 978–1–4799–6923–4, p. 1316–1322
- [46] REDMON, Joseph ; FARHADI, Ali: YOLO9000: Better, Faster, Stronger. En: *Computer Vision and Pattern Recognition 2017-January* (2016), 11, p. 6517–6525. ISBN 9781538604571
- [47] RICH, Steven I. ; WOOD, Robert J. ; MAJIDI, Carmel: Untethered soft robotics. En: *Nature Electronics 2018 1:2* 1 (2018), 2, Nr. 2, p. 102–112. – ISSN 2520–1131

- [48] ROA, Máximo A. ; SUÁREZ, Raúl: Grasp quality measures: review and performance. En: *Autonomous Robots* 38 (2015), 1, Nr. 1, p. 65–88. – ISSN 0929–5593
- [49] ROBERT, Jean M. ; TOUSSAINT, Godfried T.: Linear approximation of simple objects. En: *Computational Geometry: Theory and Applications* 4 (1994), Nr. 1, p. 27–52. – ISSN 0925–7721
- [50] SAHBANI, A ; EL-KHOURY, S ; BIDAUD, P: An overview of 3D object grasp synthesis algorithms. En: *Robotics and Autonomous Systems* 60 (2012)
- [51] SANKUR, Buğlent: Survey over image thresholding techniques and quantitative performance evaluation. En: *Journal of Electronic Imaging* 13 (2004), 1, Nr. 1, p. 146. – ISSN 1017–9909
- [52] SAXENA, A ; DRIEMEYER, J ; NG, AY: Robotic grasping of novel objects using vision. En: *The International Journal of Robotics Research* 2 (2008), p. 157–163
- [53] SCAVINO, Edgar ; WAHAB, Dzuraidah A. ; HUSSAIN, Aini ; BASRI, Hassan ; MUSTAFA, Mohd M.: Application of automated image analysis to the identification and extraction of recyclable plastic bottles. En: *Journal of Zhejiang University-SCIENCE A* 10 (2009), 6, Nr. 6, p. 794–799. – ISSN 1673–565X
- [54] SUI, Dongbao ; ZHU, Yanhe ; ZHAO, Sikai ; WANG, Tianshuo ; AGRAWAL, Sunil K. ; ZHANG, He ; ZHAO, Jie: A Bioinspired Soft Swallowing Gripper for Universal Adaptable Grasping. En: *Soft Robotics* 9 (2022), 2, Nr. 1, p. 36–56. – ISSN 21695180
- [55] TOCHON, G: Compression of color images.
- [56] TOUSSAINT, Godfried T. ; TOUSSAINT, Godfried T.: A Simple Linear Algorithm for Intersecting Convex Polygons. En: *THE VISUAL COMPUTER* 1 (1985), p. 118–123
- [57] VAHRENKAMP, N ; WESTKAMP, L: Part-based grasp planning for familiar objects. En: *2016 IEEE-RAS 16th international conference on humanoid robots (humanoids)* (2016), p. 919–925
- [58] WANG, Yong ; LIU, Zhi ; ZHANG, Yun: Image based adaptive coordinated control for cooperative manipulators. En: *2016 Chinese Control and Decision Conference (CCDC)*, IEEE, 5 2016. – ISBN 978–1–4673–9714–8, p. 5383–5388
- [59] WANG, Zhaokun ; PENG, Binbin ; HUANG, Yanjun ; SUN, Guanqun: Classification for plastic bottles recycling based on image recognition. En: *Waste Management* 88 (2019), 4, p. 170–181. – ISSN 0956–053X
- [60] ZHAO, Gang ; ZOU, Shuilong ; WU, Huijie: Improved Algorithm for Face Mask Detection Based on YOLO-v4. En: *International Journal of Computational Intelligence Systems* 16 (2023), Nr. 3, p. 104. ISBN 0123456789
- [61] ZHU, Jie ; YU, Jian ; WANG, Chaomurilige ; LI, Fan-Zhang: Object recognition via contextual color attention. En: *Journal of Visual Communication and Image Representation* 27 (2015), 2, p. 44–56. – ISSN 10473203