PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

DETECCIÓN DE OBJETOS EN IMÁGENES DE RAYOS X USANDO VISIÓN ACTIVA

VLADIMIR ALEJANDRO RIFFO BOUFFANAIS

Tesis para optar al grado de
Doctor en Ciencias de la Ingeniería

Profesor Supervisor:
DOMINGO MERY QUIROZ

Santiago de Chile, Junio, 2016

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

DETECCIÓN DE OBJETOS EN IMÁGENES DE RAYOS X USANDO VISIÓN ACTIVA

VLADIMIR ALEJANDRO RIFFO BOUFFANAIS

Tesis presentada a la Comisión integrada por los profesores:

DOMINGO MERY QUIROZ
ÁLVARO SOTO ARRIAZA
CRISTIÁN TEJOS NÚÑEZ
ROBERTO QUEVEDO LEÓN
LUIS PIZARRO QUIROZ
CRISTIÁN VIAL EDWARDS

Para completar las exigencias del grado de
Doctor en Ciencias de la Ingeniería

Santiago de Chile, Junio, 2016

Esta tesis se la dedico a mis amadas hijas Stephanie y Catherine, quienes han sido siempre fuente de inspiración y que aún en mi prolongada ausencia, con sus tiernas sonrisas, me dieron la fuerza para seguir adelante.

AGRADECIMIENTOS

Durante el proceso de elaboración de esta tesis, surgieron inconvenientes de todo tipo, lo cual hizo que no siempre fuese fácil avanzar. En esos momento de tribulación cuando no encontraba el camino, aparecieron voces amigas con consejos y comentarios que iban desde lo técnico a lo humano. También hubo muchos momentos de alegrías y crecimiento personal e intelectual. Sin duda todos los momentos experimentados en este trabajo, llevan consigo el desarrollo de amistades que han perdurado en el tiempo.

Agradezco a mi profesor supervisor, Domingo Mery, por haberme instruido en los conceptos teóricos y prácticos de esta tesis, pero principalmente por haberme escuchado y aconsejado en temas de vida. No quiero dejar pasar la oportunidad de agradecer a la comisión evaluadora, por los elementos formativos y críticas constructivas que han realizado durante todo el proceso de elaboración de esta tesis.

En algunas asignaturas y a lo largo del doctorado aparecieron buenos amigos que creyeron en mi y me dieron su apoyo e ideas en muchas circunstancias, en particular agradezco a Christian, Billy y Theo, sin ellos habría sido todo más difícil.

Quiero agradecer a mi madre Elsa, por su apoyo y amor incondicional, por darme ánimo en los días en que todo se tornaba complejo y por su ejemplo de vida, que me muestra el valor del esfuerzo.

Sin duda hay familiares, amigos y conocidos que sería difícil detallar en este instante y que han contribuido de alguna u otra forma al buen término de esta tesis doctoral, con su apoyo, consejos y críticas, ayudaron a que el camino fuese menos pedregoso.

Finalmente, agradezco a CONICYT (Comisión Nacional de Investigación Científica y Tecnológica), por haber apoyado mis estudios con la "Beca Para Estudios de Doctorado en Chile".

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

DETECCIÓN DE OBJETOS EN IMÁGENES DE RAYOS X USANDO VISIÓN ACTIVA

Tesis presentada a la Dirección de Investigación y Postgrado como parte de los requisitos para optar al grado de Doctor en Ciencias de la Ingeniería.

VLADIMIR ALEJANDRO RIFFO BOUFFANAIS

RESUMEN

Desde su origen, la inspección de imágenes de rayos X ha sido utilizada en la medicina, para detectar problemas internos en los seres humanos, sin embargo, hoy en día la inspección con rayos X tiene un sin fin de aplicaciones, de las cuales podemos destacar: el análisis de productos alimenticios, inspección de equipaje, inspección de partes y piezas de automóviles y control de calidad de las soldaduras, entre otras. Desde los atentados terroristas ocurridos en septiembre del 2001 en los Estados Unidos, la inspección de equipaje en los aeropuertos ha cobrado real importancia y ha disminuido el riesgo de crímenes y posibles ataques terroristas. Pese a la importancia de esta actividad, la inspección de equipaje en los aeropuertos, todavía es realizada por inspectores humanos con muy poco apoyo tecnológico. Esto no siempre es efectivo, ya que depende en gran medida de la pose del objeto de interés, la oclusión y las capacidades de los inspectores. Por esta razón hemos desarrollado un *framework* para la inspección de imágenes de rayos X, y que mediante visión activa, permite detectar objetos de interés, tales como una hoja de afeitar, estrellas ninja (*shuriken*) y revólveres, en una pose que sea fácilmente reconocible al ojo humano. Para llevar a cabo esta propuesta hemos desarrollado una metodología que incluye, la caracterización de los objetos mediante descriptores SIFT, algoritmos de detección basados en la similitud de características y también en la posición espacial de estas, y algoritmos

que basados en la estimación de la pose que definen la rotación de un sistema de sujeción automático o semi-automático. De esta forma, nuestra propuesta permite buscar un objeto de interés en una primera vista de una imagen de rayos X, y en caso de no ser detectado o la detección indique que el objeto se encuentra en una vista poco representativa o mala pose, el algoritmo indicará la realización del movimiento de rotación del sistema de sujeción automático o semi-automático, para provocar una segunda inspección, y de ser necesario una tercera y hasta una cuarta. Los resultados obtenidos en las diferentes etapas que constituyen a nuestra propuesta, son promisorios, y aunque el requerimiento de hardware es intenso, nos indican que el proceso de inspección no destructiva con imágenes de rayos X de forma activa, hace más eficaz la detección de objetos de interés en entornos complejos, como por ejemplo bolsos, y permite superar algunos niveles de oclusión y desórdenes. Adicionalmente, creemos haber contribuido en el estado del arte con una propuesta original de *framework*, y que puede ser mejorada en trabajos futuros, principalmente en su etapa de detección en una vista.

Palabras Claves: inspección con rayos X, inspección no destructiva, framework, visión activa, buena pose, estimación de pose, aplicación SIFT.

Miembros del Comité:
DOMINGO MERY QUIROZ
ÁLVARO SOTO ARRIAZA
CRISTIÁN TEJOS NÚÑEZ
ROBERTO QUEVEDO LEÓN
LUIS PIZARRO QUIROZ
CRISTIÁN VIAL EDWARDS

Santiago de Chile, Junio, 2016

PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA

OBJECT DETECTION IN X-RAY IMAGES USING ACTIVE VISION

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Doctor in Engineering Sciences by.

VLADIMIR ALEJANDRO RIFFO BOUFFANAIS

ABSTRACT

Since its origin, inspection of X-ray imaging has been used in medicine to detect internal problems in humans. Today, however, X-ray inspection has endless applications, among which we can highlight: the analysis of foodstuffs, baggage inspection, inspection of components and automobile parts, and quality control of welding, among others. Since the terrorist attacks in September 2001 in the United States, baggage inspection at airports has gained real importance and reduced the risk of crime and potential terrorist attacks. Despite the importance of this activity, baggage inspection at airports is still done by human inspectors with little technological support. This is not always effective because it depends largely on the position of the object of interest, occlusion and capabilities of inspectors. For this reason we have developed a *framework* for inspection of X-ray images and through active vision to detect threat objects, such as a razor blade, ninja stars and revolvers, in positions that are easily recognizable to the human eye. To carry out this proposal, we have developed a methodology that includes the characterization of objects by SIFT descriptors, detection algorithms based on the similarity of characteristics and also in the spatial position of these, and algorithms based on the estimation of the pose that define the rotation of an automatic or semi-automatic attachment system. Thus, our proposal allows searching an object of interest in a first view of an X-ray image that if not detected, or the detection indicates that the object is in an unrepresentative view or bad pose, the algorithm

indicates the completion of the rotational movement of the automatic, or semi-automatic, attachment system, to cause a second inspection and, if necessary, a third and even a fourth. The results obtained in the different stages that make up our proposal are promising, and although the hardware requirement is intense, suggest to us that the process of nondestructive inspection with X-ray images make detection of objects of interest more effective in complex environments, such as bags. It also overcomes some levels of occlusion and disorder. In addition, we have contributed to the state of the art with an original *framework* proposal, which can be improved in future work, mainly in the detection step into a single view.

Capítulo 1. INTRODUCCIÓN

En este capítulo hacemos una revisión general de nuestra propuesta de tesis, así como también la hipótesis y los objetivos que dan origen a esta investigación. Además, mostramos otros elementos introductorios que permiten un rápido acercamiento a la 'detección de objetos en imágenes de rayos X usando visión activa'. Al final de este capítulo describimos algunas ideas de trabajos futuros.

1.1. Motivación

Los "Rayos X" fueron descubiertos en 1895 por Wilhelm Röntgen, y los llamó así debido a su naturaleza incierta (Röntgen, 1895). Röntgen pensó equivocadamente que esos rayos no estaban relacionados con la luz, ya que la forma de radiación no exhibió las propiedades de reflexión ni refracción. Así, desde el descubrimiento de los Rayos X, estos han sido útiles en aplicaciones médicas para detectar problemas internos en los humanos. Posteriormente, se encontró que podrían ser útiles para ensayos no destructivos (NDT: *Non-Destructive Testing*) para los materiales y objetos, donde la tarea es analizar (no destructivamente) las partes internas que son indetectables a simple vista. En este contexto, algunas de las principales aplicaciones de inspección que hacen usos de los Rayos X son: análisis de productos alimenticios (Haff y Toyofuku, 2008), inspección de equipaje (Baştan et al., 2011, 2013), inspección de partes y piezas de automóviles (Mery, 2006), y control de calidad de las soldaduras (Silva y Mery, 2007a,b), entre otras. La principal tarea de estas aplicaciones es la detección o reconocimiento de objetos, basado en extracción de características, como por ejemplo, apariencia, forma, gradiente, etc.

La inspección con rayos X puede ser realizada por un inspector humano o por un sistema automatizado. Aunque los seres humanos, en muchos casos pueden hacer el trabajo mejor que las máquinas, estos son más lentos y se cansan rápidamente. Además, los inspectores humanos no siempre son consistentes y efectivos al evaluar objetos, porque las tareas de inspección son monótonas y agotadoras, incluso para los expertos en la materia. Por otro lado, los expertos humanos son difíciles de encontrar o mantener en una industria,

requieren entrenamiento y su aprendizaje puede tomar tiempo. Se ha publicado que la inspección humana en procesos industriales, alcanza como máximo una eficiencia del 80 % (Newman y Jain, 1995), y en otras publicaciones (Schwaninger et al., 2005; Hardmeier et al., 2006; Wales et al., 2009) asociadas a la inspección con rayos X de equipaje en aeropuertos, tiene una eficiencia que no supera el 90 % para el mejor de los casos (inspectores humanos jóvenes, entrenados).

Con el fin de lograr inspecciones con rayos X eficaces y eficientes, los sistemas automáticos de inspección están siendo desarrollados para ejecutar tareas difíciles, tediosas y a veces peligrosas. Comparada con la inspección manual, los sistemas automáticos tienen la ventaja de objetividad y reproducibilidad para cada inspección. La revisión de la literatura muestra que en esta área de investigación se han abordado diferentes enfoques, dependiendo de la aplicación. Sin embargo, la inspección automática de rayos X tiene todavía problemas no resueltos: *i*) *pérdida de generalidad*; porque los enfoques desarrollados para una aplicación no pueden ser utilizados en otros casos, *ii*) *deficiente precisión en la detección*; porque hay un compromiso entre los falsos positivos (falsas alarmas) y la no detección, *iii*) *limitada robustez*; debido a que los pre-requisitos para el uso de un método se cumplen con frecuencia sólo con estructuras simples, y *iv*) *baja capacidad de adaptación*; porque puede ser muy difícil hacer modificaciones de diseño a un sistema automatizado dado.

La inspección de equipaje usando rayos X es una tarea prioritaria que reduce el riesgo de delincuencia y atentados terroristas. Sin embargo, la tecnología existente está lejos de la perfección. Hoy en día no existen métodos completamente automáticos, y los sistemas manuales no están libres del error humano. Desde el 11 de septiembre de 2001 en los Estados Unidos, los aeropuertos y centros aduaneros de todo el mundo han intensificado las restricciones y aumentado la seguridad. La inspección de seguridad mediante escáner de rayos X, se ha convertido en un proceso de mucha importancia en los aeropuertos. Sin embargo, la inspección de bolsos, maletas, y equipajes en general es una tarea compleja ya que los elementos amenazantes son muy difíciles de detectar, cuando: se coloca en paquetes muy cercanos entre sí, ocluidos por otros objetos o girados, presentando así una

vista irreconocible (Zentai, 2008; Bolfing et al., 2008; Schwaninger et al., 2008; Michel et al., 2008). La detección manual de elementos amenazantes realizada por los inspectores humanos, es extremadamente exigente. Por un lado, es tediosa porque muy pocos equipajes contienen en realidad artículos amenazantes, y por otro lado, es estresante porque el trabajo requiere mucha concentración para identificar una amplia gama de objetos y categorías de ellos, formas y sustancias (metales, sustancias orgánicas e inorgánicas), ver Figura 1.1. Además, los inspectores humanos reciben un mínimo apoyo tecnológico, y durante las horas punta en los aeropuertos, los inspectores tienen sólo unos pocos segundos para decidir si un equipaje contiene o no un elemento que podría ser considerado una amenaza. Dado que cada operador debe examinar muchos bolsos, paquetes y equipaje de todo tipo, la probabilidad de error humano crece considerablemente durante un largo período de tiempo.

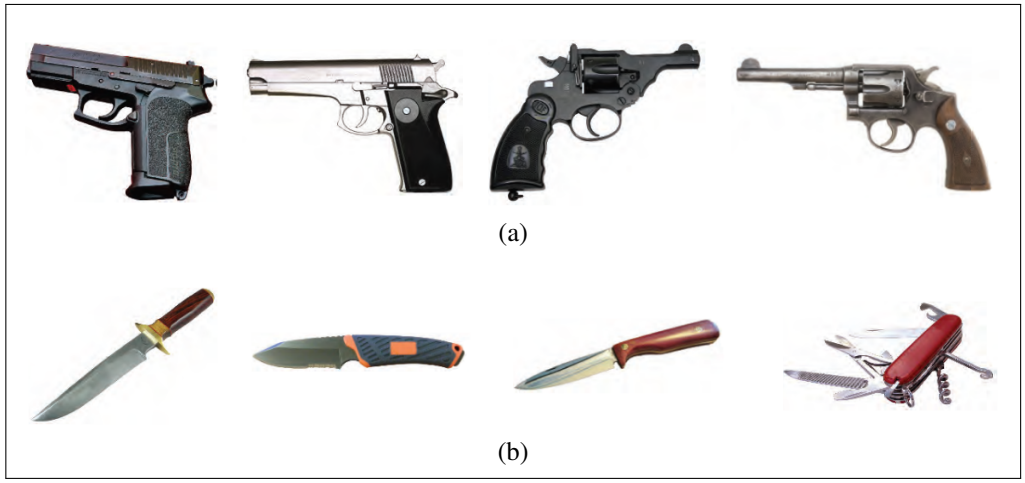


Figura 1.1. Categoría de objetos, que varia considerablemente en su apariencia visual (variaciones inter e intra categoría), a) armas de fuego, y b) armas corto-punzantes.

La detección de objetos en imágenes de rayos X es una tarea compleja para un inspector humano y también para un computador, porque las imágenes de rayos X son imágenes sombra que corresponden a proyecciones en perspectiva de los objetos. La mayoría de los sistemas de detección que usan rayos X, han sido diseñados para mejorar la calidad de las imágenes, a través de la segmentación (Megherbi et al., 2013; Heitz y Chechik, 2010) y algoritmos de pseudo-color (Abidi et al., 2006; Kase, 2002; Chan et al., 2010). Algunos enfoques utilizan detección automática de objetos amenazantes con una o múltiples vistas

usando uno o dos niveles de energía para la emisión de rayos X (Mery, 2014). Ejemplo de esto es el método para la detección de armas de fuego, basado en *bolsa de palabras visuales* (Bag of Words, BoW) el cual ha demostrado obtener buenos resultados, usado un enfoque de dos vistas, con imágenes en pseudo-color, adquiridas con dos niveles de energía (Baştan et al., 2011; Turcsany et al., 2013). También ha sido usado el algoritmo de clasificación llamado *máquinas de soporte vectorial* (Support vector machines, SVM), con múltiples vistas y dos niveles de energía en la adquisición de las imágenes de rayos X Baştan et al. (2013). En la literatura se han reportado algoritmos que permiten la detección en múltiples vistas, con el fin de eliminar falsos positivos y aumentar el rendimiento en la detección: *i*) el método que incluye una síntesis de imágenes de rayos X con efecto de profundidad cinética (Kinetic Depth Effect X-ray images, KDEX) (Abusaeeda et al., 2011), *ii*) el método que usa la reproyección en el espacio 3D de todas las detecciones en las distintas vistas y de esta forma agrupar las detecciones correctas (Franzel et al., 2012; Mery, Riffo, et al., 2013), *iii*) el método de seguimiento a través de las múltiples vistas de imágenes de rayos X, con el fin de verificar el diagnóstico realizado en una sola vista (Mery, Mondragon, et al., 2013; Mery, 2015). En todos estos enfoques, la detección de objetos amenazantes en las vistas individuales juega un rol fundamental. Sin embargo, sabemos que para realizar una detección efectiva es necesario contar con un método que sea capaz de analizar de manera eficiente los objetos desde múltiples puntos de vista.

1.2. Hipótesis

Los algoritmos de procesamiento de imágenes y visión por computador han sido de gran utilidad en la inspección con rayos X, principalmente en aplicaciones médicas y en algunos interesantes avances en inspección no destructiva de alimentos y en procesos industriales en general. En su gran mayoría las aplicaciones han sido desarrolladas para ayudar a inspectores humanos o personal del área de la salud, para que puedan tomar mejores decisiones sobre la base de lo que se observa en las imágenes de rayos X. Muy pocos algoritmos totalmente automatizados y con capacidades de decisión han sido desarrollados. Este avance poco significativo en la inspección automática con rayos X se debe a muchos factores,

pero el principal es que las imágenes de rayos X son difíciles de interpretar, ya que se trata de una proyección de muchas sombras, muchas de las cuales se encuentran en oclusión. Los avances van desde algoritmos de segmentación, hasta aplicaciones en multiples vistas: *i*) para mejorar la calidad de las imágenes, *ii*) para diferenciar una zona, sustancia u objeto de otros, y *iii*) para detectar objetos.

No siempre es posible la correcta detección, ni tampoco la detección en una pose donde el objeto de interés sea fácilmente reconocible, debido a las dificultades que se manifiestan en los procesos de inspección con rayos X, tales como: oclusión y puntos de vista inadecuados, entre otras. Según lo aquí expuesto, es que planteamos nuestra hipótesis que da origen a nuestro trabajo de investigación:

La detección automática de un objeto de interés en imágenes de rayos X, se verá ampliamente favorecida si se realiza de forma activa, es decir, modificando el punto de vista de observación del objeto de manera automática, hasta que la detección del objeto de interés, en una o múltiples vistas, sea en la posición más representativa (mejor vista). Pensamos que la inspección activa en imágenes de rayos X, hace más eficaz la detección, soslayando algunos niveles de oclusión y vistas poco reconocibles de un objeto.

1.3. Objetivos

A continuación detallamos el objetivo general y los objetivos especifico planteados en esta investigación, de tal forma de señalar las metas que deben ser cumplidas para verificar el cumplimiento de nuestra hipótesis.

A. Objetivo General

El principal objetivo de esta tesis es desarrollar un *framework* [1] para la inspección activa de imágenes de rayos X, es decir, modificando el punto de vista de observación del

[1] *framework*: define, en términos generales, un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular, que sirve como referencia para enfrentar y resolver nuevos problemas de índole similar. [Tomada de: `https://es.wikipedia.org/wiki/Framework`, septiembre del 2015].

objeto de manera automática, hasta que la detección del objeto de interés, en una o múltiples vistas, sea en la posición más representativa (mejor vista).

B. Objetivos Específicos

Para lograr nuestro principal objetivo, debemos cumplir con los objetivos específicos que se detallan a continuación:

a) Diseñar un sistema de adquisición de imágenes de rayos X y conceptualizar el modelo geométrico que lo representa,
b) Construir una base de datos de imágenes para caracterizar un objeto de interés, y una base de datos para las pruebas de detección,
c) Implementar un sistema para caracterizar objetos de interés representados en imágenes de rayos X,
d) Implementar algoritmos de visión activa para la detección de objetos de interés en imágenes de rayos X,
e) Evaluar la propuesta de inspección activa de objetos de interés, mediante índices de desempeño.

1.4. Contribuciones

Cuando los inspectores realizan la inspección radioscópica de equipaje en los aeropuertos, cuentan con sólo una imagen de rayos X (en escala de grises y/o pseudocolor) para tomar una decisión de vital importancia, es decir, encontrar un sin fin de objetos amenazantes (orgánicos e inorgánicos) que pudiesen poner en riesgo la vida de las personas que viajarán en el avión. La literatura reporta que para estos propósitos una imagen no es suficiente, pues los objetos de interés (objetos amenazantes), pueden estar ocluidos total o parcialmente y/o en posiciones que no permiten su fácil reconocimiento. En esta investigación proponemos un método de inspección en múltiples vistas de imágenes de rayos X, que realiza la búsqueda automática de algunos objetos de interés (amenazantes) de forma activa, que pueden estar presentes en un bolso u otro objeto contenedor, pero que no son

fácilmente reconocibles o detectables en una única vista (debido a la posición en la que se encuentran), así, nuestra propuesta realiza los movimientos necesarios (de un bolso u objeto contenedor), hasta encontrar en la imagen de rayos X al objeto de interés en una *Buena Pose* (*BP*[2]), ver Figura 1.2.

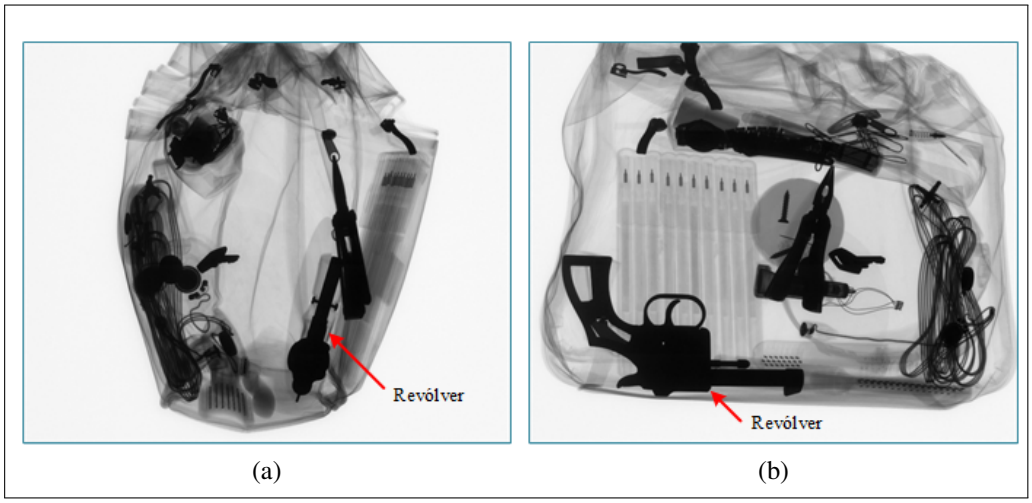


Figura 1.2. Dos vistas de de un objeto inspeccionado con rayos X, a) revólver en una mala pose, b) revólver en una buena pose.

El análisis de cada vista consiste en la detección del objeto amenzante, y para ello hemos usado dos enfoques: a) Un enfoque similar al propuesto por Lowe (2004); que consiste en hacer *matching* de descriptores SIFT con una base de datos, y b) Un enfoque que esta basado en el modelo de forma implícita (*Implicit Shape Model*, ISM) propuesto por Leibe y Schiele (2003); que en nuestro caso es una adaptación, la cual hemos denominado AISM (*Adapted Implicit Shape Model*) y considera a los objetos como un conjunto de partes independientes, pero conectada lógicamente, a través de una estructura de estrella y que nos permite detectar categoría de objetos.

Existen diferentes configuraciones que podrían ser utilizadas para la adquisición de las imágenes de rayos X y así realizar una inspección activa: a) mover sólo el emisor de rayos

[2]En este trabajo se entenderá que un objeto de interés se encuentra en una *Buena Pose* (*BP*), si este es fácilmente reconocible a simple vista o por un sistema automático de inspección, y esto ocurre cuando la superficie registrada en el plano de proyección es similar a la mayor superficie visible del objeto de interés.

X, b) mover simultáneamente el emisor y receptor de rayos X, y c) mover el objeto que será sometido a inspección, ver Figura 1.3.

En nuestro caso y por razones técnicas y económicas, hemos decidido mover el objeto que será inspeccionado, con lo cual restringimos esta propuesta a objetos rígidos o que estén en una posición de reposo. Nuestro enfoque permite detectar objetos amenazantes, y en principio podría ser útil para detectar cualquier tipo de objeto en imágenes de rayos X. Objetos que deben ser caracterizados previamente y asumiendo las condiciones básicas para la adquisición de imágenes de rayos X: imágenes de rayos X en escala de grises (sin pseudolor), con un solo punto de vista que varía según avance la inspección activa, imágenes adquiridas con solo un nivel de energía y sin ningún algoritmo de procesamiento.

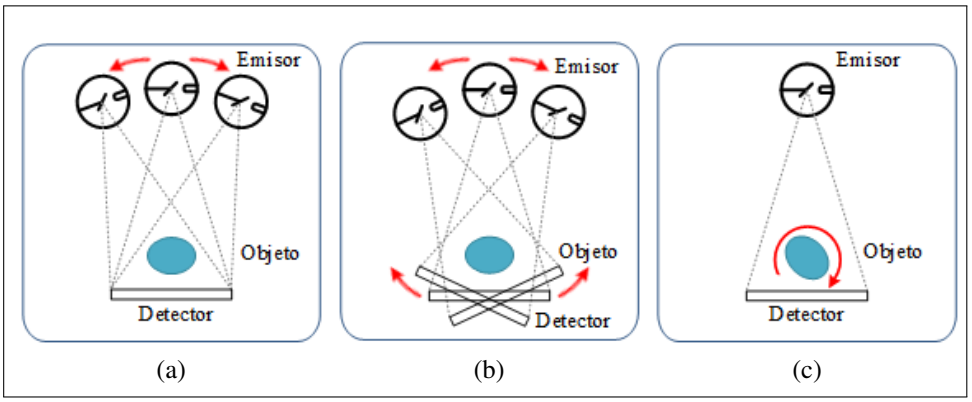


Figura 1.3. Configuraciones que permiten realizar la *Visión Activa* con rayos X, a) movimiento del emisor, b) movimiento simultáneo del emisor y detector , y c) movimiento del Objeto.

Hemos implementado y aplicado un método para detectar automáticamente y de manera activa, objetos amenazantes dispuestos al interior de bolsos, carteras, etc. Como objetos amenazantes hemos considerado: hoja de afeitar, shurikens (estrellas ninja) y revólveres. Mostramos la robustez del enfoque en la detección activa de objetos simétricos y objetos de forma regular (hoja de afeitar y shurikens) y aceptables resultados para objetos de forma irregular (revolvers). Estos resultados son prometedores y establecen un enfoque genérico para la detección de objetos en las imágenes de rayos X, y podría ser una herramienta útil para ayudar a los inspectores humanos en el aeropuerto o centros aduaneros. Creemos que los resultados podría mejorar en trabajos futuros, incorporando otros algoritmos de

detección en una vista y que se adapten al proceso de búsqueda de la siguiente vista en el algoritmo de visión activa, también creemos que es posible en trabajos futuros mejorar los resultados, si se elimina la restricción de objetos rígidos, es decir, implementando un sistema que permita mover el tubo emisor y/o detector de rayos X.

1.5. Organización del Documento

Este trabajo continua con otros cinco capítulos, que explican los aspectos más relevantes de esta investigación, y anexos con publicaciones indexadas relacionadas directamente con esta propuesta, así como también anexos con imágenes de la propuesta inicial de *framework*. El contenido de cada capítulo se describen a continuación:

- Capítulo 2. MARCO TEÓRICO: Describe los fundamentos teóricos necesarios para abordar y desarrollar la propuesta de investigación.
- Capítulo 3. ESTADO DEL ARTE: Detalla los estudios previos relacionados con la inspección con rayos X, usando una y múltiples vistas.
- Capítulo 4. METODOLOGÍA PROPUESTA: Proporciona una discusión detallada de la metodología propuesta para demostrar nuestra hipótesis, y así lograr la detección de objetos de interés en imágenes de rayos X, utilizando visión activa.
- Capítulo 5. RESULTADOS EXPERIMENTALES: Presenta el análisis experimental de este trabajo, y la evaluación de desempeño del framework que permite detectar objetos de interés en imágenes de rayos X, utilizando visión activa.
- Capítulo 6. DISCUSIÓN Y CONCLUSIONES: Presenta las conclusiones asociadas a los objetivos general y específicos del trabajo desarrollado, las contribuciones de esta investigación, y una discusión de las áreas que pueden ser exploradas en el futuro.
- ANEXOS: Presenta las publicaciones como autor y co-autor en revistas científicas indexadas, asociadas a la inspección no destructiva con rayos X. Adicionalmente se incorpora la primera imagen adquirida de cada secuencia de inspección realizada en la propuesta inicial de *framework*, usando un sistema semi-automático y un sistema automático de sujeción y rotación.

Capítulo 2. MARCO TEÓRICO

Los sistemas de inspección con rayos X son muy versátiles, pues permiten detectar una serie de objetos, sustancias orgánicas e inorgánicas, desperfectos, etc. dando garantías de calidad y seguridad. Sin embargo, la detección automática está muy poco desarrollada, y los algoritmos existentes, en su gran mayoría, operan en ambientes controlados o con muy poca variabilidad, donde es relativamente simple hacer segmentaciones, para diferenciar entre un error, defecto o elemento espurio o amenazante, de una situación normal. Por otra parte están los sistemas semi-automáticos, que permiten mejorar las características visuales de las imágenes, para que un operador o inspector humano tome decisiones asociadas a la seguridad y/o calidad. En los ambientes no controlados, los objetos de interés pueden estar en distintas posiciones e incluso con algún nivel de oclusión que impida su fácil detección, ya sea para un humano o para un algoritmo, un ejemplo de esta situación lo viven a diario los inspectores de equipaje en los aeropuertos, que utilizan rayos X para hacer inspección de bolsos, maleta, mochilas, etc. para detectar un sin número de objetos y sustancias, y sólo cuentan con ayuda de algoritmos de pseudocolor, para poder discriminar entre equipaje seguro o inseguro. Como vemos, no es simple hacer propuestas totalmente automáticas.

En esta investigación hemos visto que una de las mayores dificultades en la inspección con rayos X, es determinar el punto de vista adecuado de adquisición de las imágenes, para poder evitar que los objetos de interés o amenazantes se encuentren en posiciones irreconocibles o indetectables. Es imposible a simple vista saber el contenido y disposición de los objetos al interior de un bolso, mochila, cartera, o cualquier equipaje, y es por eso que hemos planteado esta investigación, ya que al implementar un modelo de inspección activa, que permita modificar el punto de vista de adquisición e inspección de las imágenes de rayos X, podremos soslayar malas vistas e incluso algunos niveles de oclusión. Para poder implementar esta propuesta, hemos necesitado comprender algunos fenómenos físicos asociados a la formación de imágenes de rayos X, y utilizar, modificar y desarrollar algoritmos de procesamiento de imágenes, visión por computador y reconocimiento de patrones. Una mirada general y en algunos casos con mayor detalle de todos estos aspectos, lo veremos

en este marco teórico. Para dar respuesta a nuestra hipótesis, debemos conocer y aplicar los elementos conceptuales que aquí se describen, que forman parte de la propuesta de *framework*: caracterizar, detectar, mover, procesar, clasificar, y modelar, entre otras, son algunas de las etapas que constituyen y que son necesarios para la concepción de la propuesta metodológica descrita en este trabajo.

En este capítulo se abordan los fundamentos teóricos más relevantes y necesarios para desarrollar esta tesis. El capítulo se ha organizado de la siguiente forma: en la sección 2.1 veremos la formación de las imágenes de rayos X y algunos aspectos que influyen en la calidad de dichas imágenes, en la sección 2.2 veremos los factores que afectan la inspección con rayos X, en la sección 2.3 veremos de forma general los descriptores locales SIFT y su uso en imágenes de rayos X, en la sección 2.4 veremos el modelo geométrico que describe al sistema de adquisición de imágenes de rayos X propuesto, en la sección 2.5 veremos los fundamentos que permiten establecer la correspondencia entre imágenes de rayos X, en la sección 2.6 veremos la estrategia que hemos usado en esta tesis para realizar el movimiento de los objetos y así encontrar la siguiente mejor vista, finalmente en la sección 2.7 veremos el modelo de forma implícita (ISM) para la detección de objetos, el cual hemos adaptado para hacerlo útil en la detección de objetos en imágenes de rayos X.

2.1. Formación de Imágenes de Rayos X

Las detección con rayos X es una forma de *ensayos no destructivos* (NDT), definida como una tarea que utiliza imágenes de rayos X para determinar si un *objeto de prueba* se desvía de un conjunto de especificaciones dado, sin cambiar o alterar ese objeto de alguna forma (Hellier, 2013). En los ensayos no destructivos con rayos X, la radiación pasa a través del objeto de prueba, y un detector capta una imagen de rayos X, correspondiente a la intensidad de la radiación atenuada por el objeto[1].

Una imagen radioscópica puede ser considerada como la proyección de un objeto dispuesto entre una fuente (tubo emisor) y un detector de rayos X (Quinn et al., 1980). Esta

[1]Los rayos X pueden ser *absorbidos* o *dispersados* por el objeto de prueba. En este trabajo vamos a cubrir sólo la primera interacción.

imagen tendrá diferentes niveles de gris de acuerdo al principio de *absorción fotoeléctrica* (Als-Neielsen y McMorrow, 2011).

Cuando un haz de rayo X atraviesa la materia, una porción de energía de este rayo es absorbida. La intensidad I_o de un haz de rayo X incidente se atenúa a lo largo de una dirección dz por una intensidad dI, en comparación a la intensidad inicial I. Con μ definido como *coeficiente de atenuación lineal* (Zschornack, 2007), como se ilustra en la Figura 2.1.

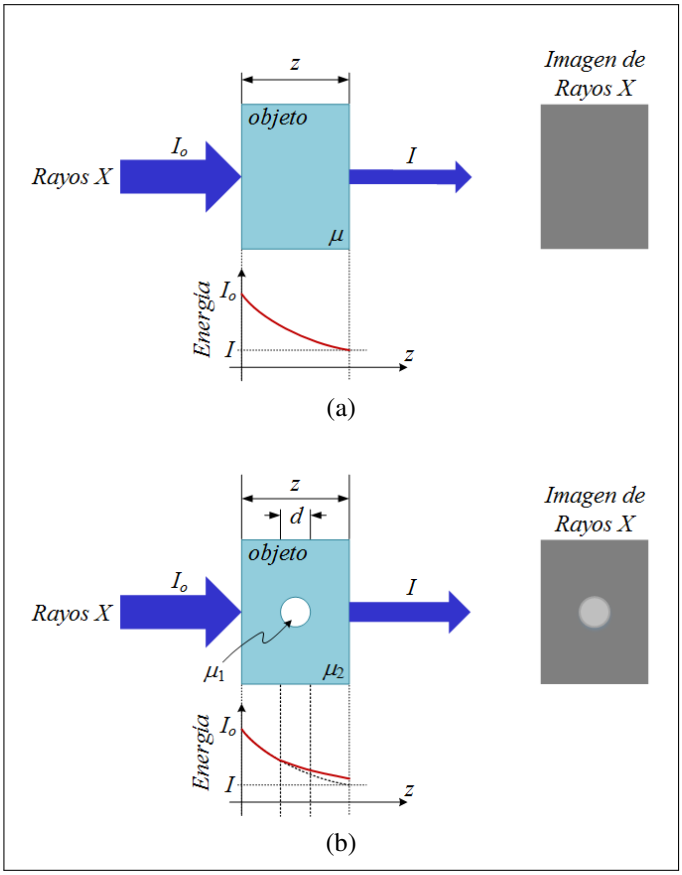


Figura 2.1. Formación de imágenes de rayos X de acuerdo con la ley de absorción, a) Imagen de rayos X de un objeto homogéneo, y b) Imagen de rayos X de un objeto con dos materiales diferentes.

$$\frac{dI}{I} = -\mu\, dz \tag{2.1}$$

el signo menos que acompaña al coeficiente de atenuación μ indica el decremento de I (medido en [fotones/seg.]) al aumentar z. Típicamente μ es expresado en $[\text{cm}]^{-1}$ por lo que z se expresa en [cm]. Integrando indefinidamente la ecuación (2.1), tendremos,

$$\ln(I) = -\mu\, z + C \qquad C\text{: Constante de integración} \tag{2.2}$$

si se tiene como condición inicial que $I = I_o$ cuando $z = 0 \longrightarrow C = \ln(I_o)$, de esta forma y reemplazando en la ecuación (2.1),

$$\ln\left(\frac{I}{I_o}\right) = -\mu\, z \tag{2.3}$$

con lo que se obtiene la expresión:

$$I = I_o\, e^{-\mu\, z} \tag{2.4}$$

La ecuación anterior representa el comportamiento de la intensidad transmitida I, que depende de la intensidad de la radiación incidente I_o, el espesor z del objeto de prueba y el coeficiente de atenuación lineal μ asociado al material.

Los sistemas de inspección con imágenes de rayos X más usados son la radiografía digital (RD) y la tomografía computarizada (TC). Por un lado, RD enfatiza un alto rendimiento, ya que utiliza sensores electrónicos (en lugar de la película radiográfica tradicional) para obtener una proyección de rayos X digital del objeto de interés, de manera simple y rápida. Un detector plano de silicio se puede utilizar como un sensor de imagen en sistemas de ensayos con rayos X. En tales detectores, que usan un semi-conductor, la energía de los rayos X se convierte directamente en una señal eléctrica que puede ser digitalizada y mostrada como una imagen digital de rayos X (Rowlands, 2002). Por otro lado, la TC proporciona una imagen de sección transversal del objeto de interés, de manera que cada objeto se separa claramente uno de otro, Sin embargo, las imágenes de TC requiere un número considerable de proyecciones para reconstruir una imagen precisa de una sección transversal, lo cual consume tiempo.

Un análisis similar al que permitió obtener I en la ecuación 2.4 puede ser realizado si la radiación de rayos X pasa a través de dos materiales diferentes, con coeficientes de atenuación μ_1 y μ_2 respectivamente, como se muestra en la Figura 2.1b. Así, la intensidad transmitida I puede ser expresado como:

$$I = I_o \, e^{-\mu_2 \, z} \, e^{d\,(\mu_2 - \mu_1)} \tag{2.5}$$

Esto explica la generación de la imagen de las regiones que están presentes en el interior del objeto de interés, como se muestra en la Figura 2.1b, donde una burbuja de gas es claramente detectable. Sin embargo, las imágenes de rayos X a veces contienen objetos que se superponen, lo que es muy difícil distinguir adecuadamente.

Una imagen radioscópica tiene representado objetos mediante niveles de gris, cuyos valores dependen de los ajustes de voltaje (V) y corriente (A) del emisor de rayos X, y del coeficiente de atenuación (μ) de los objetos. En términos físicos, el ajuste del voltaje del tubo emisor de rayos X controla la energía de los fotones de rayos X, mientras que el ajuste de la corriente en el tubo emisor afecta la cantidad de fotones de rayos X que se emite. En términos de imágenes, el voltaje controla el contraste de los objetos que se ven en la imagen y la corriente, controla la razón señal a ruido de la imagen. Una imagen radioscópica se verá más oscura para valores bajos de energía, y por el contrario, la imagen tiende a ser más clara para valores altos de energía. Como ya se ha dicho, el nivel de gris depende también de μ, entonces, no se puede decir con certeza los valores mínimos o máximos de V y A, ya que estos valores se ajustan manualmente (valores empíricos) hasta obtener imágenes con el mayor contraste y nitidez posible. El valor de voltaje se mide normalmente en kilo-voltios [KV] y el valor de corriente se mide en mili-amperes [mA], este ajuste normalmente reciben el nombre de *ajuste de la técnica*. Otros dos factores que hacen que los contornos se vean más o menos nítidos, son a) el tamaño del punto focal del emisor de rayos X, que idealmente debiese ser un punto cuyo diámetro tienda a cero, y b) la distancia del objeto en estudio respecto al punto focal del emisor de rayos X. Ambos efectos son visualizados en la figura 2.2

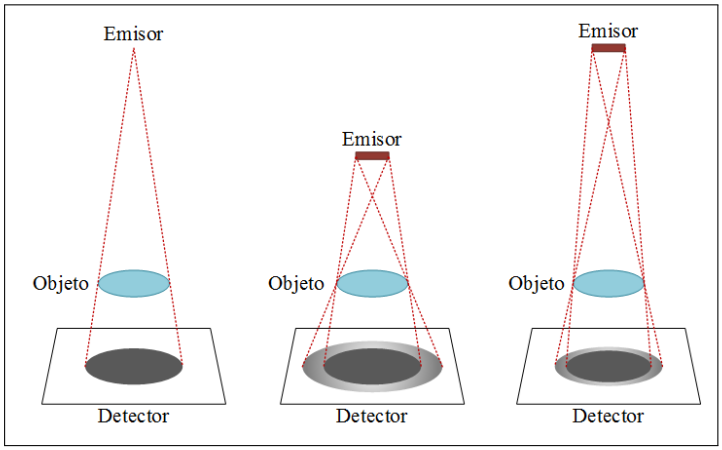


Figura 2.2. Efecto del tamaño del punto focal del emisor de rayos X y de la distancia de este punto hacia el objeto, (Quinn et al., 1980).

2.2. Factores que afectan la Inspección con Rayos X

En la inspección radioscópica de objetos complejos existen tres factores importantes que pueden influir en la búsqueda de algún objetos de interés, estos tres factores han sido descritos en el marco de la inspección de equipaje en los aeropuertos (Schwaninger, 2003; Zhigang et al., 2005): dificultades visuales, superposición, y complejidad del objeto.

***A*. Dificultades Visuales:** Esto se debe principalmente a las dificultades que tienen los inspectores humanos para visualizar imágenes de rayos X de objetos de interés en poses que impiden su reconocimiento. Influyen también en este sentido, la predisposición del inspector a buscar elementos que pueda considerar de mayor amenaza (en el caso de objetos prohibidos) y el nivel de entrenamiento y experiencia en dichos proceso visual.

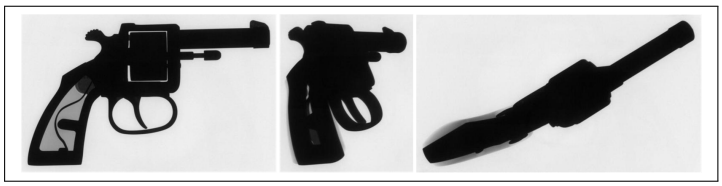


Figura 2.3. Dificultades Visuales: Los objetos rotados son mas difíciles de detectar (dependencia del punto de vista).

***B*. Superposición:** Este factor hace referencia al grado de oclusión en que puede encontrarse el objeto de interés. Existirán niveles de oclusión que harán imposible la detección para un inspector humano e incluso para un sistema automático.

Figura 2.4. Superposición: Cuando los objetos de interés están superpuestos por otros objetos, el desempeño de la detección decrece (efecto de superposición).

***C*. Complejidad del Objeto:** Este factor está asociado a la disposición espacial del objeto de interés y los elementos que lo rodean, que pueden afectar a la dirección visual de inspección. Además, influye en la complejidad, el hecho de que algunos objetos tienen mayor o menor nivel de absorción a los rayos X, lo que puede causar que sean mas perceptibles objetos distintos al objeto de interés. Entonces, los dos sub-factores son: *Desorden* y *Opacidad*.

Figura 2.5. Complejidad del Objeto: La identificación de un objeto amenazante en un bolso o equipaje compacto es más difícil de realizar (efecto de la complejidad del equipaje).

Un factor adicional que influye en menor medida es el ***Tamaño del Objeto***, que hace que la inspección visual abarque una mayor superficie, con un incremento en el tiempo de búsqueda. El desempeño en la detección es determinado por la capacidad del inspector humano para detectar elementos prohibidos o amenazantes (objeto de interés) a pesar de la pose, superposición y la complejidad del bolso o equipaje.

2.3. Descriptores Locales SIFT en Imágenes de Rayos X

SIFT (*Scale Invariant Feature Transform*) (Lowe, 2004) es un algoritmo que permite detectar y describir características locales en imágenes. La detección y descripción de características locales en las imágenes es útil en el reconocimiento de objetos. Las características locales SIFT se basan en la apariencia del objeto y en los puntos de interés, y son invariantes a la rotación y escalamiento de la imagen. También son robustos a los cambios en la iluminación, el ruido, y cambios menores en el punto de vista. Además de estas propiedades, son muy distintivos, son relativamente fáciles de extraer, para permitir la correcta identificación de objetos con baja probabilidad de desajuste y son fáciles de parear (hacer *matching*) con grandes base de datos de características locales. Los descriptores SIFT son también robustos a la oclusión parcial.

El costo de la extracción de características se minimiza teniendo un enfoque de filtrado en cascada, en que las operaciones más costosas se aplican sólo en los lugares que pasan una prueba inicial. A continuación se muestran las principales etapas de cálculo que permiten generar el conjunto de características de la imagen.

a) Detección de regiones '*extrema*' en escala-espacio,
b) Localización de *keypoints*,
c) Asignación de orientación, y
d) Generación de *keypoints*.

a) Detección de regiones '*extremas*' en escala-espacio

En esta etapa se identifican regiones y escalas que pueden ser asignadas de manera repetitiva bajo diferentes vistas del mismo objeto. Lo anterior se realiza buscando características estables a través de todas las escalas posibles, utilizando una función continua de escala conocida como: función de escala-espacio (Witkin, 1984). El único núcleo escala-espacio posible es la función Gaussiana, por lo que la escala-espacio de una imagen se define como una función $L(x, y, \sigma)$ producida de la convolución de una variable de escala

Gaussiana $G(x, y, \sigma)$, con una imagen de entrada $I(x, y)$:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2.6}$$

donde '$*$' es la operación de convolución, x e y son las distancia desde el origen en el eje horizontal y vertical respectivamente, σ es la desviación estándar de la distribución Gaussiana, con:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{2.7}$$

La función es eficiente ya que las imágenes suavizadas L, se calculan necesariamente para la descripción de características escala-espacio, y D se puede obtener con una simple substracción, ver Figura 2.6.

$$\begin{aligned} D(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) && (2.8) \\ &= L(x, y, k\sigma) - L(x, y, \sigma) && (2.9) \end{aligned}$$

donde k es fijo y depende de la finura de la discretización del escala-espacio.

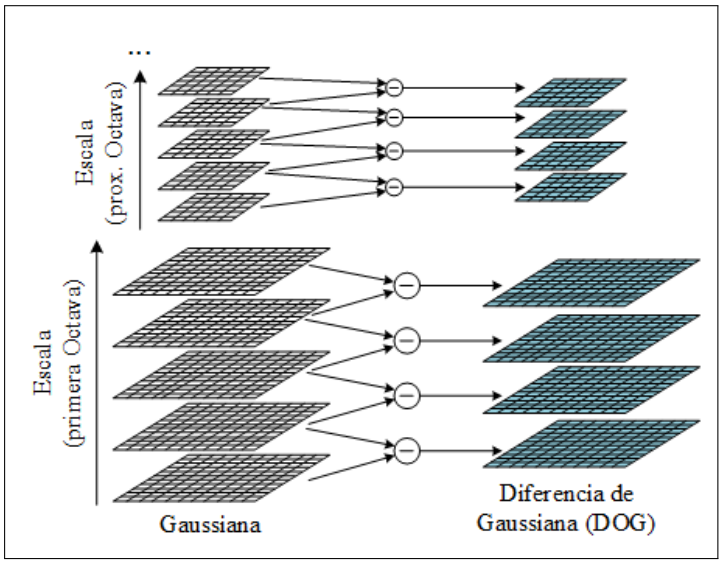


Figura 2.6. Para cada conjunto de imágenes suavizadas con el operador σ (Octava del escala-espacio), la primera imagen es sometida a repetidas convoluciones con la función Gaussiana (σ aumenta al doble en cada convolución). Las imágenes Gaussianas adyacentes se restan para producir la diferencia de imágenes Gaussianas, (Lowe, 2004).

En resumen una pirámide de Diferencia de Gaussianas (DoG), se construye a partir de la pirámide de las imágenes Gaussianas. Entonces se realiza una búsqueda exhaustiva sobre la pirámide para identificar puntos de interés sobre escala-espacio, ver Figura 2.7.

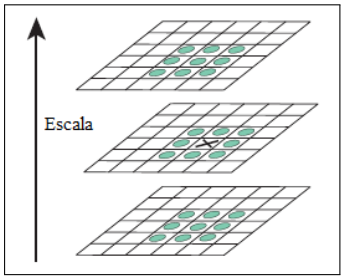


Figura 2.7. Máximo y mínimo de la diferencia de imágenes Gaussianas que se detectan mediante la comparación de un pixel (marcado con X), con sus 26 vecinos en las regiones de 3×3 en la escala actual y adyacentes (marcado con círculos), (Lowe, 2004).

b) Localización de *keypoints*

En esta etapa se rechazan puntos con bajo contraste y puntos que están localizados a lo largo de un borde, luego se procede a ajustar un punto de interés (*keypoint*) a algún dato cercano para efectos de localización, escala y radio de las curvaturas principales. Nuevamente se procede a rechazar puntos con bajo contraste y puntos que estén localizados junto a un borde. Para la localización de *keypoints* se siguen dos enfoques: *Enfoque inicial*, que consiste en localizar los *keypoints* en la localidad y escala del 'punto muestra central'. Y el *Nuevo enfoque*, que consiste en tratar de calcular la localidad de interpolación del máximo, mejorando de esta manera las correspondencias y estabilidad. Se utiliza la *Expansión Cuadrática de Taylor* de la función DoG, $D(x, y, \sigma)$, para que el origen esté en ese punto de muestreo (x es el desplazamiento de este punto):

$$D(\mathrm{x}) = D + \frac{\partial D^{\mathrm{T}}}{\partial \mathrm{x}}\,\mathrm{x} + \frac{1}{2}\mathrm{x}^{\mathrm{T}}\frac{\partial^2 D}{\partial \mathrm{x}^2}\,\mathrm{x} \tag{2.10}$$

donde, D y sus derivadas se evalúan en el punto de muestreo y $\mathrm{x} = (x, y, \sigma)^{\mathrm{T}}$ es el desplazamiento desde este punto. La ubicación del extremo (estimación de x), $\hat{\mathrm{x}}$, se determina tomando la derivada de esta función con respecto a x y se establece a cero, dando,

$$\hat{\mathrm{x}} = -\frac{\partial^2 D^{-1}}{\partial \mathrm{x}^2}\,\frac{\partial D}{\partial \mathrm{x}} \tag{2.11}$$

de esta forma, si $\hat{\mathrm{x}} > 0.5$, indica que el extremo está más cerca a un punto diferente y necesita ser recalculado. Por el contrario, para conseguir el extremo estimado, se debe sumar el desplazamiento a la localidad del punto muestreado.

Para el rechazo del *keypoint* se calcula el valor de D en el punto extremo $\hat{\mathrm{x}}$, y si $|D(\hat{\mathrm{x}})| < 0.03$, se descarta el *keypoint* por tener un contraste bajo. Por estabilidad aún es necesario eliminar aquellos *keypoints* que correspondan a *peaks* poco delineados en la DoG, para mas detalles, ver sección 4.1 en (Lowe, 2004).

c) Asignación de orientación

A cada *keypoint* se le asigna una o más orientaciones, basándose en las direcciones del gradiente local de la imagen. Luego los datos son transformados en relación a la orientación asignada, escala y localidad, y es así como se provee invariancia a estas transformaciones. Para el cálculo del gradientes, se utiliza la escala de un *keypoint* para seleccionar la imagen suavizada L en la que se realizan las distintas operaciones (imagen con la escala más cercana). Todos los cálculos se ejecutan en una forma invariante a la escala y se calcula la magnitud del gradiente y orientación utilizando la diferencia de pixeles:

$$m(x,y) = \sqrt{((L(x+1,y) - L(x-1,y))^2 + ((L(x,y+1) - L(x,y-1))^2} \tag{2.12}$$

$$\theta(x,y) = \tan^{-1}\left(\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right) \tag{2.13}$$

El histograma de orientación consta de 36 barras (cada barra cubre 10 grados) y cada muestra añadida a alguna barra del histograma es valorizada por la magnitud de su gradiente y por una ventana circular de peso Gaussiano con $\sigma = 1.5$ veces la escala del *keypoint*. Se detectan los máximos globales en el histograma y luego los máximos locales que están en el 80 % del punto más alto, y se utilizan estos para asignar (una o más) orientaciones. En resumen la orientación también se calcula en la escala (más próxima) en la cual el *keypoint* es detectado. A partir de la imagen Gaussiana suavizada de la escala apropiada (calculada antes, al momento de construir la pirámide Gaussiana), la magnitud $m(x,y)$ y orientación

$\theta(x, y)$ del gradiente se calculan a partir de la ecuación (2.12) y (2.13) respectivamente. Para asignar la orientación a un *keypoint*, se construye un histograma de las orientaciones de los gradientes de los puntos alrededor del *keypoint*. Barras de 10 grados de ancho se utilizan y la asignación se realiza después del ajuste de la parábola al máximo del histograma. Si el histograma muestra varios máximos, puede haber múltiples *keypoints* en la misma posición con diferentes orientaciones, ver Figura 2.8.

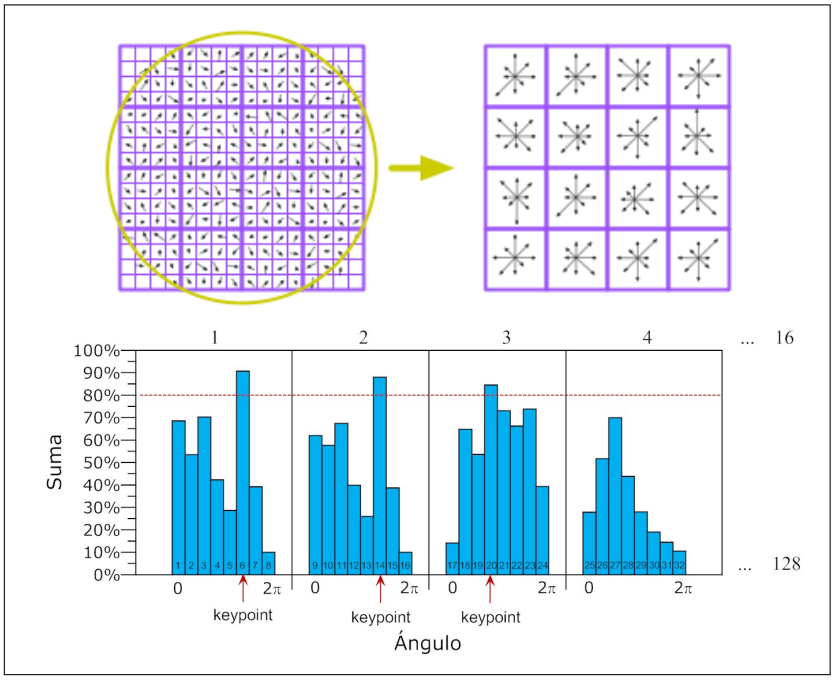


Figura 2.8. Asignación de orientación de uno o varios *keypoint*. Imagen basada y adaptada desde: (Lowe, 2004).

d) Generación de *keypoints*

Para un descriptor se determina una escala (maximizando DoG en escala y en espacio) y la orientación local como la dirección de gradiente dominante. Se utiliza esta escala y orientación para hacer más cálculos invariantes a escala y rotación. Se calculan los histogramas de orientación por gradiente de varias ventanas pequeñas (128 valores para cada punto) y se normaliza el descriptor para hacerlo invariante al cambio de intensidad. Para representar un descriptor; se usa la escala del *keypoint* para seleccionar el grado de 'borrosidad' Gaussiana. Luego se muestra la magnitud y orientación del gradiente alrededor

del *keypoint* y se asigna un peso a la magnitud utilizando la función de peso Gaussiano, $\theta = 0.5\, w$ (w es el ancho de la ventana del descriptor). Esto provee un cambio gradual y da menor énfasis a los gradientes, lejos del *keypoint*.

Los descriptores SIFT han sido ampliamente utilizados en la detección y reconocimiento de objetos sobre imágenes fotográficas ópticas (luz visible). En los últimos años se ha demostrado su versatilidad sobre imágenes de rayos X (Chan et al., 2009), y los resultados demuestran que los descriptores SIFT son suficientemente discriminativos como para lograr alta dispersión inter-clases y baja dispersión intra-clase. El éxito de SIFT en las imágenes de rayos X, puede estar asociada con dos factores: el contraste local y la similitud local, que afectan directamente a la detección de la extremos locales. Estos factores también juegan un papel importante en la búsqueda de correspondencias en aplicaciones de luz visible. Sin embargo, la propiedad de transparencia de las imágenes de rayos X, sumado a movimientos de rotación que eventualmente provocan grandes oclusiones entre dos o más objetos, hace que la correspondencia entre estas imágenes, usando SIFT directamente, no sea siempre lo más adecuado. Así, en esta tesis se plantea la incorporación de SIFT, sumado a un modelo que considere rotar al objeto hasta lograr disminuir el nivel de oclusión. Un ejemplo de extracción de características SIFT se ve en la Figura 2.9.

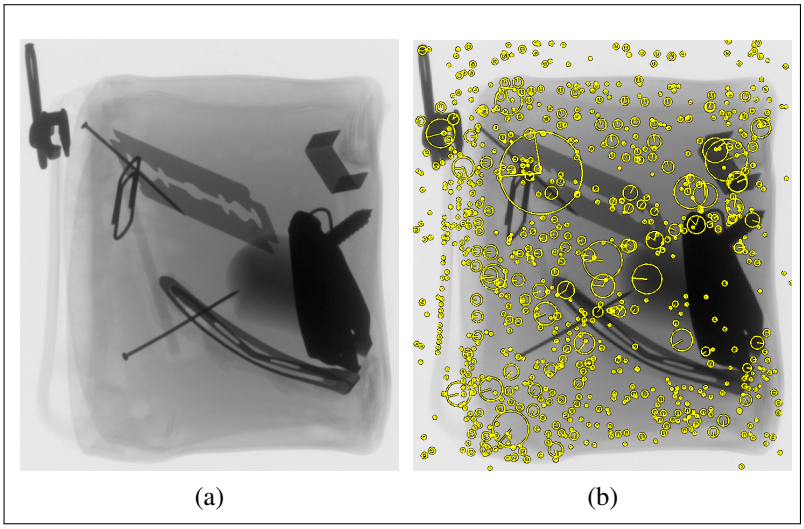


Figura 2.9. Extracción de característica SIFT, a) Imagen de Rayos X, b) Imagen de rayos X con descriptores SIFT.

2.4. Modelo Geométrico de un Sistema de Inspección con Rayos X

La imagen de rayos X de un objeto corresponde a una proyección en perspectiva, en donde un punto 3D del objeto de interés es visto como un pixel 2D en la imagen digital de Rayos X, como se muestra en la Figura 2.10. Con el fin de obtener imágenes de rayos X, que pueden ser útiles para la detección y/o la reconstrucción tridimensional, con una o múltiples vistas, se requieren algunos dispositivos, que debe ser dispuesto en una configuración especial. Esta configuración se puede caracterizar en un modelo geométrico que permita obtener la relación de un punto 3D y su proyección en una imagen como un punto 2D (mapeo 3D $\mapsto$ 2D).

A continuación describimos el modelo geométrico de nuestro sistema integrado, el cual permite adquirir imágenes de rayos X, de forma similar a lo desarrollado en la literatura (Hartley y Zisserman, 2003; Faugeras et al., 2001) y enfocados en sistemas radioscópicos (Mery y Filbert, 2002; Mery, 2003, 2014).

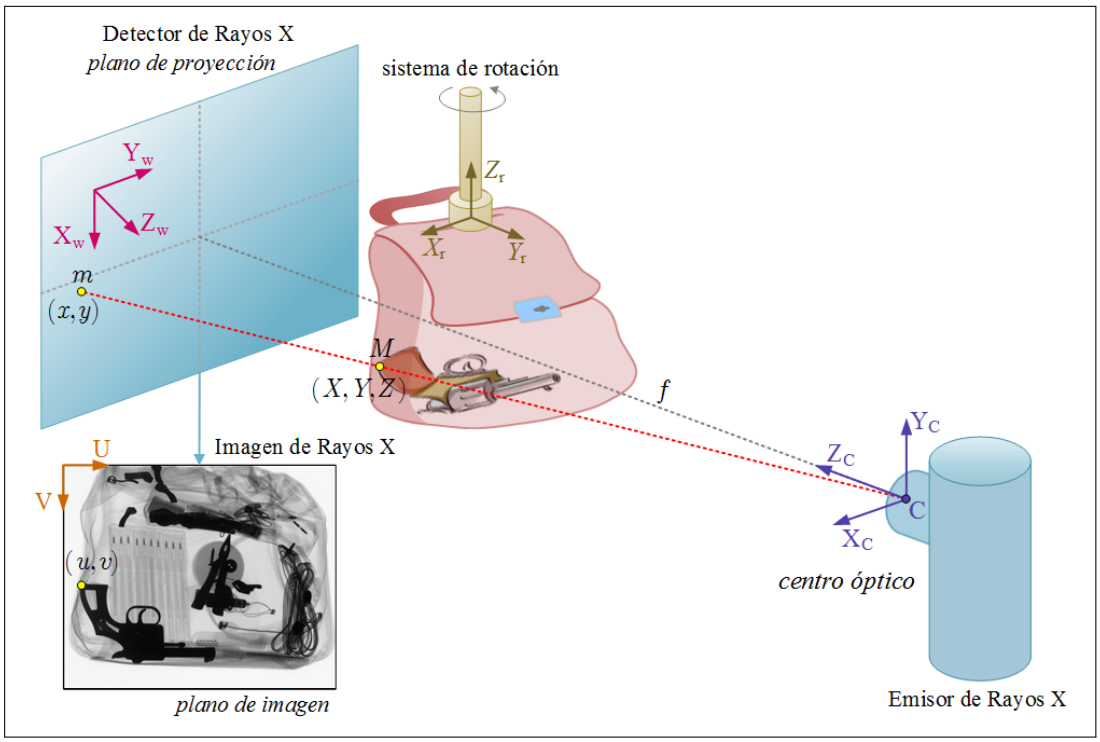


Figura 2.10. Modelo geométrico de un sistema de inspección con Rayos X.

Para el modelo geométrico propuesto, se utiliza un sistema de cinco coordenadas:

- OCS (X, Y, Z): Sistema de coordenadas del objeto, donde un punto 3D $M = (X, Y, Z)$ es definido usando coordenadas adheridas al objeto de interés.
- MCS (X', Y', Z'): Sistema de coordenadas del manipulador, $X_{\mathrm{r}}Y_{\mathrm{r}}Z_{\mathrm{r}}$, localizado en el sistema que permite rotar y trasladar al objeto sometido a inspección (en nuestro caso el sistema es un manipulador robótico).
- WCS (X'', Y'', Z''): Sistema de coordenadas del mundo, $X_{\mathrm{w}}Y_{\mathrm{w}}Z_{\mathrm{w}}$, localizado en el *plano de proyección*.
- PCS (x, y): Sistema de coordenadas de proyección, donde el punto 3D $M = (X, Y, Z)$ es proyectado sobre el *plano de proyección* y visto como un punto 2D $m = (x, y)$.
- ICS (u, v): Sistema de coordenadas de imagen, UV, localizado en el *plano de imagen* definido en pixeles, donde un punto 2D proyectado m con coordenadas (x, y) es visto en la imagen como un punto 2D con coordenadas en pixeles (u, v).

El modelo geométrico que permite mapear $(X, Y, Z) \mapsto (u, v)$, es decir, OCS $\mapsto$ ICS, puede ser expresado como un mapeo lineal en coordenadas homogéneas:

$$\lambda \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{P} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \tag{2.14}$$

donde λ es un factor escalar y $\mathbf{P}$ es la *matriz de proyección* de 3×4, la cual puede ser modelada siguiendo la siguiente transformación:

a) OCS $\mapsto$ MCS, es decir, transformación $\mathbf{T_1} : (X, Y, Z) \mapsto (X', Y', Z')$, usando una matriz de rotación 3D, $\mathbf{R_r}$ y un vector de traslación 3D, $\mathbf{t_r}$, del manipulador robótico,

b) MCS $\mapsto$ WCS, es decir, transformación $\mathbf{T_2} : (X', Y', Z') \mapsto (X'', Y'', Z'')$, usando una matriz de rotación 3D, $\mathbf{R_w}$ y un vector de traslación 3D, $\mathbf{t_w}$, del sistema de coordenadas del mundo. En este caso, $\mathbf{R_w}$ y $\mathbf{t_w}$ son los parámetros extrínsecos, los cuales son obtenidos por un proceso de calibración,

c) WCS $\mapsto$ PCS, es decir, transformación $\mathbf{T_3} : (X", Y", Z") \mapsto (x, y)$, usando una matriz de proyección que depende de la distancia focal f, y finalmente,

d) PCS $\mapsto$ ICS, es decir, transformación $\mathbf{T_4} : (x, y) \mapsto (u, v)$, usando el factor de escala α_x y α_y, y el punto principal 2D de la imagen (u_o, v_o) en pixeles (usualmente el centro de la imagen). En este caso, el factor de escala α_x y α_y, el punto principal (u_o, v_o) y también la distancia focal f, son los parámetros intrínsecos, los cuales son obtenidos directamente desde el proceso de calibración.

Las cuatro transformaciones OCS $\overset{\mathbf{T_1}}{\longmapsto}$ MCS $\overset{\mathbf{T_2}}{\longmapsto}$ WCS $\overset{\mathbf{T_3}}{\longmapsto}$ PCS $\overset{\mathbf{T_4}}{\longmapsto}$ ICS son expresadas como:

$$\mathbf{P} = \underbrace{\begin{bmatrix} \alpha_x & 0 & u_o \\ 0 & \alpha_y & v_o \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{T_4}} \underbrace{\begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\mathbf{T_3}} \underbrace{\begin{bmatrix} \mathbf{R_w} & \mathbf{t_w} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}}_{\mathbf{T_2}} \underbrace{\begin{bmatrix} \mathbf{R_r} & \mathbf{t_r} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}}_{\mathbf{T_1}}. \tag{2.15}$$

Los parámetros incluidos en la matriz $\mathbf{P}$ pueden ser estimados usando un enfoque como el descrito en Hartley y Zisserman (2003) y como se muestra en la Figura 2.11.

Para obtener las múltiples vistas del objeto, se necesitan n diferentes proyecciones del objeto sometido a inspección, lo cual se puede lograr mediante la rotación y/o traslación (para esta tarea se puede utilizar un manipulador robótico). Para la k-ésima proyección, con $k = 1 \ldots n$, el modelo geométrico $\mathbf{P}_k$ usado en la ecuación (2.14) se calcula desde la ecuación (2.15) que incluyen las matrices 3D de rotación $\mathbf{R}_k$ y traslación $\mathbf{t}_k$ asociadas a las transformaciones $\mathbf{T_1}$ y $\mathbf{T_2}$. Las matrices $\mathbf{P}_k$ pueden ser estimadas usando un patrón de calibración, proyectado en las n diferentes posiciones (Mery, 2003), o usando un algoritmo de *bundle adjustment*, donde el modelo geométrico es obtenido desde las n imágenes de rayos X del objeto inspeccionado (Mery, 2011a).

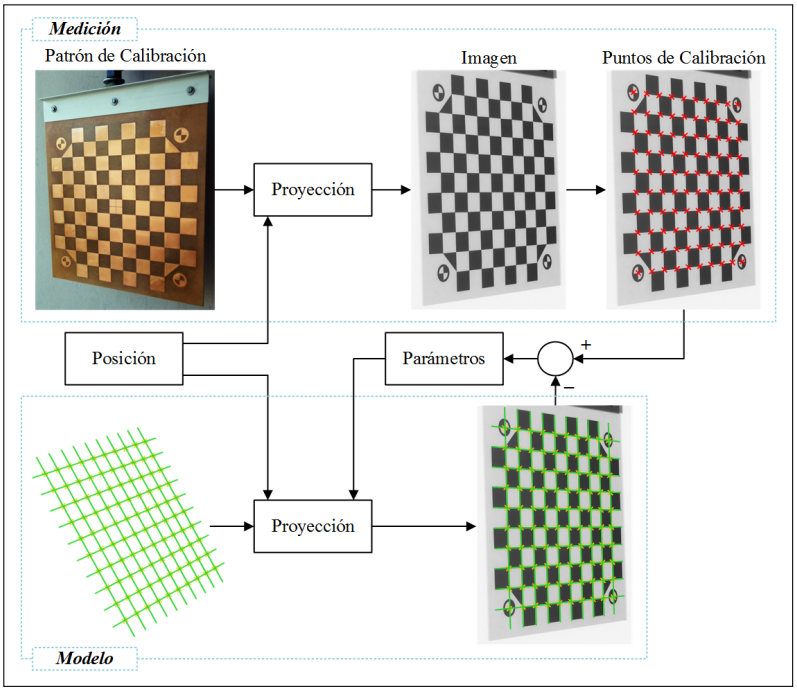


Figura 2.11. Calibración geométrica de un sistema de visión por computador para imágenes de Rayos X: la proyección de un modelo 3D coincide con la geometría real del patrón de calibración.

2.5. Correspondencia en Imágenes de Rayos X

En el análisis de imágenes de rayos X, la aplicación de algoritmos de detección de objetos, suele ser incierta cuando sólo se dispone de una imagen para dicho análisis. Esta incerteza se debe fundamentalmente a la presencia de falsos positivos (FP), o falsas alarmas. La eliminación de los falsos positivos implica realizar sintonizaciones de algunos parámetros o valores de umbral en los algoritmos detectores, lo cual no es simple de realizar, ya que existe un compromiso entre la cantidad de verdaderos positivos (VP) presentes en la imagen y la cantidad de falsos positivos que realmente se eliminarán, ya que eventualmente podríamos aumentar o disminuir demasiado estos valores de umbral que perderíamos detecciones válidas.

En el enfoque propuesto en esta tesis se hace uso de la visión activa para la detectar un objeto en la mejor vista (mejor pose visual del objeto), y cuando la detección no ocurre o

cuando la pose no es la mejor, entonces se hace necesario hacer un análisis en multiples vistas, para validar la detección y eliminar falsas alarmas. El concepto detrás de este enfoque, es que sólo las detecciones reales pueden rastrearse a lo largo de una secuencia de imágenes; un verdadero objeto detectado implica una relación espacio-temporal en las diferentes vistas en los que aparece, mientras que una falsa alarma corresponde a un evento aleatorio (Carrasco y Mery, 2011). Por esta razón, utilizamos la restricción geométrica de múltiples vista para reducir el número de búsqueda de candidatos entre detecciones monoculares. En nuestro enfoque, las restricciones geométricas se establecen a partir de la geometría bifocal (epipolar) y trifocal (Hartley y Zisserman, 2003).

Dada una secuencia de imágenes $\mathbb{I} = \{\mathbf{I}_i\}_{i=1}^{n}$, y un conjunto de puntos detectados $\mathbb{m} = \{\mathbf{m}_i\}_{i=1}^{n}$, para simplificar el análisis, sólo habrá un punto detectado por imagen (sólo un objeto de interés detectado por imagen, cuyo centroide es el punto en cuestión). Así, para un punto $\mathbf{m}_i$ en la imagen $\mathbf{I}_i$, le corresponde un punto $\mathbf{m}_j$ en la segunda imagen $\mathbf{I}_j$, el cual debe estar sobre su línea epipolar, estimada utilizando los tensores bifocales (o matriz fundamental $\mathcal{F}_{ij}$, entre las vistas i y j). Posteriormente si fuese necesario una nueva adquisición, existirá un punto en correspondencia $\mathbf{m}_k$ en la tercera imagen $\mathbf{I}_k$, punto cuya posición es estimada usando los tensores trifocales $\mathcal{T}_{ijk}$, entre las vistas i, j y k. Los tensores multifocales pueden ser estimados a partir de las matrices de proyección $\mathbb{P} = \{\mathbf{P}_i\}_{i=1}^{n}$, donde $\mathbf{P}_i$ es usada para calcular la proyección 3D de un punto $\mathbf{M}$ del objeto de interés (que debe ser detectado) en un punto 2D $\mathbf{m}_i$ de la imagen $\mathbf{I}_i$.

Para hacer la correspondencia de los puntos $\mathbb{m}$ y el seguimiento de estos a través de las imágenes $\mathbb{I}$, se debe hacer el análisis bifocal y trifocal.

A. Correspondencia Entre Dos Vistas

Dado los puntos 2D detectados $\mathbf{m}_i$ y $\mathbf{m}_j$ en las imágenes $\mathbf{I}_i$ e $\mathbf{I}_j$ respectivamente, que son debido a la proyección del punto $\mathbf{M}$ del objeto de interés en dos posiciones distintas (objeto rotado), como se muestra en la Figura 2.10, y de acuerdo al principio de geometría

de multiples vistas (Hartley y Zisserman, 2003), los puntos $\mathbf{m}_i$ y $\mathbf{m}_j$ están en correspondencia si existe la matriz fundamental $\mathcal{F}_{ij}$, tal que se cumpla:

$$\mathbf{m}_j^\top \mathcal{F}_{ij} \mathbf{m}_i = 0 \tag{2.16}$$

donde, $\mathcal{F}_{ij}$ es la matriz fundamental, que depende sólo de las matrices de proyección $\mathbf{P}_i$ y $\mathbf{P}_j$.

La restricción epipolar señala que para que $\mathbf{m}_i$ y $\mathbf{m}_j$ sean puntos correspondientes, el punto $\mathbf{m}_j$ debe estar en la línea epipolar generada por $\mathbf{m}_i$. Esto no quiere decir que todos los puntos en la línea epipolar generada por $\mathbf{m}_i$ son correspondientes a $\mathbf{m}_i$, ya que sólo un punto en la imagen $\mathbf{I}_j$ es correspondiente a $\mathbf{m}_i$, y en este caso es la proyección de $\mathbf{M}$ en la imagen $\mathbf{I}_j$. La restricción epipolar es entonces una condición necesaria, pero no suficiente. A pesar de que no sea una condición suficiente, es de gran utilidad saber que el punto correspondiente a $\mathbf{m}_i$ en la imagen $\mathbf{I}_j$ está sobre una línea y no está ubicado en cualquier parte de la imagen. Esto representa una reducción considerable en la dimensionalidad del problema de búsqueda de puntos correspondientes, ya que en vez de buscar en toda la imagen $\mathbf{I}_j$ (de dos dimensiones) se busca sólo a lo largo de una línea (una dimensión). Las líneas epipolares l_i y l_j se determinan mediante las siguientes ecuaciones:

$$l_j = \mathcal{F}_{ij} \mathbf{m}_i = [a_1 \; a_2 \; a_3]^\top \tag{2.17}$$

$$l_i = \mathcal{F}_{ij}^\top \mathbf{m}_j = [b_1 \; b_2 \; b_3]^\top \tag{2.18}$$

donde, $[a_1 \; a_2 \; a_3]$ y $[b_1 \; b_2 \; b_3]$ son los coeficientes de las líneas epipolares l_j y l_i respectivamente.

En la práctica $\mathbf{m}_j$ no está exactamente sobre la línea epipolar l_j, sino que está muy cerca. Por esta razón es necesario utilizar otro criterio de correspondencia; se dice que $\mathbf{m}_i$ y $\mathbf{m}_j$ pueden ser puntos correspondientes si la distancia mínima de $\mathbf{m}_j$ a l_j es menor que una distancia d_0. Esta distancia se calcula a partir de una línea perpendicular a la que pase por $\mathbf{m}_j$.

De esta manera, se obtiene que la restricción epipolar práctica se expresa como:

$$\frac{|\mathbf{m}_j^\top \mathcal{F}_{ij} \mathbf{m}_i|}{\sqrt{a_1^2 + a_2^2}} < d_0 \tag{2.19}$$

Restricción que podemos representar gráficamente, tal como se muestra en la Figura 2.12.

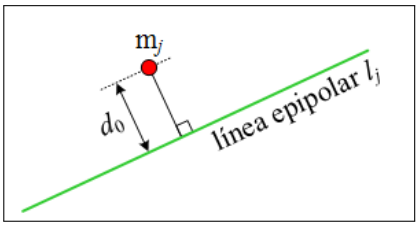


Figura 2.12. Restricción epipolar práctica: distancia del punto $\mathbf{m}_j$ a la línea epipolar l_j.

B. Correspondencia Entre Tres Vistas

De forma similar al caso de dos vistas, en la correspondencia en tres vistas se debe disponer de los puntos 2D detectados $\mathbf{m}_i$, $\mathbf{m}_j$ y $\mathbf{m}_k$ en las imágenes $\mathbf{I}_i$, $\mathbf{I}_j$ e $\mathbf{I}_k$ respectivamente, que son debido a la proyección del punto $\mathbf{M}$ del objeto de interés en tres posiciones distintas (objeto rotado), como se muestra en la Figura 2.10, a partir de esto podemos aplicar el análisis trifocal que permite modelar todas las relaciones geométricas existentes en tres vista, y es independiente de la estructura contenida en cada imagen (Hartley y Zisserman, 2003). El tensor, tiene una estructura de matriz similar a la matriz fundamental, sólo depende del movimiento entre las imágenes y de los parámetros intrínsecos del sistema. Su principal ventaja es que puede ser calculado a partir de las correspondencias de las imágenes sin ningún conocimiento a priori del movimiento o de calibración del objeto. Esta característica justifica el porque la estimación de la matriz fundamental no siempre eliminar todos los falsos positivos.

Basado en los resultados del análisis en dos vistas, para confirmar que una correspondencia bifocal efectivamente representa una detección válida, tratamos de descubrir una nueva correspondencia utilizando una tercera vista, con la ayuda de los tensores trifocales.

Sea el tensor trifocal $\mathbf{T}^{jk} = [T_1^{jk}\ T_2^{jk}\ T_3^{jk}]^\top$ para $j, k = 1, 2, 3$ que codifica el movimiento relativo entre las tres vistas y que puede ser calculado a partir de las matrices de proyección $\mathbb{P}$, como lo propone Hartley y Zisserman (2003). Luego, se puede estimar la posición hipotética $\hat{\mathbf{m}}_k$ de una detección en un tercera vista k usando las correspondencias $\mathbf{m}_i$, $\mathbf{m}_j$ y el tensor $\mathbf{T}^{jk}$ como:

$$\hat{\mathbf{m}}_k = \begin{bmatrix} \hat{x}_k \\ \hat{y}_k \\ 1 \end{bmatrix} = \frac{1}{\mathbf{m}_i^\top(\mathbf{T}^{13} - x_2\mathbf{T}^{33})} \begin{bmatrix} \mathbf{m}_i^\top(\mathbf{T}^{11} - x_2\mathbf{T}^{31}) \\ \mathbf{m}_i^\top(\mathbf{T}^{12} - x_2\mathbf{T}^{32}) \\ \mathbf{m}_i^\top(\mathbf{T}^{13} - x_2\mathbf{T}^{33}) \end{bmatrix}. \tag{2.20}$$

Comparamos la posición estimada $\hat{\mathbf{m}}_k$ con todas las potenciales detecciones $\mathbf{m}_k$ de la vista k. Así, se considerará a la potencial detección k como una verdadera detección si:

$$\|\hat{\mathbf{m}}_k - \mathbf{m}_k\| < d_1, \qquad \text{para pequeños valores de } d_1 > 0. \tag{2.21}$$

Si se cumple esta restricción, tomamos la potencial detección como una detección real, puesto que cumple con la correspondencia en tres vistas. En el caso de que la potencial detección en la tercera vista no coinciden con la proyección del tensor, se descarta, ya que no cumple con la restricción trifocal. En general, dado que la restricción trifocal se analiza para las secuencias que cumplen la condición bifocal, se reduce el número de falsos positivos generados en dos vistas.

2.6. Búsqueda de la Siguiente Mejor Vista

La visión activa en los procesos de inspección con rayos X requiere de una estrategia de búsqueda que permita encontrar un objeto de interés en la mejor vista posible. La búsqueda consiste en determinar el siguiente movimiento que debe realizar un manipulador robótico o sistema de sujeción, es decir, una vez detectado el objeto de interés y estimada su pose, se decide o no realizar un movimiento, siempre y cuando la pose estimada no sea la visualmente adecuada (mala pose). Las poses adecuadas o mejores poses son definidas como un conocimiento a priori, y el objeto sometido a inspección deberá ser rotado hasta

detectar al objeto de interés en una de estas poses previamente definidas (Riffo y Mery, 2012). En esta propuesta se utilizan dos estrategias de búsqueda de la siguiente pose, es decir como realizar el movimiento del manipulador; *i*) **Mediante estimación de tipo heurística**: basada en la forma de los objetos para realizar el movimiento y obtener una mejor pose, aquí, en cada movimiento realizado al objeto que contiene al objeto de interés, se aplica un detector, y una vez detectado el objeto de interés, se realiza un pareo (*matching*) de descriptores SIFT encontrados en la imagen, con los descriptores almacenados en una base de datos, en donde la mayor cantidad de coincidencias indicaran la pose del objeto y el ángulo de rotación del manipulador cuando la pose detectada no es la mejor. Una explicación más detallada de este método se pueden encontrar en la sección 4.1.5. Un segundo método es el *ii*) **Aprendizaje por refuerzo Q-Learning**: que es un método más eficiente para la estimación de movimiento, y así lograr una mejor vista del objeto de interés en la menor cantidad de movimientos y dotando a la metodología de búsqueda y detección de una *pseudo inteligencia* (C. J. Watkins y Dayan, 1992).

El aprendizaje por refuerzo, esquematizado en la Figura 2.13, consiste en aprender qué acciones realizar, dado el estado actual del ambiente, con el objetivo de maximizar una señal de recompensa numérica (Sutton y Barto, 1998). En el aprendizaje por refuerzo la única información de retroalimentación es un escalar conocido como recompensa, donde dicha recompensa la recibe un agente posteriormente a la ejecución de una acción. Una característica importante de este tipo de aprendizaje, es que al final lo que se busca es una función que indique cual es la mejor secuencia de acciones necesarias para alcanzar una meta predefinida.

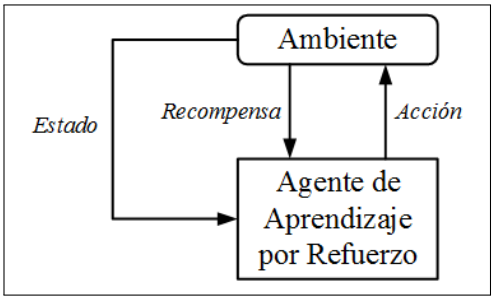


Figura 2.13. Esquema de aprendizaje por refuerzo.

Dentro del aprendizaje por refuerzo existen varias técnicas o algoritmos como *Q-Learning*, el cual es un algoritmo de aprendizaje automático, desarrollado para resolver tareas que pueden ser modeladas como procesos de decisión Markovianos (o MDP: Markov Decision Processes). Los MDP, también conocidos como programas dinámicos estocásticos o problemas de control estocástico, son modelos usados para la toma de decisiones en secuencia, cuando las salidas del sistema controlado son inciertas (Puterman, 2014). *Q-Learning* fue propuesto en (C. J. C. H. Watkins, 1989) para resolver MDP's con información incompleta. En *Q-learning* un agente intenta aprender la política óptima de su historia de interacciones con el medio ambiente. Una historia de un agente es una secuencia de estado-acción-recompensas: $\langle s_0, a_0, r_1, s_1, a_1, r_2, s_2, a_2, r_3, s_3, a_3, r_4, s_4 \ldots \rangle$, lo cual significa que el agente estuvo en el estado s_0 e hizo la acción a_0, lo que resultó en la recepción de la recompensa r_1 y estar en el estado s_1; luego se hizo una acción a_1, recibió la recompensa r_2, y terminó en el estado s_2; luego hizo la acción a_2, recibió r_3 de recompensa, y terminó en el estado s_3; y así sucesivamente. Trataremos a esta historia de interacción como una secuencia de experiencias, donde la experiencia es una tupla de la forma $\langle s, a, r, s' \rangle$, lo que significa que el agente estuvo en el estado s, se hizo una acción a, recibió la recompensa r, y que entró en el estado s'. Estas experiencias serán los datos de los que el agente puede aprender qué hacer. Así, el objetivo del agente es maximizar su valor, que suele ser la recompensa con descuento.

Dado $Q^*(s, a)$, que es el valor esperado (recompensa con descuento acumulativo) de hacer a en el estado s y siguiendo la política óptima. En *Q-learning* se utilizan diferencias temporales para estimar el valor de $Q^*(s, a)$. En *Q-learning*, el agente mantiene una tabla de $Q[S, A]$, donde S es el conjunto de estados y A es el conjunto de acciones. $Q[s, a]$ representa la estimación actual de $Q^*(s, a)$.

Una tupla $\langle s, a, r, s' \rangle$ proporciona un punto de datos para el valor de $Q(s, a)$. El punto de datos denomina *retorno* es el valor futuro $r + \eta V(s')$ que el agente recibió, donde $V(s') = \max_{a'} Q(s', a')$ es la recompensa real actual más el valor futuro estimado con descuento y η es un número entre 0 y 1 $(0 \leq \eta \leq 1)$ llamado *factor de descuento*. El

agente puede actualizar su estimación de $Q(s, a)$ como:

$$Q[s, a] \leftarrow Q[s, a] + \alpha \left(r + \eta \max_{a'} Q[s', a'] - Q[s, a] \right), \tag{2.22}$$

o su equivalente:

$$Q[s, a] \leftarrow (1 - \alpha)\, Q[s, a] + \alpha \left(r + \eta \max_{a'} Q[s', a'] \right). \tag{2.23}$$

donde α es la *tasa de aprendizaje* $(0 \leq \alpha < 1)$. Esto supone que α es fijo; si α es variable, habrá un conteo diferente para cada par estado-acción y el algoritmo también tendría que realizar un seguimiento de este conteo.

Q-learning aprende una política óptima, no importando cual política actualmente el agente está siguiendo, es decir, cual acción a se selecciona para cualquier estado s, siempre y cuando no haya ningún límite en el número de veces que se intenta una acción en cualquier estado (no siempre se hace el mismo subconjunto de acciones en un estado). A este método se le denomina *método fuera de la política* (off-policy method), debido a que este método aprende una política óptima y no importa la política que está llevando a cabo (Poole y Mackworth, 2010).

2.7. Modelo de Forma Implícita Para la Detección de Objetos

El concepto *modelo de forma implícita* (ISM: Implicit Shape Model) fue originalmente propuesto por Leibe et al. (2004) y ha mostrado buenos resultados en la detección de una serie de categorías de objetos, como por ejemplo, automóviles, vacas y motocicletas. Como el nombre ya sugiere, no se trata de definir un modelo explícito para todas las formas posibles que puede tomar una clase de objeto, pero en su lugar se definen formas permitidas implícitamente en término de los cuales las apariciones locales son coherentes entre sí. A diferencia de otros enfoques, el ISM original considera la detección de objetos y segmentación de éste del fondo, como dos procesos entrelazados, que colaboran estrechamente. El enfoque se basa en un modelo de objetos generativo con una representación flexible del objeto, en forma de estrella. Para ello, la disposición espacial de los descriptores locales

(características) de la imagen se aprende con respecto al centro del objeto. Teniendo en cuenta que la configuración espacial de los descriptores locales puede diferir significativamente entre los miembros de la categoría. En consecuencia, los descriptores locales de las imágenes de diferentes ejemplos de entrenamiento se pueden combinar con el fin de generalizar a toda la categoría del objeto, a partir de los ejemplos observados.

En esta sección, vamos a explicar los aspectos más relevantes del modelo de forma implícita, como originalmente fue descrito por Leibe et al. (2004). En el capítulo 4 explicaremos los detalles del algoritmo y las adaptaciones al modelo, para aplicarlo a la tarea de detección de objetos amenazantes en imágenes de rayos X.

A. Representación de Formas

Como representación básica del enfoque, se introduce el modelo de forma implícita $ISM(C) = (\mathbf{C}, \mathcal{P}_\mathbf{C})$, que consiste en un alfabeto $\mathbf{C}$ de clase específica (el vocabulario visual: *codebook*) de las apariencias locales que son representativas para la categoría C del objeto (ver Figura 2.14), y de una distribución de probabilidad espacial $\mathcal{P}_\mathbf{C}$, que especifica el lugar donde cada entrada del vocabulario visual puede ser encontrado en el objeto.

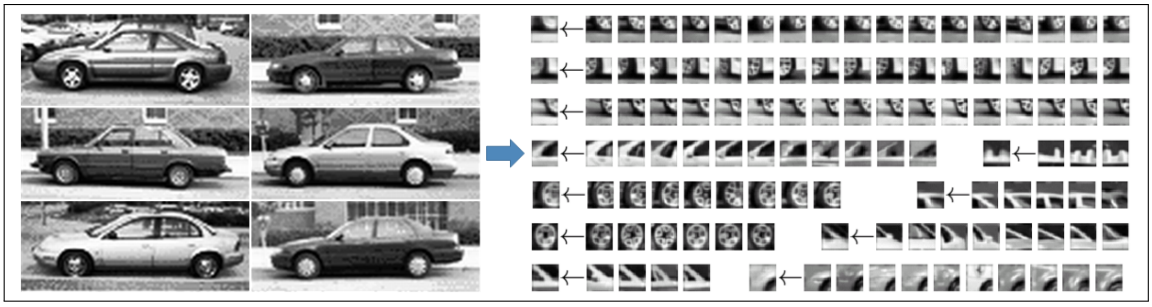


Figura 2.14. Generación de un vocabulario visual de apariencias a partir de imágenes de entrenamiento. Adaptada desde (Leibe y Schiele, 2003).

Se tienen dos opciones explícitas de diseño para la distribución de probabilidad $\mathcal{P}_\mathbf{C}$. *La primera* es que la distribución se define independientemente para cada entrada del vocabulario visual. Esto hace que el enfoque sea flexible, ya que permite combinar partes del objeto durante el reconocimiento, que inicialmente se observaron, en diferentes ejemplos

de entrenamiento. Además, permite aprender modelos de reconocimiento desde relativamente pequeños conjuntos de entrenamiento. *La segunda* limitación, es que la distribución de probabilidad espacial para cada entrada del vocabulario visual, se calcula de una manera no paramétrica. Esto permite que el método para modelar la verdadera distribución sea más detallado, según lo permitan los datos de entrenamiento, en vez de asumir una distribución gaussiana simplificada.

B. Modelo de Aprendizaje de la Forma

Dado el vocabulario visual de apariencia aprendido $\mathbf{C}$, el primer paso es aprender la distribución de probabilidad espacial $\mathcal{P}_{\mathbf{C}}$. Para ello, realizamos una segunda iteración sobre todas las imágenes de entrenamiento y pareamos (matching) las entradas del vocabulario visual con las imágenes (usando una medida de similaridad). Aquí, no sólo activamos el mejor pareo con el vocabulario visual, sino que todas las entradas de dicho vovabulario, cuya similitud sea superior a un valor de umbral θ_u, usado en el proceso formación del vocabulario visual. Para todas las entradas del vocabulario visual, almacenamos todas las posiciones que se activaron, en relación con el centro del objeto.

En este paso, se modela la incertidumbre en el proceso de generación del vocabulario. Si un vocabulario visual es "perfecto" en el sentido de que cada descriptor se puede asignar de forma única a exactamente un conjunto (cluster), entonces el resultado es equivalente a una estrategia de búsqueda del vecino más cercano. Sin embargo, en aplicaciones prácticas, es poco realista esperar que los datos sean limpios. Por lo tanto, mantenemos cada posible asignación, pero con su peso, con la probabilidad de que esta asignación sea la correcta. Es fácil ver que para las puntuaciones de similitud más pequeñas que θ_u, la probabilidad de que este descriptor haya sido asignado al conjunto durante el proceso de generación del vocabulario visual es cero; por lo tanto, no necesitamos considerar esos pareos. Las ubicaciones de la ocurrencia almacenadas, por otra parte, reflejan la distribución espacial de una entrada del vocabulario visual sobre la zona del objeto, en una forma no paramétrica.

C. Enfoque de Reconocimiento

Dada una imagen nueva de prueba, una vez más aplicamos un detector de punto de interés y extraemos descriptores alrededor de los lugares seleccionados. Los descriptores extraídos son entonces pareados con el vocabulario visual para activar las entradas de dicho vocabulario, utilizando el mismo mecanismo que señalamos anteriormente, es decir, usando una medida de similaridad. Desde el conjunto de todos los descriptores pareados, seleccionamos configuraciones consistentes, realizando una Transformada generalizada de Hough (Hough, 1962; Ballard, 1981). Cada entrada activada emite votos para posibles posiciones del centro del objeto, de acuerdo con la distribución espacial aprendida $\mathcal{P}_{\mathrm{C}}$. Las hipótesis consistentes son luego buscadas como máximos locales en el espacio de votación. En el ejercicio de este enfoque, es importante evitar cuantificaciones. A diferencia de la práctica habitual (por ejemplo, (Lowe, 1999)), no discretizamos los votos, pero mantenemos los valores continuos originales. La máxima en este espacio continuo se puede encontrar con precisión y eficiencia utilizando el modo de estimación Mean-Shift (Cheng, 1995; Comaniciu et al., 2001). Una vez que una hipótesis ha sido seleccionada, se reúnen todos los descriptores que contribuyeron a ella (ver Figura 2.15), de esta manera se visualiza a lo que el sistema reacciona.

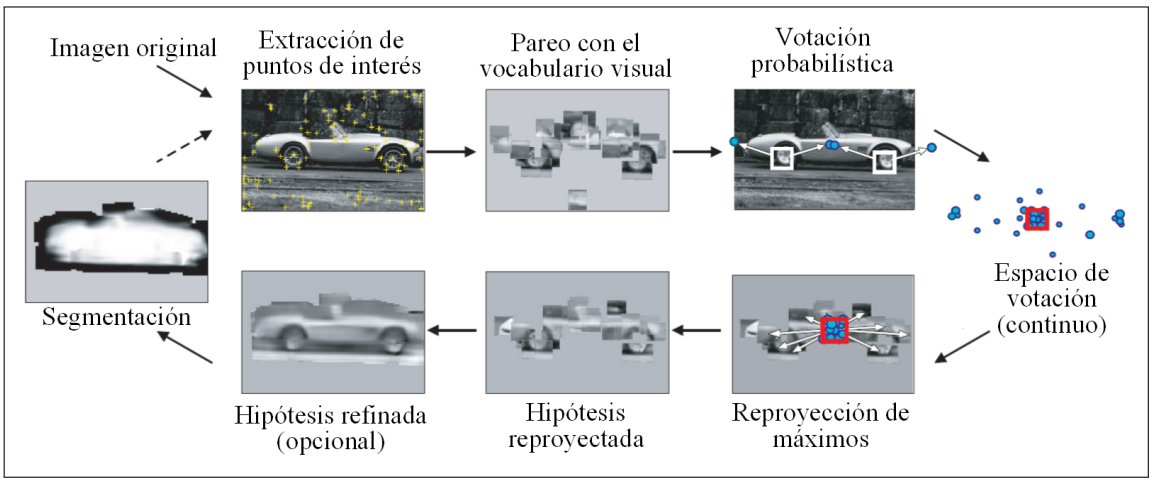


Figura 2.15. Proceso para reconocer objetos usando ISM original. El proceso comienza arriba a la derecha y continua en sentido de las manecillas del reloj, (Leibe y Schiele, 2003).

Como resultado, se obtiene una representación del objeto que incluye un área determinada. Esta representación opcionalmente se puede refinar aún más, mediante el muestreo más localizado de las características. La repuesta retroalimentada, servirá posteriormente de base para el cálculo de una segmentación de categoría de objeto especifica.

Capítulo 3. ESTADO DEL ARTE

Como ya hemos dicho, los rayos X han sido desarrollados para ser usados en imágenes médicas de seres humanos. Sin embargo, este campo de utilización ha sido ampliado a ensayos no destructivos (NDT) para los materiales u objetos, donde las partes interiores, que son indetectables a simple vista, son analizadas *no destructivamente* con rayos X. Las pruebas o ensayos con rayos X se utilizan en muchas aplicaciones, tales como: análisis de productos alimenticios, inspección de equipaje, inspección de partes y piezas de automóviles, y control de calidad de soldaduras, entre otras. En este capítulo presentamos el estado del arte y una mirada general de las metodologías de visión por computador que se han utilizado en la inspección con rayos X.

Los ensayos no destructivos con rayos X y que usan visión por computador se presentan como un modelo general, de acuerdo con la Figura 3.1. Estas aplicaciones siguen este esquema, y dependiendo de la forma en que las imágenes de rayos X son adquiridos y analizadas, cada bloque puede (o no puede) ser utilizado. Por ejemplo, en la inspección de equipaje hay métodos que analizan una o múltiples vistas, utilizando sistemas mono-energía o dual-energía. Para cada una de estas cuatro combinaciones, solamente los bloques correspondientes de la Figura 3.1 se van a utilizar.

A continuación describiremos el estado del arte de las aplicaciones más relevantes en la inspección no destructiva con rayos X. Este capítulo cubre la inspección con rayos X en:

a) Fundición,
b) Soldadura,
c) Equipaje,
d) Alimentos, y
e) Cargamento.

Enfoques de cada aplicación se resumen en las Tablas 3.1, . . . , 3.5, mostrando cómo usan los diferentes pasos del esquema general de acuerdo con la Figura 3.1.

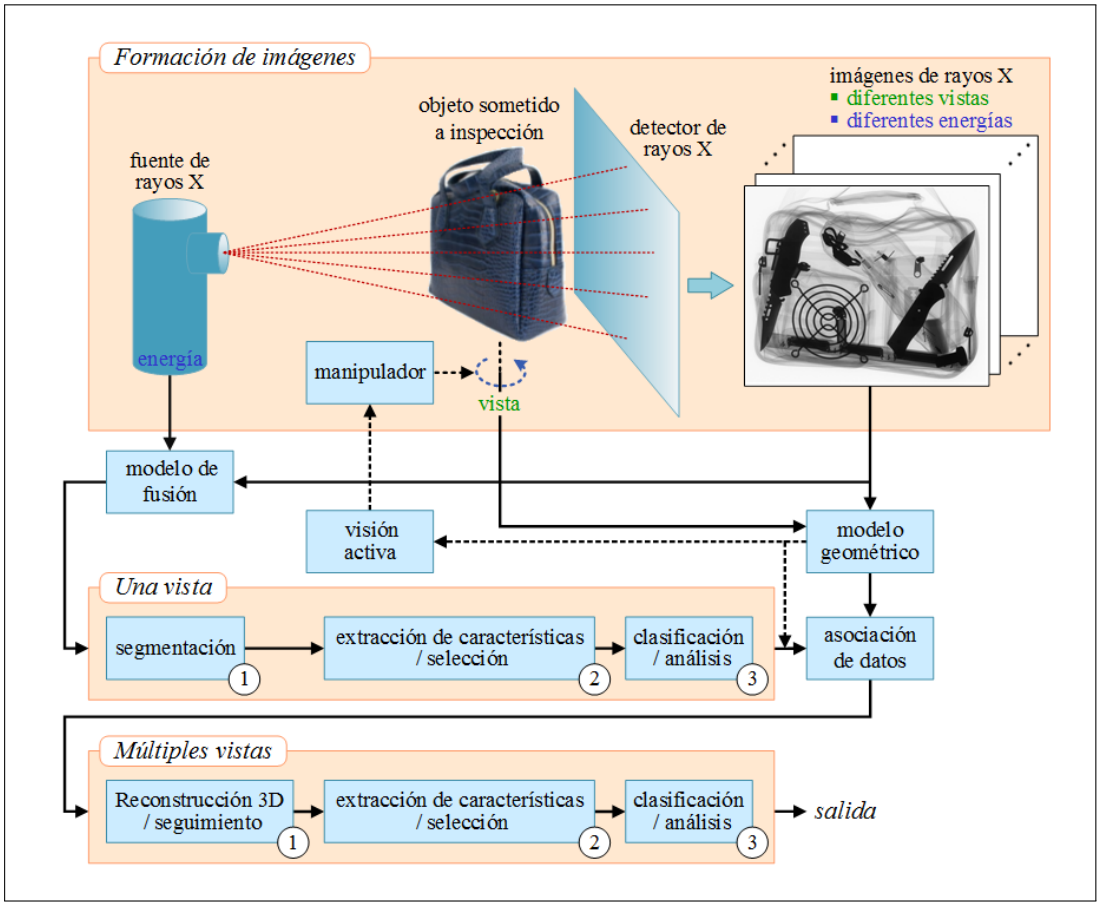


Figura 3.1. Esquema general para los ensayos no destructivos con rayos X, utilizando visión por computador. Las imágenes de rayos X de un objeto de prueba, pueden ser generadas en diferentes posiciones y diferentes niveles de energía. Dependiendo de la aplicación, cada bloque de este diagrama puede (o no puede) ser utilizado.

3.1. Fundición

Las piezas fundidas de aleación ligera producidas para la industria automotriz, tales como llantas, pivotes, rótulas de dirección, y cajas de cambio, se consideran componentes importantes para la inspección técnica general de vehículos. Regiones no-homogéneas se pueden formar dentro de la pieza de trabajo en el proceso de producción. Esto se manifiesta, por ejemplo, por vacíos en forma de burbuja, fracturas, inclusiones o formación de escoria. Para garantizar la seguridad de la piezas o partes construida, es necesario comprobar minuciosamente cada una de ellas, mediante el uso de la inspección con rayos X. En la inspección de piezas fundidas, los sistemas automatizados de rayos X, no sólo han elevado

la calidad, a través de inspecciones objetivas, repetitivas y procesos mejorados, ya que también han aumentado la productividad y la coherencia mediante la reducción de los costos laborales. Un ejemplo se ilustra en la Figura 3.2. Un estudio de estas aplicaciones se puede encontrar en (Mery, 2006). Hemos seleccionado distintos enfoque, los cuales resumimos en la Tabla 3.1. Llegamos a la conclusión de que los sistemas automatizados son muy eficaces en esta área, porque la tarea de inspección es rápida y logra un alto desempeño.

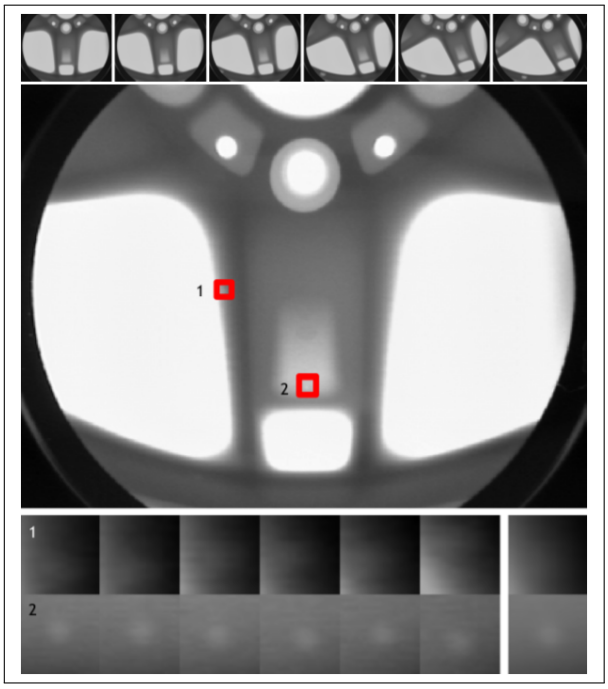


Figura 3.2. Detección de pequeñas fallas en llantas de aluminio, usando múltiples vistas, (Mery, 2011a).

Tabla 3.1. Aplicación de los rayos X en inspección de piezas metálicas fundidas.

Referencia	Energía		Modelo	Una Vista(**)			Visión	Múltiples Vistas(**)		
	Mono	Dual	Geométrico(*)	①	②	③	Activa	①	②	③
Carrasco y Mery (2011)	✓	–	N	✓	✓	✓	–	✓	✓	✓
Li et al. (2006)	✓	–	–	✓	✓	–	–	–	–	–
Mery y Filbert (2002)	✓	–	C	✓	✓	✓	–	✓	✓	✓
Mery (2011a)	✓	–	N	✓	✓	✓	–	✓	✓	✓
Pieringer y Mery (2010)	✓	–	C	✓	✓	✓	–	✓	✓	✓
Pizarro et al. (2008)	✓	–	N	✓	✓	✓	–	✓	✓	✓
Tang et al. (2009)	✓	–	–	✓	–	–	–	–	–	–

(*) C: Calibrado, N: No calibrado, –: no usado. (**) Ver ①, ②, ③ en Figura 3.1.

3.2. Soldadura

En el proceso de soldadura de alta calidad, se requiere obligatoriamente la inspección mediante ensayos no destructivos (NDT) con rayos X, con el fin de detectar defectos tales como; porosidad, inclusión, falta de fusión, falta de penetración, y grietas (ver Figura 3.3).

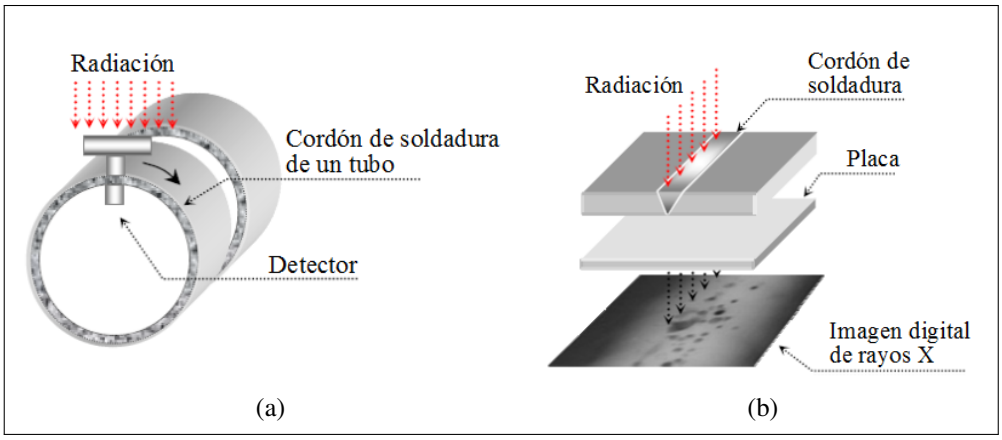


Figura 3.3. Inspección de soldadura, a) Esquema de detección de defectos a través de la radiación de la soldadura, b) Detección de defectos en la soldadura, irradiando y capturando la imagen en una placa radiográfica. (Carrasco y Mery, 2006).

Las imágenes de rayos X industriales de soldaduras son ampliamente utilizadas para la detección de defectos en la industria: petrolera, química, nuclear, naval, aeronáutica, y de construcción civil, entre otras. Un ejemplo de la utilización de algoritmos de procesamiento de imágenes y visión por computador para la inspección de soldaduras, se ilustra en la Figura 3.4. Otras de estas aplicaciones se puede encontrar en (Silva y Mery, 2007a,b).

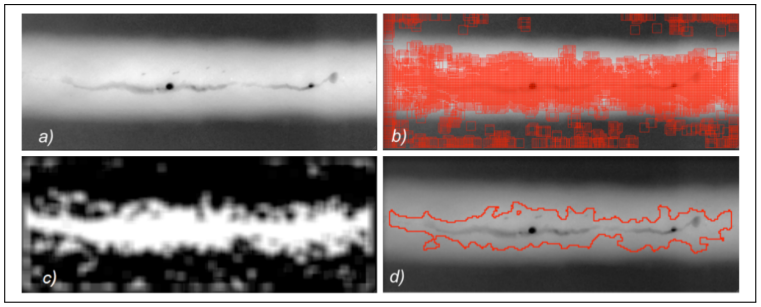


Figura 3.4. Inspección de soldadura usando ventana deslizante: a) imagen de rayos X, b) ventanas detectadas, c) mapa de activación, d) detección, (Mery, 2011b).

Hemos seleccionado distintos enfoque, los cuales resumimos en la Tabla 3.2. Como se puede ver, hay mucha investigación sobre la inspección de soldaduras. El desempeño alcanzado de los algoritmos desarrollados todavía no es lo suficientemente alto, por lo cual no son adecuados para la inspección completamente automatizada.

Tabla 3.2. Aplicación de los rayos X en inspección de soldadura.

Referencia	Energía		Modelo	Una Vista(**)			Visión	Múltiples Vistas(**)		
	Mono	Dual	Geométrico(*)	①	②	③	Activa	①	②	③
Liao (2008)	✓	–	–	✓	✓	✓	–	–	–	–
Liao (2009)	✓	–	–	✓	✓	✓	–	–	–	–
Mery y Berti (2003)	✓	–	–	✓	✓	✓	–	–	–	–
Mery (2011b)	✓	–	–	✓	✓	✓	–	–	–	–
Vilar et al. (2009)	✓	–	–	✓	✓	✓	–	–	–	–
Wang et al. (2008)	✓	–	–	✓	✓	✓	–	–	–	–
Shi et al. (2007)	✓	–	–	✓	✓	✓	–	–	–	–

(*) C: Calibrado, N: No calibrado, –: no usado. (**) Ver ①, ②, ③ en Figura 3.1.

3.3. Equipaje

Desde los ataques terroristas del 11 de septiembre del 2011 en los Estados Unidos, los sistemas automáticos (o semi-automáticos) de detección de objetos en el equipaje, usando imágenes de rayos X, se han convertido en un tema muy importante (Michel et al., 2008; Zentai, 2008). El proceso de inspección es muy complejo porque los elementos amenazantes son muy difíciles de detectar cuando se coloca en bolsos compactos cerrados, superpuestos por otros objetos y/o girados, mostrando una vista irreconocible. Sin embargo, durante la última década, la investigación pertinente se ha llevado a cabo usando el análisis en múltiples vista y la adquisición de imágenes con dos niveles de energía (energía dual). El uso de información extraída desde múltiples vista produce una mejora significativa en el desempeño, debido a que ciertos objetos son difíciles de reconocer usando sólo un punto de vista (von Bastian et al., 2010). Por otro lado, mediante el uso de energía dual es posible identificar la composición del material, típicamente para explosivos, drogas y materiales orgánicos (Singh y Singh, 2003).

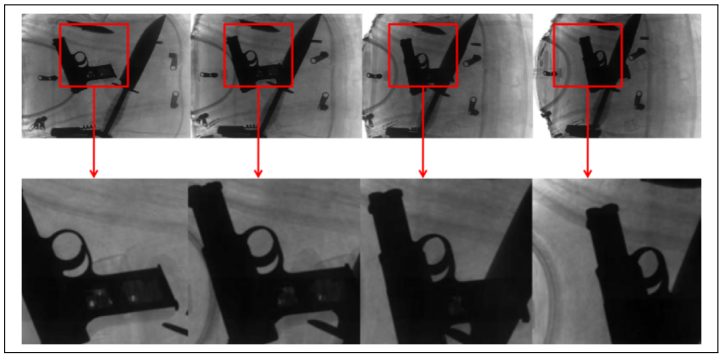


Figura 3.5. Detección en múltiples vistas de una pistola, basada en la identificación del gatillo, (Mery, Mondragon, et al., 2013).

Un ejemplo del uso de multiples vistas se ilustra en la Figura 3.5 y un estudio de la detección de explosivos se puede encontrar en (Singh y Singh, 2003; Wells y Bradley, 2012), además, distintos enfoque son resumidos en la Tabla 3.3. Podemos concluir que en la inspección de equipaje, donde la seguridad humana juega un papel importante y la complejidad de inspección es muy alta, los inspectores humanos siguen siendo necesarios. Para condiciones intrincados, se requiere de inspección con rayos X en múltiples vistas y el uso de energía dual.

Tabla 3.3. Aplicación de los rayos X en inspección de equipaje.

Referencia	Energía		Modelo	Una Vista(**)			Visión	Múltiples Vistas(**)		
	Mono	Dual	Geométrico(*)	①	②	③	Activa	①	②	③
Abusaeeda et al. (2011)	✓	✓	C	–	✓	–	–	–	✓	–
Baştan et al. (2011)	✓	✓	–	✓	✓	✓	–	–	–	–
Baştan et al. (2013)	✓	✓	–	✓	✓	✓	–	✓	✓	✓
Chen et al. (2005)	✓	✓	–	✓	✓	–	–	–	–	–
Ding et al. (2006)	✓	✓	–	✓	✓	✓	–	–	–	–
Franzel et al. (2012)	✓	✓	C	✓	✓	✓	–	✓	✓	✓
Heitz y Chechik (2010)	✓	✓	–	✓	✓	✓	–	✓	✓	✓
Mansoor y Rajashankari (2012)	✓	✓	–	✓	✓	✓	–	–	–	–
Mery (2011a)	✓	–	N	✓	✓	✓	–	✓	✓	✓
Mery, Mondragon, et al. (2013)	✓	–	N	✓	✓	✓	–	✓	✓	✓
Mery, Riffo, et al. (2013)	✓	–	N	✓	✓	✓	–	✓	✓	✓
Lu y Conners (2006)	✓	✓	–	✓	✓	–	–	–	–	–
Riffo y Mery (2012)	✓	–	N	✓	✓	–	✓	✓	✓	✓
Riffo y Mery (2016)	✓	–	N	–	✓	✓	–	–	–	–
Schmidt-Hackenberg et al. (2012)	✓	✓	–	✓	✓	✓	–	–	–	–
Turcsany et al. (2013)	✓	✓	–	✓	✓	✓	–	–	–	–

(*) C: Calibrado, N: No calibrado, –: no usado.

(**) Ver ①, ②, ③ en Figura 3.1.

3.4. Alimentos

Para la industria alimentaria se han desarrollado varias aplicaciones que garantizan la inspección segura de los alimentos. Las dificultades inherentes a la detección de defectos y contaminantes en los productos alimenticios envasados, han limitado el uso de rayos X. Sin embargo, la necesidad de inspección no destructiva (NDT), ha motivado un esfuerzo considerable de investigación en este campo, que abarca muchas décadas (Haff y Toyofuku, 2008).

Los avances más importantes son: detección de objetos extraños en los alimentos envasados (Kwon et al., 2008); detección de espinas en el pescado (Mery et al., 2011); identificación de plagas de insectos en los cítricos (Jiang et al., 2008); detección de larvas de polilla en las manzanas (Haff y Toyofuku, 2008); inspección de calidad en frutas, tal como, detección de cuescos partidos, distribución del contenido de agua y de la estructura interna (Ogawa et al., 2003); y la detección de fases larvarias del gorgojo granero en semillas de trigo (Haff y Slaughter, 2004). En estas aplicaciones, sólo se requiere el análisis de una sola vista. Un ejemplo que muestra la utilización de los rayos X para detectar espinas en filetes de spescado (salmón) se ilustra en la Figura 3.6. Un estudio se puede encontrar en (Haff y Toyofuku, 2008). Finalmente, en la Tabla 3.4, se resumen las aplicaciones mencionadas.

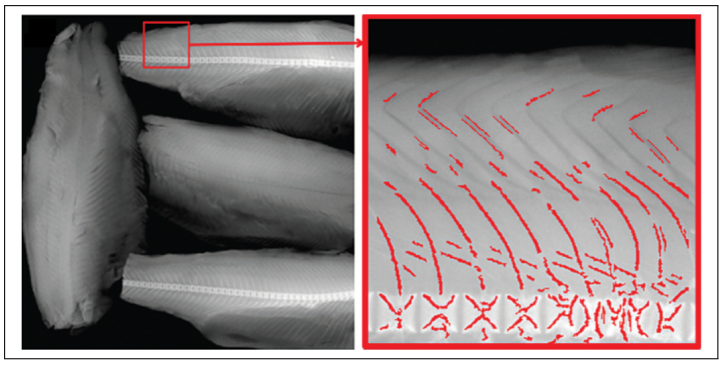


Figura 3.6. Detección de espinas de pescado, usando ventanas deslizantes, (Mery et al., 2011).

Tabla 3.4. Aplicación de los rayos X en inspección de alimentos.

Referencia	Energía		Modelo	Una Vista(**)			Visión	Múltiples Vistas(**)		
	Mono	Dual	Geométrico(*)	①	②	③	Activa	①	②	③
Haff y Slaughter (2004)	✓	–	–	✓	✓	✓	–	–	–	–
Jiang et al. (2008)	✓	–	–	✓	✓	✓	–	–	–	–
Ogawa et al. (2003)	✓	–	–	✓	✓	✓	–	–	–	–
Kwon et al. (2008)	✓	–	–	✓	✓	✓	–	–	–	–
Mery et al. (2011)	✓	–	–	✓	✓	✓	–	–	–	–

(*) C: Calibrado, N: No calibrado, –: no usado. (**) Ver ①, ②, ③ en Figura 3.1.

3.5. Cargamento

Con el continuo desarrollo del comercio internacional, la inspección de cargamentos se vuelve cada vez más importante. La inspección con rayos X se utiliza para la evaluación de los contenidos de cargamentos, camiones, contenedores, y vehículos de pasajeros, para detectar la posible presencia de muchos tipos de contrabando (ver Figura 3.7). Algunos enfoques se presentan en la Tabla 3.5, pero todavía no hay mucha investigación asociada a la inspección de cargamento, debido a la alta complejidad de esta tarea. Por esta razón, concluimos que estos sistemas de rayos X son todavía sólo semi-automático y requieren supervisión humana.

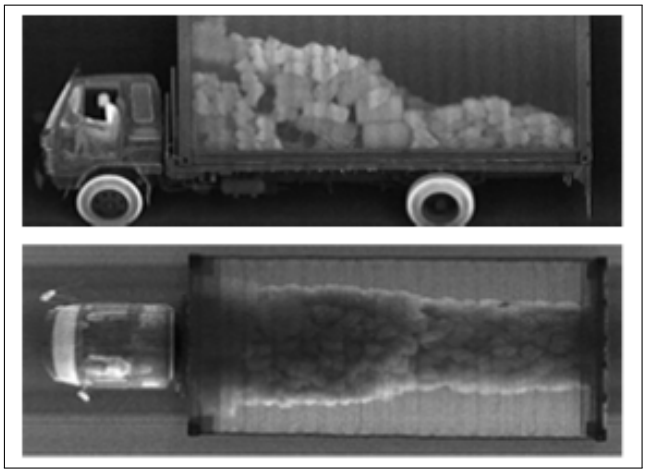


Figura 3.7. Imágenes de rayos X de un camión desde dos vistas distintas. Imagen tomada desde: `http://www.as-e.com/` [mayo del 2015].

Tabla 3.5. Aplicación de los rayos X en inspección de cargamento.

Referencia	Energía		Modelo	Una Vista(**)			Visión	Múltiples Vistas(**)		
	Mono	Dual	Geométrico(*)	(1)	(2)	(3)	Activa	(1)	(2)	(3)
Duan et al. (2009)	✓	–	–	✓	✓	✓	–	✓	✓	–
Frosio et al. (2011)	✓	–	–	✓	✓	✓	✓	✓	✓	–
Zhigang et al. (2005)	✓	–	–	✓	✓	✓	–	✓	✓	–
Zhigang et al. (2010)	✓	–	–	✓	✓	✓	–	✓	✓	–

(*) C: Calibrado, N: No calibrado, –: no usado. (**) Ver (1), (2), (3) en Figura 3.1.

3.6. Conclusiones

En este capitulo, hemos presentado una visión general de las metodologías de visión por computador que se han utilizado en las inspecciones con rayos X, como se ilustra en la Figura 3.1. Las aplicaciones presentadas y que usan pruebas de rayos X siguen este esquema general, y dependiendo de la forma en que las imágenes de rayos X son adquiridos y analizadas, cada paso puede (o no puede) ser utilizado.

En las aplicaciones presentadas, se observa que hay algunas áreas, como la inspección de fundición, donde los sistemas automatizados son muy eficaces, y otras áreas de aplicación, tales como la revisión de equipaje, donde aún se requiere la inspección humana. Además, hay ciertas áreas de aplicación, tales como soldadura y la inspección de cargamentos, donde la inspección es semi-automática. Por último, hay algo de investigación en el análisis de alimentos, donde está empezando a ser desarrollada utilizando imágenes de rayos X.

Está claro que muchas direcciones de investigación han sido explotadas, algunos principios muy diferentes se han adoptado y una amplia variedad de algoritmos se han desarrollado para aplicaciones muy diferentes. Sin embargo, la inspección automática con rayos X sigue siendo una cuestión abierta, ya que todavía sufre de: *i*) la pérdida de generalidad, porque los enfoques desarrollados para una aplicación no pueden utilizarse en otra; *ii*) la deficiente precisión en la detección, porque normalmente hay un *trade-off* fundamental entre las falsas alarmas y pérdidas en la detección; *iii*) la limitada robustez, debido a los

requisitos previos para el uso de un método que a menudo se cumplen sólo en casos sencillos; y *iv*) baja capacidad de adaptación, ya que puede ser muy difícil de acomodar un sistema automatizado para modificaciones de diseño o diferentes objetos.

En comparación con la inspección de rayos X manual, los sistemas automatizados ofrecen las ventajas de la objetividad y reproducibilidad para cada inspección. Las desventajas fundamentales son, sin embargo, la complejidad de su configuración, la inflexibilidad a cualquier cambio en el proceso de evaluación, y a veces, la imposibilidad de analizar imágenes intrincadas, que es algo que la gente en general, puede hacer bien. La investigación y el desarrollo continua en los procesos adaptativos automatizados para dar cabida a modificaciones.

Capítulo 4. METODOLOGÍA PROPUESTA

En este capítulo se describe la metodología que se utilizó para demostrar nuestra hipótesis, mediante la aplicación de un enfoque, el cual permite analizar imágenes de rayos X para detectar un objeto de interés en una buena pose, aplicando visón activa. Esta metodología incluye: la caracterización del objeto de interés que será detectado, un algoritmo de detección, un algoritmo de movimiento a partir de la pose del objeto detectado (búsqueda de la siguiente mejor vista) y un algoritmo de eliminación de falsas alarmas. Adicionalmente, nuestro enfoque incluye la construcción de un entorno o sistema de inspección, modelado geométricamente (ver sección 2.4), en el cual hemos realizado los procesos de caracterización y luego de inspección activa. Nuestra propuesta metodológica describe las distintas etapas que dan origen al *framework* desarrollado, y en tal sentido hemos realizado una propuesta inicial (Riffo y Mery, 2012), y posteriormente algunas mejoras en las principales etapas: ampliación del proceso de inspección a un modelo calibrado; implementación de un detector que permite categoría de objetos (Riffo y Mery, 2016), basado en el modelo de forma implícita (ISM: *Implicit Shape Model*); perfeccionamiento de la etapa de estimación del siguiente movimiento del objeto de prueba (que contiene al objeto de interés) para lograr una mejor pose, mediante un algoritmo de *Q-learning*; implementación de un algoritmo para eliminar falsas alarmas, que utiliza la teoría epipolar y trifocal. Los experimentos y la evaluación del desempeño de nuestra propuesta inicial de *framework*, así como también, los hallazgos preliminares de la unión de todas las mejorías a dicho *framework*, se verán en el Capítulo 5.

4.1. Propuesta de *Framework* de Inspección Activa

En nuestra propuesta, un objeto de prueba, es decir, el objeto de interés que será detectado, puede estar situado en el interior de un objeto contenedor. Por lo general, hay muchos objetos en el interior de un objeto contenedor, y sólo unos pocos de ellos, –si es que los hay–, son objetos de interés. Por esta razón, decimos que todo objeto que será inspeccionado, es un objeto de inspección complejo. Con el fin de diseñar e implementar nuestro

framework, hemos utilizado –sin pérdida de generalidad– una hojas de afeitar (objetos de interés) en el interior de diferentes objetos contenedores, ya que presenta la ventaja de ser un objeto con características interesantes de abordar, como lo es su delgadez, la poca absorción a los rayos X, la simetría en todos sus cuadrantes, su fácil portabilidad y la de ser considerado un elemento peligroso en aeropuertos (ver Figura 4.1).

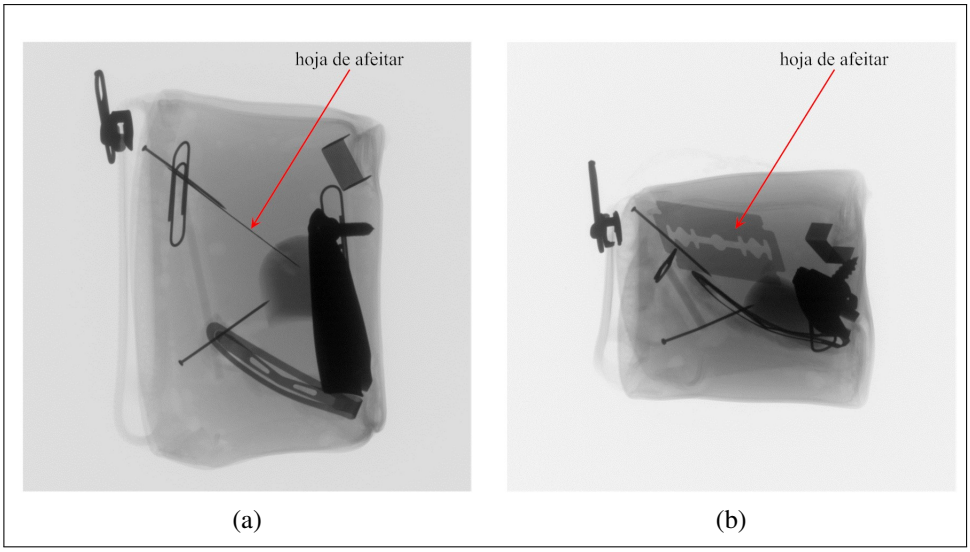


Figura 4.1. Dos vistas radioscópicas de un mismo objeto, en donde el objeto de interés (*hoja de afeitar*) se encuentra en: a) pose que impide detección, y b) mejor pose que facilita detección.

4.1.1. *Framework* General

El *framework* general, intenta encontrar una imagen en la cual un objeto de interés, es visto en una buena pose, que asegure su detección. Las buenas poses del objeto de interés, corresponden a aquellas vistas adquiridas, que deben tener una alta probabilidad de detección. En nuestro enfoque, las buenas poses de una hoja de afeitar corresponden a las vistas frontales. Por lo tanto, la idea principal es rotar y/o trasladar el objeto sometido a inspección, desde una posición inicial, a una nueva, en que la probabilidad de detección del objeto de interés es mayor. Está claro que si la posición inicial corresponde a una buena vista, no será requerida una nueva posición, y en estos casos la inspección se realiza con sólo una imagen de rayos X, evitando así, el análisis de más imágenes.

El algoritmo propuesto consta de dos partes (A y B), como se ilustra en la Figura 4.2:

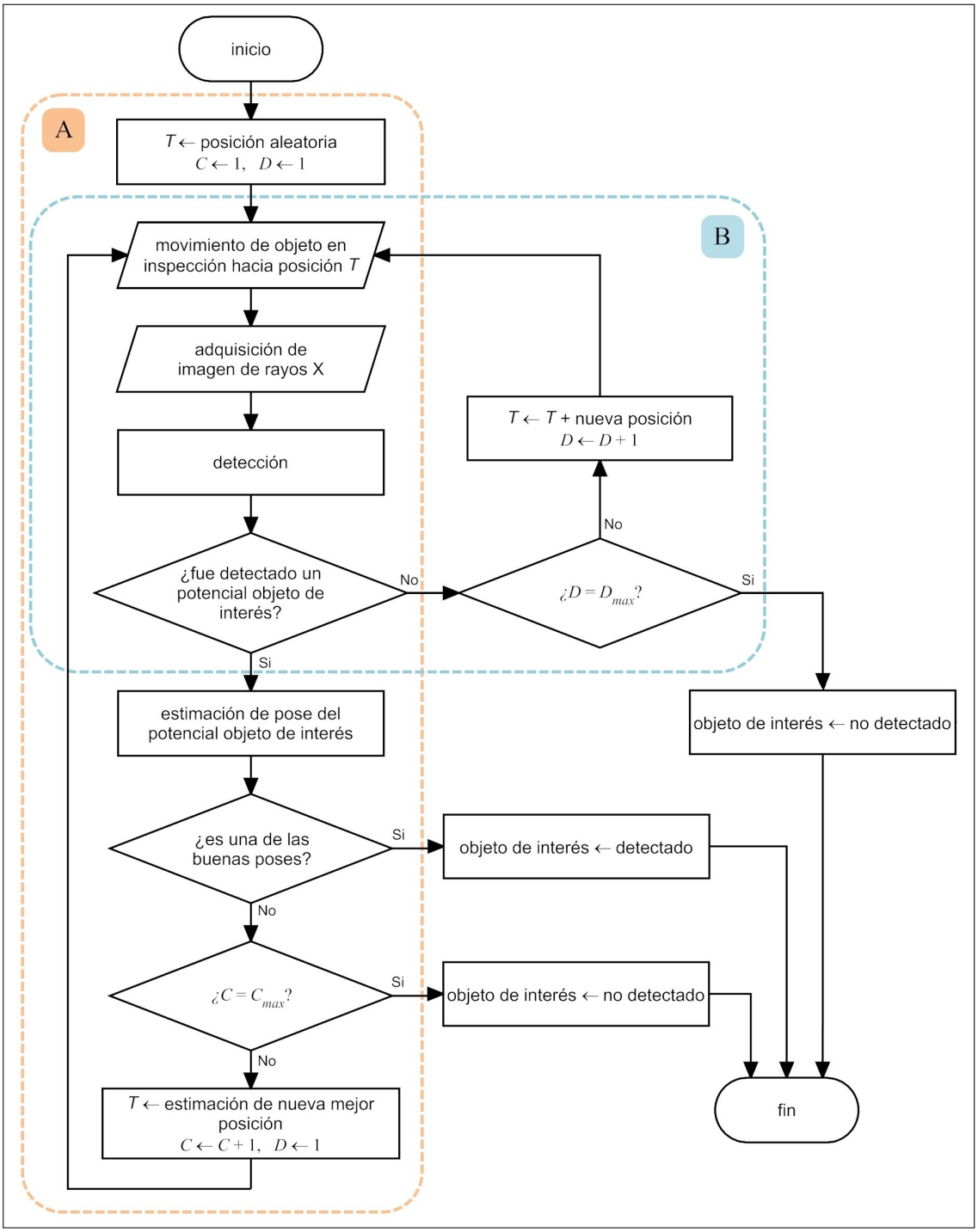


Figura 4.2. Framework para inspección activa con rayos X.

En la **parte A** se escoge una posición inicial arbitraria para el objeto en inspección y se adquiere una imagen de rayos X. En la etapa de detección, se busca un objeto de interés, y si se detecta un potencial objeto de interés, se estima su pose. Si la pose estimada corresponde a una de las buenas poses, luego de la inspección se termina la detección, de lo contrario el objeto en inspección se debe mover –usando la pose estimada–, de modo que la siguiente pose corresponda a una buena pose. Este proceso se interrumpe después de C_{max} veces, con el fin de evitar bucles infinitos. Por otro lado, en la **parte B**, si no se detecta un potencial objeto de interés –en la etapa de detección–, entonces el objeto es movido arbitrariamente a un nueva posición muy diferente a la primera, repitiendo la etapa de detección. La anterior situación puede repetirse hasta D_{max} veces, con el fin de garantizar la inspección de todos los puntos de vista relevantes. En nuestros experimentos, $C_{max} = 4$ y $D_{max} = 3$.

Esta propuesta de *framework* es general y fácilmente adaptable. Sin embargo, el diseño de los algoritmos para la detección y estimación de pose, dependen de la aplicación. Como se mencionó anteriormente, en nuestros experimentos utilizamos una hoja de afeitar como un objeto de interés. Por esta razón, los detalles de nuestro método para esta aplicación en particular, son explicados a continuación.

4.1.2. Caracterización del Objeto de Interés

Debe ser definido un algoritmo para detectar automáticamente y en una sola imagen los potenciales objetos de interés. Como se mencionó anteriormente, con el fin de probar nuestro *framework*, hemos desarrollado un algoritmo que es capaz de detectar hojas de afeitar. Antes de poder aplicar un algoritmo de detección es necesario caracterizar al objeto de interés. Este algoritmo se basa en la bien conocida técnica utilizado por la comunidad de visión por computador, llamada SIFT (*Scale Invariant Feature Transform*), propuesta por Lowe (2004). SIFT es capaz de detectar y extraer –en ciertos puntos de una imagen– descriptores de características locales que son muy robustos al ruido, cambios en la escala, rotación, punto de vista y contraste. Un descriptor SIFT es típicamente un vector con 128 elementos, calculados a partir de 16 histogramas de gradientes, en 8 direcciones en la vecindad a un punto (*Keypoint*). Por lo tanto, la correspondencia de *Keypoints* entre dos

imágenes diferentes del mismo objeto, se realiza de manera eficiente comparando sus descriptores SIFT, es decir, mediante la búsqueda de la distancia Euclídea mínima entre los descriptores. En nuestro enfoque, usamos una caracterización SIFT del objeto de interés, en distintas poses, logradas mediante la rotación en dos ejes, en nueve pasos por eje, tal como se muestra en la Figura 4.3. Así, un *keypoint* k tendrá un descriptor $\mathbf{f}_k$ localizado en $\mathbf{z}_k = (x_k, y_k)$. Cada descriptor $\mathbf{f}_k$ tiene asociada una pose r_k. Para la hoja de afeitar, $r_k \in [1, 81]$, es decir, 9×9 poses.

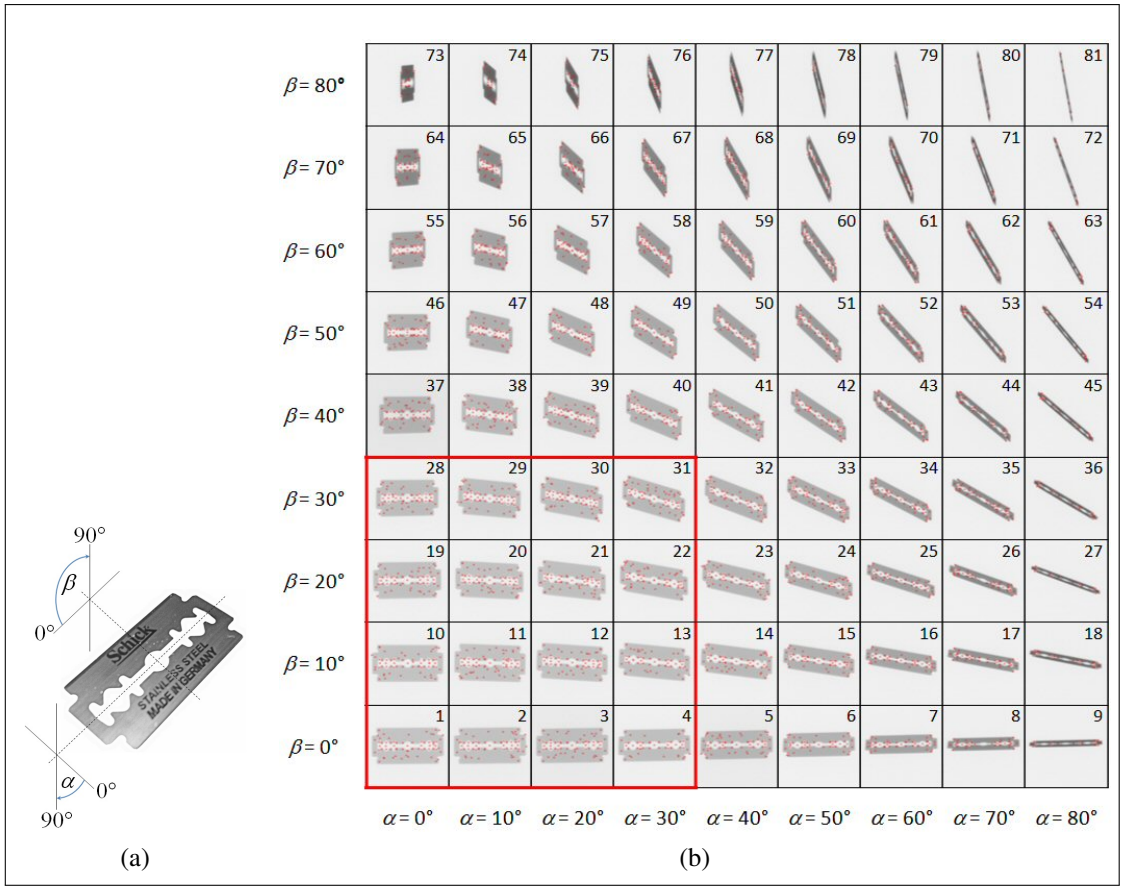


Figura 4.3. Caracterización: (a) Rotaciones aplicadas al objeto de interés para su caracterización ($\alpha = 0°, 10°, ..., 80°$ y $\beta = 0°, 10°, ..., 80°$) y (b) Plantilla de caracterización con descriptores SIFT (puntos rojos) e indicación de pose. El recuadro rojo ubicado abajo a la izquierda contiene las mejores poses de la hoja de afeitar.

4.1.3. Detección del Objeto de Interés

La detección se realiza siguiendo la idea sugerida en (Lowe, 2004), que consiste en realizar el *matching* entre los descriptores $\mathbf{f}_t$ obtenidos de la imagen inspeccionada de rayos X y los descriptores $\mathbf{f}_k$ obtenidos y almacenados en el proceso de caracterización. Se encuentran los mejores descriptores $\mathbf{f}_t$ que hagan *matching* con algún $\mathbf{f}_k$, esto es, que cumplan con $d_E(\mathbf{f}_k, \mathbf{f}_t) < \theta_u$, donde $d_E(\mathbf{f}_k, \mathbf{f}_t) = ||\mathbf{f}_k - \mathbf{f}_t||$ es la distancia Euclídea entre $\mathbf{f}_k$ y $\mathbf{f}_t$, y θ_u es un umbral de distancia mínima. Además, cada descriptor $\mathbf{f}_t$ encontrado se denominará $\hat{\mathbf{f}}$, con $\hat{\mathbf{f}} \subseteq \mathbf{f}_t$. De esta forma $\hat{\mathbf{f}}$ tendrá asociada la pose r_k del descriptor $\mathbf{f}_k$ que hizo el mejor *matching*.

Cada imagen sometida a inspección tendrá ahora sólo los descriptores $\hat{\mathbf{f}}$ con las poses r_k, que tienen la menor distancia a algún descriptores $\mathbf{f}_k$. A partir de esto, la detección se realiza en dos etapas: *localización* y *encuadramiento*.

a) Localización: En esta etapa se seleccionan algunos de los descriptores $\hat{\mathbf{f}}$ que están próximos entre sí y que tienen una pose común. Para esto, se define una pequeña ventana de tamaño $w \times h$ en pixeles, W_B, cuyo centro corresponde a la posición $\hat{\mathbf{z}} = (\hat{x}, \hat{y})$ de cada descriptor. Así, sólo permanecerán los descriptores $\hat{\mathbf{f}}$ en la posición $\hat{\mathbf{z}} = (\hat{x}, \hat{y})$ que tengan un número θ_B de descriptores situados dentro de W_B y que tengan la misma pose (ver ejemplo en Figura 4.4b). En nuestros experimentos se escogió $\theta_B = 3$, $w = h = 80$ pixeles. En nuestro método hemos definido que un descriptor $\hat{\mathbf{f}}$ permanece en la imagen, si y sólo si, a lo menos existe $\theta_B = 3$ descriptores con la misma pose dentro de la ventana W_B de tamaño 80×80 pixeles. Si lo anterior ocurre, W_B se establece como ventana válida.

b) Encuadramiento: Todas las ventanas W_B que se encuentren unidas o traslapadas, formarán una nueva ventana W_G de tamaño variable, mayor o igual que el tamaño de W_B, ver Figura 4.4b, 4.4c.

4.1.4. Estimación de Pose

Para estimar la pose, se exige a cada ventana detectada W_G, contenga a lo menos θ_G descriptores con la misma pose (en nuestros experimentos se escogió $\theta_G = 8$), y la pose que tiene mayor frecuencia, será la pose asignada a W_G. Si lo anterior no ocurre, se descarta la W_G, y por el contrario, si hay más de una W_G con pose asignada, se selecciona como detección válida a la W_G que tenga mayor número de descriptores con igual pose, ver Figura 4.4d.

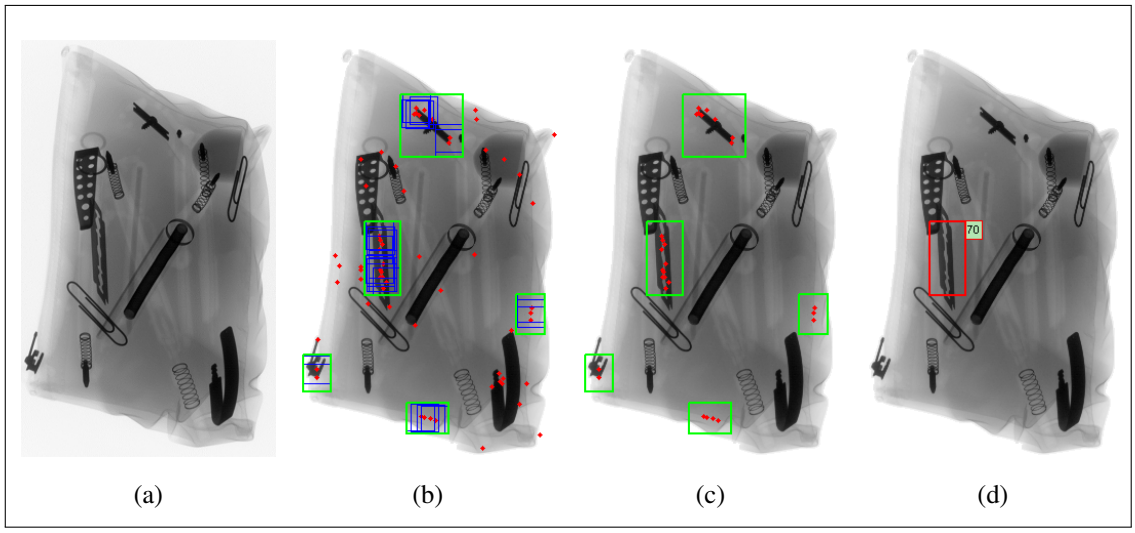


Figura 4.4. Etapas para detectar un objeto de interés y estimar su pose: a) imagen inicial, b) localización de ventanas W_B (ventanas azules) y encuadramiento de ventanas W_G (ventanas verdes), c) ventanas W_G encuadradas, y d) pose detectada = 70 (se puede comparar con pose 70 mostrada en la figura 4.3).

4.1.5. Estimación de Movimiento

Un objeto sometido a inspección con rayos X, puede estar en distintas posiciones, sin embargo, no todas serán adecuadas para la detección de algún objeto de interés. Sumado a este inconveniente, está el hecho de que un objeto de interés tiene vistas que dificultan su reconocimiento visual. Es necesario entonces establecer una estrategia de rotación del objeto y que permita detectar al objeto de interés en un reducido número de imágenes (*visión activa*). La rotación del objeto se realiza considerando dos situaciones: *no detección* y *detección*, y en ambas se realizan rotaciones (a no ser que se hayan producido un número

máximo C_{max} de no detecciones *ND* sucesivas, que en nuestro caso es 3), de tal forma de colocar al objeto de interés en una mejor vista (ver mejores vistas en Figura 4.3). De forma general, se podrían realizar rotaciones en torno a los tres ejes X, Y, Z del sistema cartesiano del objeto. Sin embargo, en nuestra propuesta tenemos restricción en los grados de libertad, por lo cual, sólo es posible definir dos rotaciones; una en torno al eje Z y otra en torno al eje X, cuando se usa el sistema semi-automático[1] de sujeción, y sólo rotación en torno al eje X, cuando se usa el brazo robótico como sistema de sujeción, considerando que cualquier mecanismo de inspección (manual, semi-automático y automático) tendrá limitados grados de libertad sin causar oclusión en las imágenes de rayos X. Cuando el objeto de interés no es detectado (*ND*), ya sea por la no existencia al interior del objeto, o por la posición en que este se encuentra, se realizará sólo una rotación de $\alpha = -40°$ en torno al eje X. Esto se ha determinado de manera heurística, ya que al realizar una rotación de 40° (positiva o negativa) en torno al eje X, se podría aproximar rápidamente a una zona de mejor vista, según Figura 4.3.

Para detectar el objeto de interés y estimar su pose, se ha requerido que a lo menos θ_G descriptores $\hat{\mathbf{f}}$ hayan hecho *matching* con los descriptores $\mathbf{f}_k$. Estos descriptores $\hat{\mathbf{f}}$ son útiles para calcular la rotación γ en torno al eje Z. Esta rotación permite dejar al objeto de interés de forma horizontal o lo más próximo a esa posición, es decir con $\beta = 0°$ (ver Figura 4.3). Para esto es necesario ajustar a una elipse la posición $(\hat{x}, \hat{y})$ de los descriptores $\hat{\mathbf{f}}$, y determinar el ángulo ϕ del eje mayor de la elipse con la horizontal. Si se cumple que el ángulo $\phi \geq 90° \Rightarrow \gamma = 180° - \phi$, y si el ángulo $\phi < 90° \Rightarrow \gamma = -\phi$, tal como se muestra en la Figura 4.5.

Cuando se utiliza el sistema semi-automático luego que el objeto de interés es detectado y se ha estimado su pose, se provoca la rotación en Z tal como ya fue explicado. Con la estimación de la pose, el algoritmo busca en la base de datos de todas las poses (ver Figura 4.3), la rotación en torno al eje X que dio origen a dicha pose, entonces la rotación será la misma, pero en sentido contrario, es decir $-\alpha$.

[1]Nuestro sistema semi-automático consta de: giroscopio de acrílico que permite mover un objeto en forma manual y en forma automática a través de un motor.

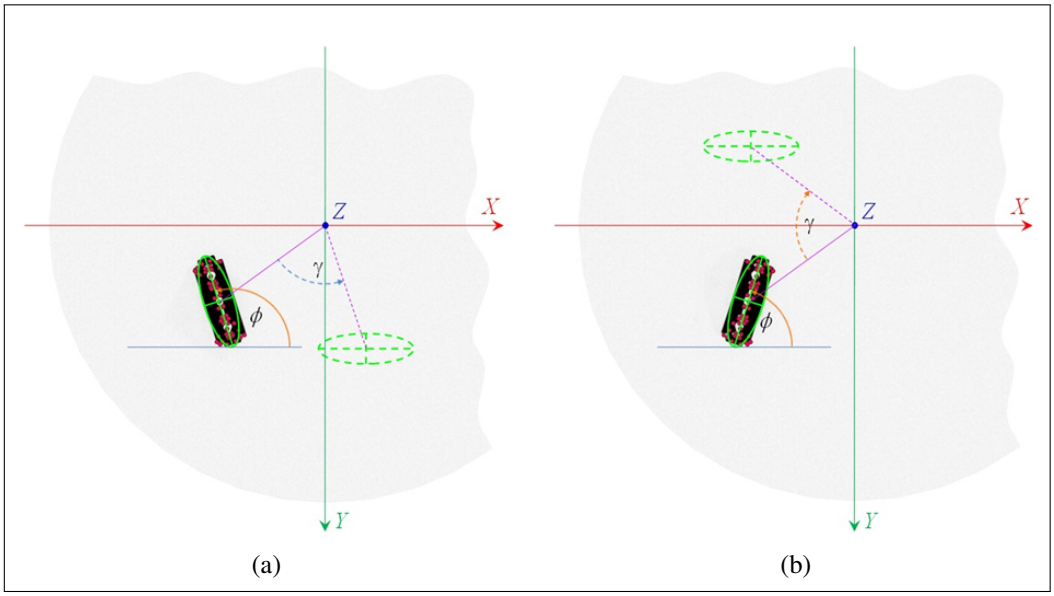


Figura 4.5. Rotación en Z: a) cuando $\phi \geq 90° \Rightarrow \gamma = 180° - \phi$, b) cuando $\phi < 90° \Rightarrow \gamma = -\phi$.

Cuando se utiliza el brazo robótico no es posible realizar rotación en Z, debido a las características del brazo robótico, y en este caso sólo es posible realizar rotación en X sin causar oclusión. Entonces, al igual que con el sistema semi-automático, con la detección y estimación de la pose, se determina la rotación en torno al eje X.

Independiente del sistema de sujeción del objeto, las rotaciones terminan cuando a) en tres ocasiones sucesivas no se ha logrado la detección, b) se alcanza una mejor vista, y c) excepcionalmente se ha alcanzado una pose con $\alpha = 0°$ (restricción de rotaciones).

Con este enfoque, se intenta obtener una nueva posición del objeto sometido a inspección, en donde la pose del objeto de interés debe ser aquella que permita una buena percepción de él (buena pose). Es importante señalar que debido a la perspectiva, las imágenes de rayos X de la hoja de afeitar, obtenidas desde cuatro diferentes puntos de vista $(\pm\alpha, \pm\beta)$ son muy similares. Por esta razón, la estimación de pose, y por lo tanto, la estimación de las próximas mejores posiciones, puede fallar. En estos casos, la siguiente vista puede no ser una buena vista y por lo tanto, el objeto de interés no será detectado o no se obtendrá una buena pose. Hay dos alternativas para hacer frente a esta posición que fue erróneamente estimada: i) devolverse en la rotación y luego corregir la estimación de pose,

y ii) tratar de obtener una nueva próxima mejor posición desde esta posición errónea. Desde un punto de vista práctico, y debido a que la opción i) se puede repetir hasta tres veces, nos decidimos por la opción ii). Nuestros experimentos validaron esta opción.

Teóricamente, en el caso de la inspección con un sistema de manipulación con sólo un eje de rotación (caso del brazo robótico), la detección podría fallar, esto es: que la hoja de afeitar esté dispuesta en forma perpendicular al eje de rotación, en cuyo caso la detección nunca ocurrirá. Para superar dicha situación, podría otorgarse la capacidad de movimiento, al algoritmo de control del robot, para que una vez ocurridas las tres no detecciones ND sucesivas, el brazo robótico coloque el objeto sobre el detector y lo tome en otra posición, y así ejecutar una nueva secuencia de inspección.

4.2. Mejoras a la Propuesta de *Framework* de Inspección Activa

En esta sección abordaremos algunas mejoras a nuestra propuesta inicial de *Framework* presentada en la sección 4.1, para así aumentar la robustez del algoritmo de inspección activa con rayos X. Las mejoras que proponemos son las siguientes:

a) Implementar un detector de categorías de objetos (objetos amenazantes),
b) Implementar un estimador de movimiento con Q-learning, e
c) Implementar un sistema que permita eliminar falsas alarmas.

4.2.1. Detector de Objetos Amenazantes

En esta sección, se explica el enfoque propuesto, que se puede utilizar para detectar automáticamente objetos de interés (objetos amenazantes) en las imágenes de rayos X. Hemos adaptado el modelo de forma implícita ISM (Leibe et al., 2008) para incrementar su eficacia y robustez en el procesamiento de imágenes de rayos-X (Riffo y Mery, 2016). A nuestro método propuesto lo hemos denominado, modelo de forma implícita adaptado AISM (AISM: *Adapted Implicit Shape Model*), el cual tiene dos pasos principales: *A*) caracterización del objeto de interés y *B*) detección de objetos de interés. El objeto de interés

que será detectado en las imágenes de rayos X es un objeto amenazante. Los pasos corresponden a las etapas de entrenamiento y prueba del detector, respectivamente. Una discusión más detallada de estos pasos se proporciona a continuación.

A. Caracterización del Objeto de Interés

La caracterización consiste en tres pasos: *1) Adquisición de imágenes de entrenamiento*: adquisición de imágenes de rayos X representativas del objeto amenazante, *2) Generación de un codebook*: creación de un vocabulario visual usando *keypoints* y descriptores visuales locales, y *3) Ocurrencia*: estimación de la posición de los *keypoints* relacionados con cada palabra del vocabulario visual.

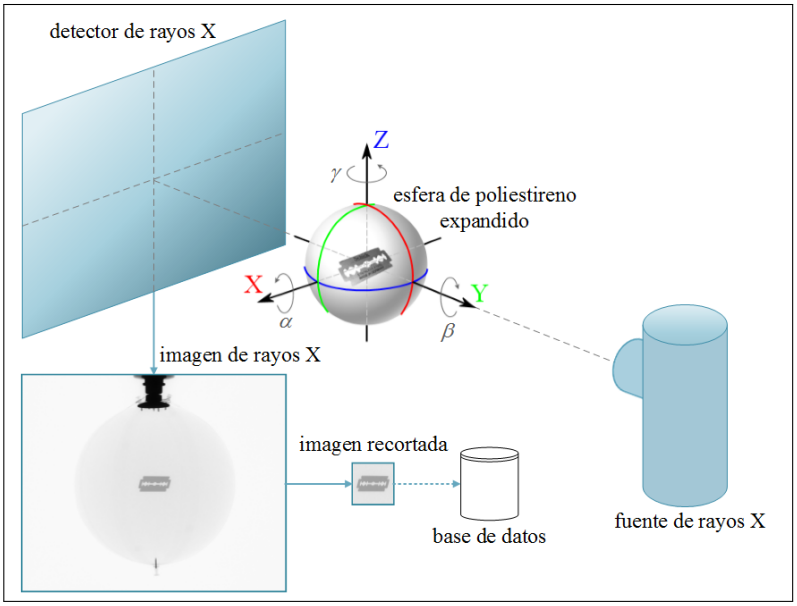


Figura 4.6. Sistema de adquisición de imágenes de rayos X para la caracterización de objetos de interés.

1) Adquisición de imágenes de entrenamiento:

Se utiliza una base de datos de entrenamiento de N imágenes de rayos X. Con el fin de adquirir imágenes de rayos X representativas de un objeto de interés en diferentes poses, es necesario implementar un sistema de adquisición, que puede adquirir imágenes de rayos X desde diferentes puntos de vista, como se muestra en la Figura 4.6 para una hoja de

afeitar. El objeto debe estar ubicado dentro de una esfera de poliestireno expandido (EPS). Se utilizó una esfera de EPS debido a su bajo coeficiente de absorción de rayos X.

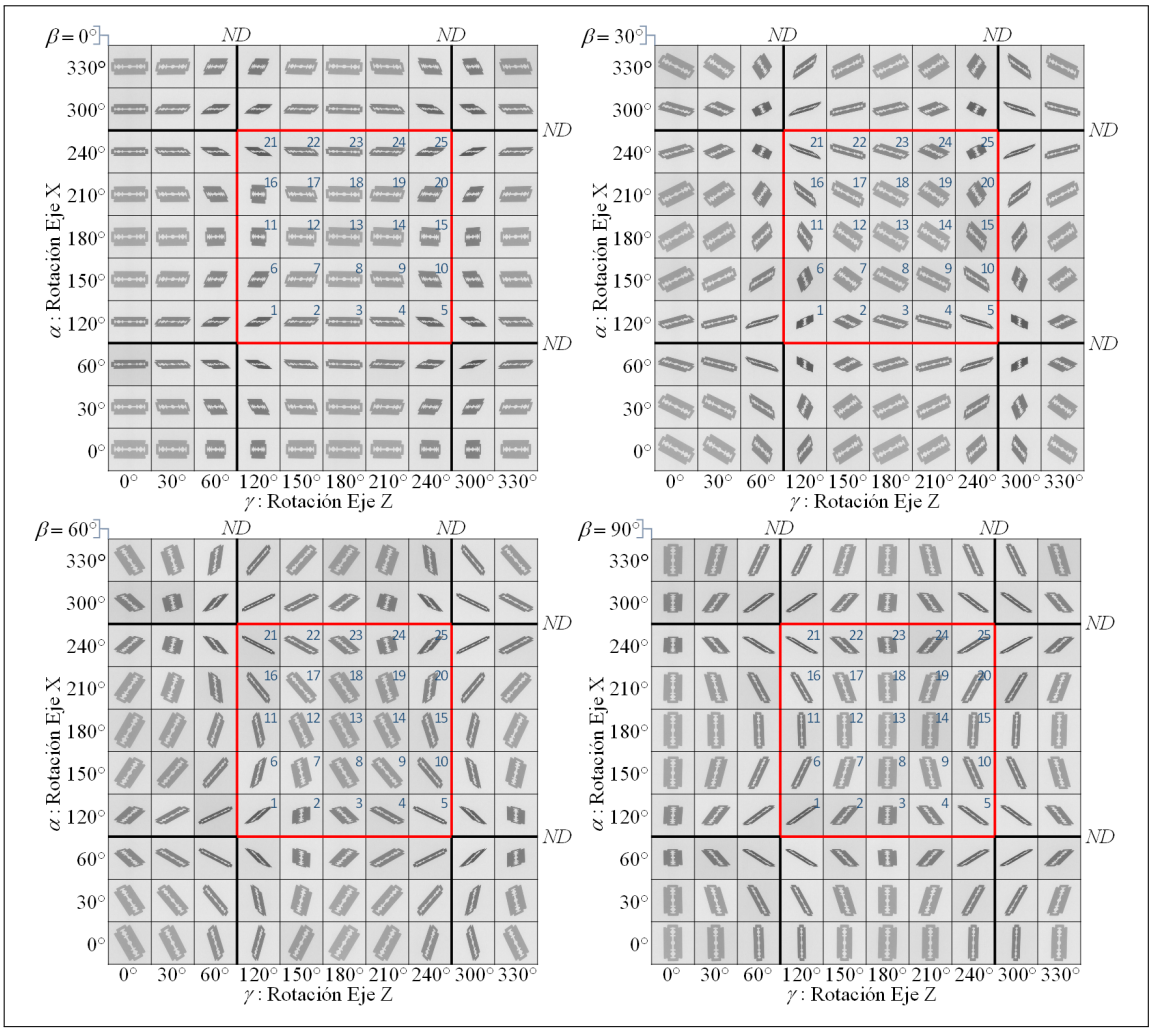


Figura 4.7. Imágenes de rayos X de un objeto de interés (hoja de afeitar) adquiridas usando diferentes ángulos α, β y γ. Las imágenes de rayos X útiles para la caracterización están encerradas en el cuadro rojo, es decir, las imágenes entre α: [120°, 150°, 180°, 210°, 240°] y γ: [120°, 150°, 180°, 210°, 240°]. Cada imagen encerrada en el cuadro rojo se asocia con una pose (1 a 25).

El sistema propuesto permite a los usuarios adquieren imágenes de un objeto en muchas poses, mediante la modificación de los ángulos de rotación; α, β y γ, asociado a cada eje, X, Y y Z de la esfera, respectivamente. Todas las imágenes de la hoja de afeitar se muestran en la Figura 4.7. Sin embargo, hay cuadrantes de imágenes que visualmente se

repiten. Por lo tanto, no todas las imágenes son útiles y sólo unas pocas se almacenan en la base de datos de entrenamiento. La base de datos incluye las imágenes adquiridas en los siguientes valores de ángulos (ver Figura 4.7): $\alpha \in \{120°, 150°, 180°, 210°, 240°\}$, $\gamma \in \{120°, 150°, 180°, 210°, 240°\}$ y $\beta \in \{0°, 30°, 60°, 90°\}$. Las imágenes auto-ocluidas o imágenes con posiciones que no permiten la extracción de características discriminativas (por ejemplo, $\alpha = 90°$ o $\gamma = 270°$) se han eliminado. Se presentan en la Figura 4.7 usando el acrónimo *ND* (No Detección). Se debe tener en cuenta que la base de datos se muestra en la Figura 4.7 se hizo utilizando una hoja de afeitar que no tiene variabilidad en su categoría de objeto. Para un objeto con grandes variaciones intra-clase, la base de datos debe incluir un conjunto representativo de imágenes de la categoría de objeto. Este procedimiento también se realizó para estrellas ninja (*shurikens*) y revólveres.

Este proceso nos permitió obtener una base de datos de N imágenes de entrenamiento. Para cada imagen de entrenamiento, el centro del objeto (c_x^j, c_y^j) medido en píxeles, se calcula para $j = 1, \ldots, N$.

2) Generación de un codebook:

En esta etapa, un objeto de interés es representado usando un vocabulario visual de partes (*codebook* de apariencia de una categoría específica). Se extraen automáticamente los *keypoints* y sus descriptores visuales locales desde todas las imágenes de entrenamiento del objeto de interés, utilizando el conocido enfoque SIFT (Lowe, 2004). Un *keypoint* es un punto distinguible en una imagen, es decir, que representa un área de imagen sobresaliente, la cual puede ser reconocida al cambiar su punto de vista, la orientación y la escala. Por lo tanto, los *keypoints* SIFT son puntos distinguibles de una imagen, con un alto contenido de datos, en términos de la variación local en la señal. Un *keypoints* k tiene un descriptor $\mathbf{f}_k$ y una ubicación $\mathbf{z}_k = (x_k, y_k)$ con respecto al centro (c_x^j, c_y^j) del objeto de interés, visto en su correspondiente imagen de entrenamiento j, como se muestra en la Figura 4.8.

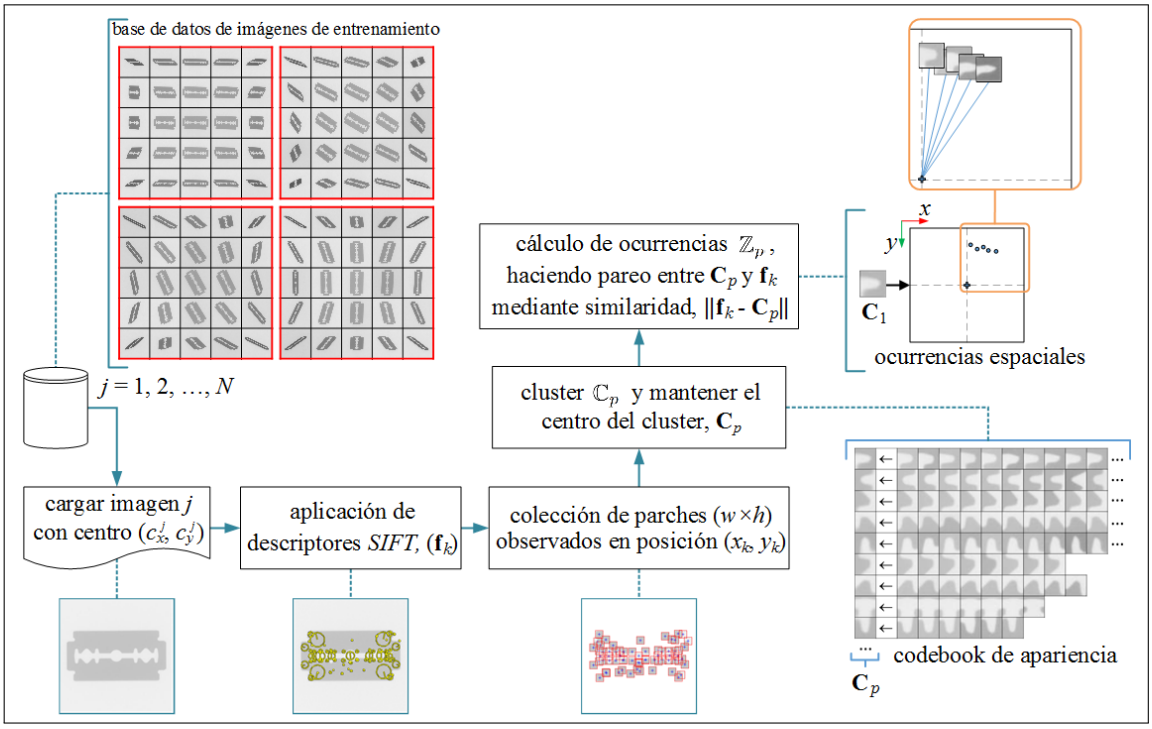


Figura 4.8. Caracterización de un objetos de interés: imágenes de entrenamiento, generación de *codebook* y cálculo de las ocurrencia de una hoja de afeitar. Aquí los *keypoints* y descriptores se visualizan como pequeños parches.

Para garantizar que un efectivo vocabulario visual sea desarrollado para las diversas partes de los objetos, es importante que las partes visualmente similares y con características similares, se agrupan para crear un *codebook* de apariencias locales. En nuestro enfoque, usamos una estrategia de *clustering* aglomerativo (Leibe et al., 2008). Para iniciar el proceso de *clustering* aglomerativo, cada característica SIFT $\mathbf{f}_k$, es decir, un vector de 128 elementos, se le asigna a un *cluster*. El par más similar de *clusters* se fusiona de forma secuencial hasta que se cumpla un criterio de parada. El par más similar de *clusters* se establece mediante la minimización de una métrica de distancia entre los *clusters*. Formalmente, un *cluster* p se define como un conjunto $\mathbb{C}_p$ que contiene todos las características SIFT $\mathbf{f}_k$ fusionadas. Como métrica de distancia entre dos *clusters*, p y q, utilizamos:

$$d(\mathbb{C}_p, \mathbb{C}_q) = \min(d_E(\mathbf{f}_p, \mathbf{f}_q)) \quad \text{para todo } \mathbf{f}_p \in \mathbb{C}_p, \mathbf{f}_q \in \mathbb{C}_q, \tag{4.1}$$

donde $d_E(\mathbf{f}_p, \mathbf{f}_q) = ||\mathbf{f}_p - \mathbf{f}_q||$ es la distancia Euclídea entre la características $\mathbf{f}_p$ y $\mathbf{f}_q$.

El *clustering* aglomerativo se detiene cuando se obtiene un cierto número de *clusters* o cuando todas las distancias inter-*cluster* están por encima de un determinado umbral. En nuestra aplicación, los mejores resultados se lograron mediante el primer criterio, donde el número predefinido de *clusters* (M) para cada categoría de objeto (hoja de afeitar, *shuriken* y revolver) se establece arbitrariamente en 400. El centro de masa del *cluster* p es una palabra de nuestro vocabulario visual, definido como:

$$\mathbf{C}_p = \frac{1}{n_p} \sum_{\mathbf{f}_p \in \mathbb{C}_p} \mathbf{f}_p \qquad \text{para } p = 1, \ldots, M. \tag{4.2}$$

donde n_P es el número de *keypoints* en el *cluster* $\mathbb{C}_p$. El *codebook* se define como el centro de masa de cada *cluster* $\mathbf{C}_p$, y las muestras que pertenecen a cada cluster $\mathbb{C}_p$, para $p = 1, \ldots, M$.

3) Ocurrencia:

En esta etapa, se calcula una estructura denominada "ocurrencia" para cada *cluster* de un objeto de interés. La ocurrencia del *cluster* p, denotada como el conjunto $\mathbb{Z}_p$, para $p = 1, \ldots, M$, contiene todos los *keypoints* de las imágenes de entrenamiento, cuyos descriptores SIFT son lo suficientemente similar al centro de masa del *cluster* $\mathbf{C}_p$ estimado en la ecuación (4.2). Formalmente, $\mathbb{Z}_p$ es un conjunto de coordenadas $\mathbf{z}_k = (x_k, y_k)$ definidos de la siguiente manera:

$$\mathbb{Z}_p = \{\mathbf{z}_k \ : \ d_E(\mathbf{f}_k, \mathbf{C}_p) < \theta\} \tag{4.3}$$

donde $\mathbf{f}_k$ es el descriptor SIFT del *keypoint* $\mathbf{z}_k$. Como se indicó anteriormente, las coordenadas de los *keypoint* se definen con respecto al centro del objeto.

B. Detección de Objetos de Interés

Nuestro método de detección de objetos de interés en imágenes de rayos X, tiene cuatro etapas principales: *1) Extracción de características*, *2) Entradas coincidentes del codebook y espacio de votación*, *3) Fusión de candidatos detectados*, y *4) Detección*. La Figura 4.9

muestra un resumen de este proceso. Una explicación detallada de cada etapa se presenta a continuación.

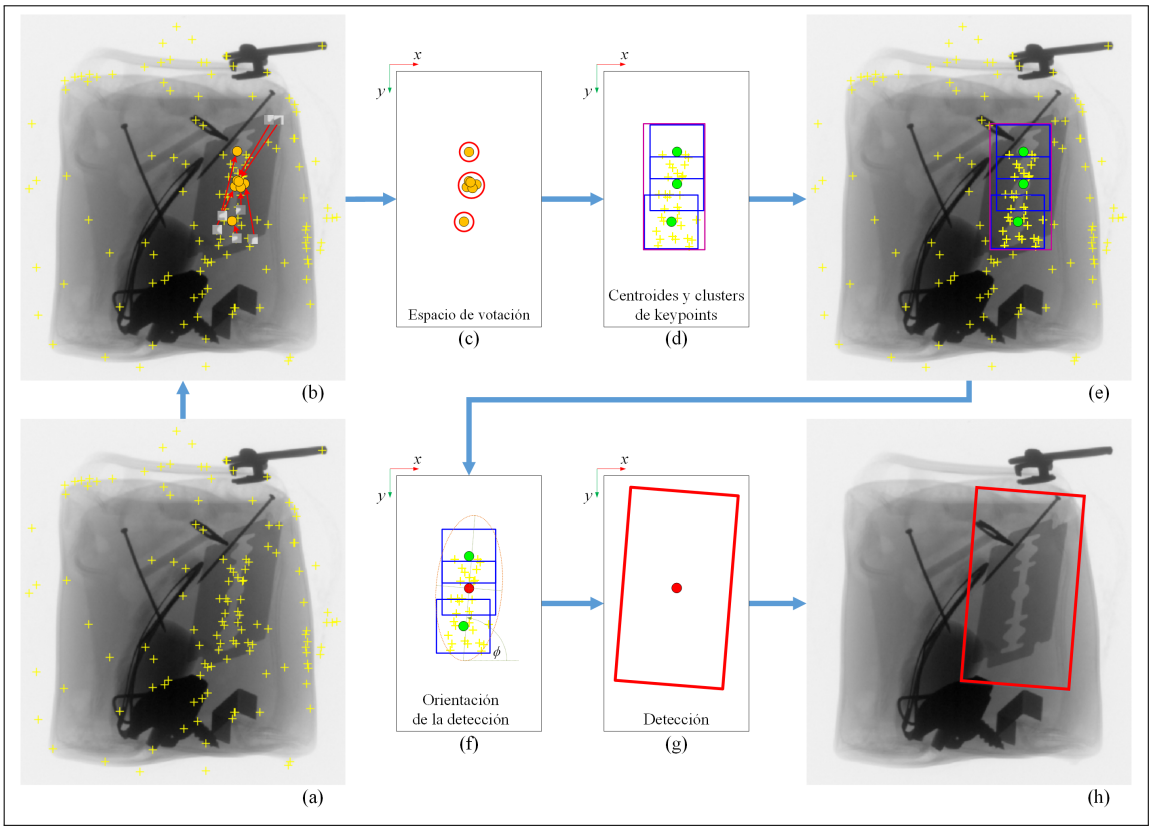


Figura 4.9. Proceso de detección para objetos de interés en imágenes de rayos X, usando el método propuesto AISM. (a) Imagen con sólo *keypoints* útiles $\hat{\mathbf{f}}$, (b) Imagen con entradas del *codebook* que hicieron *matching* y espacio de votación, (c) Detalle del espacio de votación, con los candidatos máximos detectados, (d) Detalle de espacio de votación con los centros de los candidatos máximos seleccionados y *cluster* de *keypoints* (con subventanas azules W_B), (d) y (e) *Clusters* fusionados (con subventana magenta W_m), (e) *Clusters* previo a la detección, (f) Ajuste a una Elipse para estimar el ángulo de orientación, (g) y (h) Detección final, h) Hoja de afeitar detectada en la imagen de rayos X en.

1) Extracción de características:

Aplicamos un detector de *keypoints* en la imagen de rayos X que es inspeccionada, con el fin de extraer toda la característica $\mathbf{f}_t$. En general, las imágenes de rayos X proporcionan un gran número de *keypoints*. En primer lugar, realizamos el pareo (*matching*) entre todos los *keypoints* $\mathbf{f}_k$ almacenada durante el proceso de entrenamiento y todos los *keypoints* $\mathbf{f}_t$

obtenidos de la imagen inspeccionada, utilizando la expresión: $d_E(\mathbf{f}_k, \mathbf{f}_t) < \theta_u$, donde θ_u es el umbral mínimo de la distancia permitida entre $\mathbf{f}_k$ y $\mathbf{f}_t$. Esos puntos clave que cumplen esta expresión se designan como $\hat{\mathbf{f}}$. Así, $\hat{\mathbf{f}} \subseteq \mathbf{f}_t$. Por lo tanto, eliminamos los *keypoints* innecesarios de la imagen inspeccionada y mantenemos sólo el conjunto útil de *keypoints* $\hat{\mathbf{f}}$. La Figura 4.9(a) muestra un ejemplo de los *keypoints* útiles resultantes. Usamos los descriptores SIFT como medida la apariencia de cada *keypoint*.

2) Entradas coincidentes del codebook y espacio de votación:

Esta etapa de AISM puede ser descrita utilizando el enfoque probabilístico propuesto en (Leibe y Schiele, 2003; Leibe et al., 2008). Dado $\hat{\mathbf{f}}$ será el descriptor de un *keypoint* útil observada en la ubicación $\hat{\mathbf{z}} = (\hat{x}, \hat{y})$ de la imagen inspeccionada de rayos X. En primer lugar, cada *keypoint* $\hat{\mathbf{f}}$ es mapeado a la palabra más cercana $\mathbf{C}_p$ de nuestro vocabulario visual aprendido en la 'etapa de generación del *codebook*' (ver etapa *A.2*). La probabilidad de que $\hat{\mathbf{f}}$ haga *matching* con alguna palabra $\mathbf{C}_p$ del vocabulario visual, se puede expresar como $\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}}, \hat{\mathbf{z}})$. Luego, la transformada de Hough generalizada (Ballard, 1981) captura la configuración coherente de varias palabras visuales. Cada palabra $\mathbf{C}_p$ que hace *matching*, genera votos para las instancias de la categoría de objeto o_n, en diferentes posiciones $\lambda = (\lambda_x, \lambda_y)$, de acuerdo a su distribución espacial de ocurrencia aprendida $\mathcal{P}(o_n, \lambda|\mathbf{C}_p, \hat{\mathbf{z}})$. Esta distribución de probabilidad corresponde a los lugares de ocurrencia de *keypoints*, aprendidos durante el proceso de entrenamiento. La distribución anterior se puede expresar formalmente por la siguiente marginación:

$$\mathcal{P}(o_n, \lambda|\hat{\mathbf{f}}, \hat{\mathbf{z}}) = \sum_p \mathcal{P}(o_n, \lambda|\hat{\mathbf{f}}, \mathbf{C}_p, \hat{\mathbf{z}})\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}}, \hat{\mathbf{z}}). \tag{4.4}$$

para $p = 1, \ldots, M$, donde M es el número de palabras en nuestro *codebook*. Puesto que hemos reemplazado el desconocido descriptor $\hat{\mathbf{f}}$ en la imagen de rayos X que es inspeccionada, con una conocida interpretación $\mathbf{C}_p$, el primer término de la ecuación (4.4) pueden ser tratados como independientes de $\hat{\mathbf{f}}$. Además, hacemos *matching* de descriptores con el *codebook* independiente de su ubicación. Por lo tanto, la ecuación se reduce a:

$$\mathcal{P}(o_n, \lambda|\hat{\mathbf{f}}, \hat{\mathbf{z}}) = \sum_p \mathcal{P}(o_n, \lambda|\mathbf{C}_p, \hat{\mathbf{z}})\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}}), \tag{4.5}$$

$$= \sum_p \mathcal{P}(\lambda|o_n, \mathbf{C}_p, \hat{\mathbf{z}})\mathcal{P}(o_n|\mathbf{C}_p, \hat{\mathbf{z}})\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}}). \tag{4.6}$$

donde, $\mathcal{P}(\lambda|o_n, \mathbf{C}_p, \hat{\mathbf{z}})$ es el voto probabilístico Hough, para una posición del objeto λ, dada su etiqueta de clase o_n, la palabra $\mathbf{C}_p$ y la ubicación $\hat{\mathbf{z}}$ del *keypoint*. La probabilidad $\mathcal{P}(o_n|\mathbf{C}_p, \hat{\mathbf{z}})$ especifica la confianza que la palabra $\mathbf{C}_p$ en la posición $\hat{\mathbf{z}}$ del *keypoint*, haga *matching* en la categoría de objeto o_n. Finalmente, $\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}})$ refleja la probabilidad de *matching* entre el descriptor de imagen $\hat{\mathbf{f}}$ y la palabra $\mathbf{C}_p$. Si una característica de imagen $\hat{\mathbf{f}}$ que se encuentra en la ubicación $\hat{\mathbf{z}} = (\hat{x}, \hat{y})$ hace *matching* con una entrada del *codebook* que ha sido observada en la posición (x_k, y_k) de una imagen de entrenamiento (almacenada en la ocurrencia $\mathbb{Z}_p$), se produce el vota en las siguientes coordenadas:

$$x_{voto} = \hat{x} - x_k \text{ y } y_{voto} = \hat{y} - y_k. \tag{4.7}$$

Así, la distribución del voto $\mathcal{P}(\lambda|o_n, \mathbf{C}_p, \hat{\mathbf{z}})$ se obtiene al emitir un voto para cada observación almacenada desde la distribución de ocurrencia aprendida. El conjunto de todos esos votos reunidos, se utiliza para obtener una estimación no paramétrica de la densidad de probabilidad para la posición del centro del objeto.

El puntaje de la hipótesis $h = (o_n, \lambda)$ de una detección de objeto, se obtiene al marginalizar a todos los descriptores que contribuyen a esta hipótesis. Teniendo en cuenta la probabilidad de la hipótesis de un sólo descriptor votado, llegamos a la siguiente ecuación:

$$\mathcal{P}(o_n, \lambda) = \sum_i \mathcal{P}(o_n, \lambda|\hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i)\mathcal{P}(\hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i), \tag{4.8}$$

para $i = 1, \ldots, N_f$, donde, N_f es el número de descriptores útiles, $\mathcal{P}(\hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i)$ es la probabilidad del descriptor $(\hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i)$ siendo muestreada por el detector de punto de interés para el objeto o_n y la ubicación λ. No obstante, tenemos que tolerar pequeñas deformaciones de

la forma, con el fin de ser robusto a la variación intra-clase del objeto. Logramos esta flexibilidad mediante la integración de votos sobre un tamaño fijo de la ventana de búsqueda $W(\lambda)$ mediante una búsqueda Mean-Shift (Comaniciu et al., 2001). En lugar de *keypoints* del *cluster*, con el algoritmo Mean-Shift y para cualquier posición, establecemos que un *keypoint* sólo puede ser unido con otro *keypoint* si se encuentran próximos entre si, de acuerdo con el parámetro de distancia fija. Hemos ajustado este parámetro al 10 % del tamaño del objeto de entrenamiento. Si la decisión se basa en los votos de un solo descriptor y asumimos uniformidad a priori para los descriptores, hemos aproximado la función de probabilidad $\mathcal{P}(o_n, \lambda)$ con la siguiente puntuación, denominada 'score':

$$\text{score}(o_n, \lambda) = \sum_i \sum_{\lambda_j \in W(\lambda)} \mathcal{P}(o_n, \lambda_j | \hat{\mathbf{f}}_i, \hat{\mathbf{z}}_i). \tag{4.9}$$

Para evitar cualquier sesgo sistemático, requerimos que cada descriptor muestreado tenga el mismo peso a priori. Por lo tanto, debemos normalizar el peso de los votos, de manera que $\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}})$ y $\mathcal{P}(\lambda|o_n, \mathbf{C}_p, \hat{\mathbf{z}})$ sumen en total uno. Así, el peso $\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}})$ se propaga de manera uniforme sobre todos los descriptores válidamente interpretados $\mathbf{C}_p$, mediante el ajuste de $\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}}) = \frac{1}{|C^*|}$, donde $|C^*|$ es el número de entradas que hicieron *matching* en el *codebook*. Pero también sería posible hacer que la distribución $\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}})$ refleje el puntaje de los *matching* relacionados. Además de las coordenadas de la votación (x_{voto}, y_{voto}) –ver ecuación (4.7)–, se calcula su peso w, ajustando el peso de la ocurrencia $\mathcal{P}(o_n, \lambda|\mathbf{C}_p, \hat{\mathbf{z}}) = \frac{1}{|\mathbb{Z}_p|}$, donde $|\mathbb{Z}_p|$ es el número de ocurrencias de cada cluster que hizo *matching*. Por lo tanto, un voto se emite en las coordenadas (x_{voto}, y_{voto}) con el peso $w = \mathcal{P}(o_n, \lambda|\mathbf{C}_p, \hat{\mathbf{z}})\mathcal{P}(\mathbf{C}_p|\hat{\mathbf{f}})$. El proceso Mean-Shift y la normalización de los votos se representa en la Figura 4.9(b) y 4.9(c) aplicado en la inspección de imágenes de rayos X.

3) Fusión de candidatos detectados:

El enfoque original de ISM define al candidato con la puntuación más alta como la detección válida. Sin embargo, esto no es necesariamente cierto en las imágenes de rayos X, ya que tienen muchas zonas oscuras que son similares entre sí, y el objeto de interés

pueden tener múltiples posibles orientaciones espaciales al interior de un contenedor, como por ejemplo, un equipaje, es decir, descriptores similares con diferentes ocurrencias. En nuestro caso, encontraron que los candidatos con las puntuaciones más altas se ubican generalmente cerca unos de otros. Es así que proponemos mantener los centros (s_x, s_y) de cada candidato seleccionado, cuya puntuación es mayor que un valor umbral de θ_s. Este valor de umbral se ajusta durante el análisis de desempeño del detector.

El proceso de fusión es similar al procedimiento que hemos explicado en la sección 4.1. En este contexto seguimos dos pasos:

i) *Clustering*: Los centros de cada candidato seleccionado $(> \theta_s)$ se almacenan en $\mathbf{C}_m$. Para cada candidato en $\mathbf{C}_m$, definimos subventanas W_B que tienen por lo menos θ_B *keypoints* de la misma pose. La Figura 4.9(d), 4.9(e) y 4.9(f) muestran ejemplos de estos cuadros azules, W_B. La pose asociada a cada descriptor útil se almacenó en la primera etapa de limpieza de descriptores innecesarios. Hay 25 poses válidas.

ii) *Fusión*: Todas las subventanas W_B que están conectadas o traslapadas, se fusionarán en una nueva subventana más grande W_m (ver rectángulo magenta en 4.9(d) y 4.9(e)).

4) Detección:

La detección final se obtiene de la siguiente manera: En primer lugar, la subventana W_m que encierra el mayor número de candidatos (cuyos centros $(s_x, s_y) \in \mathbf{C}_m$) será seleccionada, si este número es mayor que el umbral θ_m. Si la subventana no cumple esta condición, no se detecta ningún potencial objeto de interés. Con este valor de umbral, validamos la subventana W_m como la detección final. Sin embargo, el tamaño y la orientación de W_m no necesariamente corresponden con el tamaño y la orientación del objeto de interés detectado. A continuación, se estima la orientación de la última ventana de detección W_F, mediante el cálculo del centro de gravedad de los candidatos seleccionados, encerrados en W_m, mediante el uso del conocido algoritmo de clustering *K-means* (MacQueen, 1967), lo que minimiza la distancia entre cada centro (s_x, s_y), utilizando un único centroide $k = 1$. Esto se muestra en la Figura 4.9(f) y 4.9(g), donde el centroide está representado por un

círculo rojo. Con el fin de estimar la orientación de la ventana final W_F, una elipse se ajusta a los *keypoints* contenidos en la subventana seleccionada W_m, como se muestra en la Figura 4.9(f). El ángulo entre el eje mayor de la elipse y el eje horizontal x se representa por ϕ. Este es precisamente el valor de la orientación de la ventana final W_F. La estimación del ángulo se determina para cada objeto de interés, teniendo en cuenta las dimensiones del objeto en la imagen de entrenamiento con mejor vista, es decir, la pose número 13 (ver pose 13 en la Figura 4.7, donde $\alpha = 180°$, $\beta = 0°$ y $\gamma = 180°$). Un ejemplo de detección final orientada del objeto de interés, se muestra en la Figura 4.9(h).

4.2.2. Estimación de Movimiento con Q-Learning

En nuestro trabajo, la estimación de movimiento que permite lograr la siguiente mejor vista del objeto de interés, se basa en el algoritmo de aprendizaje por refuerzo *Q-learning*, descrito en la sección 2.6. Este algoritmo encuentra la política óptima de giros que debe realizar el manipulador robótico para lograr la mejor vista en la menor cantidad de movimientos posibles. Después de aplicar la política basada en *Q-learning*, será solicitada la adquisición de una nueva imagen de rayos X (desde un nuevo punto de vista) sobre la cual se aplicará el detector del objeto de interés, el cual evaluará, si la imagen contiene al objeto buscado, y si este se encuentra en una buena pose. En caso de no ser así, o bien, no se encuentre el objeto de interés en esta nueva imagen (ND: No Detectado), se repite el ciclo, evaluando una nueva pose determinada las rotaciones del manipulador robótico, dadas por *Q-learning*, para luego adquirir una nueva imagen y repetir el proceso completo, hasta tener suficiente confianza en la detección del objeto de interés.

Hemos establecido como política de término, realizar tres intentos sucesivos en los cuales no se encuentre el objeto de interés buscado y/o cuatro detecciones del objeto de interés en una malas poses. En ambos casos se deja de buscar el objeto y se presume que no está presente o no fue posible decir con certeza que la detección corresponde al objeto buscado, terminando de esta forma la búsqueda y ejecución del algoritmo con un estado de no detección (ND).

La estimación de movimiento para lograr la siguiente mejor posición del objeto de interés puede ser resumida en cuatro etapas principales, las cuales se describen a continuación:

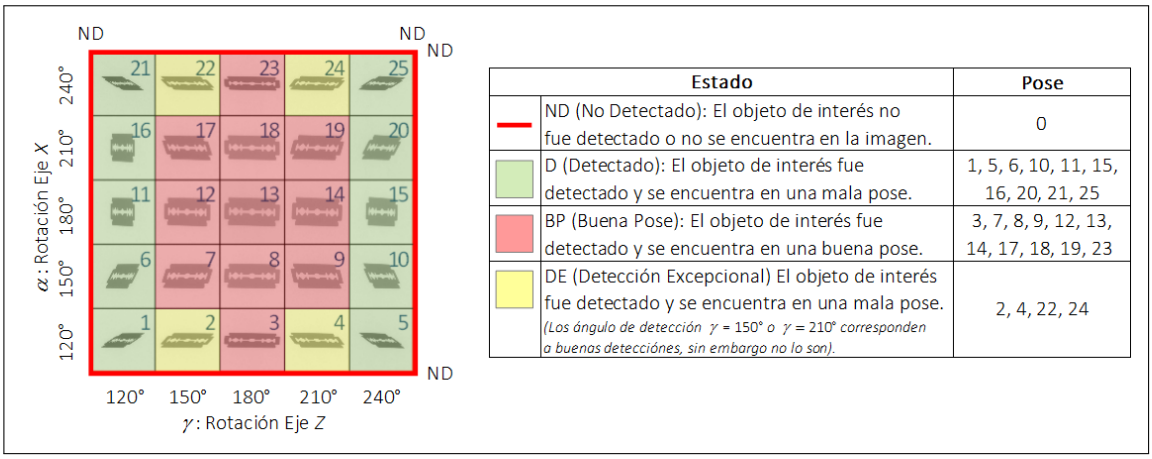
A. Detección de Objeto de Interés y Estimación de Pose

En esta etapa se hace uso del detector descrito en la sección 4.2.1. A este detector, que analiza las imágenes de rayos X, le hemos incorporado un estimador de pose que define como pose válida del objeto de interés, a la pose asociada a los descriptores SIFT $\hat{\mathbf{f}}$ que hicieron *match* con los descriptores SIFT $\mathbf{f}_k$ almacenados en el proceso de entrenamiento, con mayor frecuencia y que se encuentren contenidos en la subventana W_m. Si el objeto de interés es detectado, el detector entregará como salida la pose estimada en la cual éste se encuentra. En caso de no detectar el objeto de interés, el detector entregará un valor de pose predefinido igual a cero, el cual representará un estado de no detección ND. En ambos casos, ya sea el número de la pose estimada o el valor fijo de cero, será el parámetro de entrada a utilizar en el siguiente proceso de inspección.

B. Entrenamiento

La fase de entrenamiento de nuestro algoritmo se realiza de manera *off-line*. Para el entrenamiento utilizamos una base de datos de imágenes rayos X de una hoja de afeitar y una base de datos de imágenes de revólveres, en distintas posiciones en los tres ejes del sistema cartesiano (ver en Figura 4.7 base de datos de imágenes de una hoja de afeitar). A partir de estas imágenes determinamos los distintos estados para el entrenamiento de *Q-Learning*. Cabe señalar que debido a las distintas rotaciones α, β y γ se producen replicas de los cuadrantes, por lo cual, para el proceso de entrenamiento sólo se utilizan las imágenes del cuadrante principal, es decir, las imágenes comprendidas entre los ángulos α: [120°, 150°, 180°, 210°, 240°] y γ: [120°, 150°, 180°, 210°, 240°], para los cuatro ángulos β: [0°, 30°, 60°, 90°], así definimos como un conocimiento a priori las vistas que serán consideradas

'Buenas Poses' (BP) y que el algoritmo debiese encontrar (ver en Figura 4.10 las diferentes poses y estados asociados de una hoja de afeitar).



Estado	Pose
ND (No Detectado): El objeto de interés no fue detectado o no se encuentra en la imagen.	0
D (Detectado): El objeto de interés fue detectado y se encuentra en una mala pose.	1, 5, 6, 10, 11, 15, 16, 20, 21, 25
BP (Buena Pose): El objeto de interés fue detectado y se encuentra en una buena pose.	3, 7, 8, 9, 12, 13, 14, 17, 18, 19, 23
DE (Detección Excepcional) El objeto de interés fue detectado y se encuentra en una mala pose. *(Los ángulo de detección $\gamma = 150°$ o $\gamma = 210°$ corresponden a buenas detecciónes, sin embargo no lo son).*	2, 4, 22, 24

Figura 4.10. Detalle de un sector de la base de datos de entrenamiento de una hoja de afeitar ($\beta = 0°$), utilizado en *Q-Learning* para definir los estados. Los mismos estados aplican para la base de datos del revolver.

A partir de las 25 poses asociadas a cada imágenes, incluyendo además la No Detección (ND), se procede a modelar el ambiente *no continuo* o *discreto* (debido a que las imágenes se encuentran rotadas 30° entre ellas): Agrupamos las distintas imágenes con su pose, considerando el ángulo γ y le asignamos un nombre, tal como se muestra en la Figura 4.10, comenzando por D (Detectado, D1 y D2), BP (Buena Pose), DE (Detección Excepcional, DE1 y DE2) y ND cuando no ocurre detección (No Detectado). Podemos simplificar esta representación mediante un gráfico (ver Figura 4.11), el cual es una colección de nodos y curvas. La longitud y rectitud de las curvas (enlaces) no tiene importancia para el modelo. Del mismo modo, el tamaño y la forma del nodo, tampoco tiene importancia. Representamos cada grupo de imágenes como un nodo y cada rotación γ como un enlace. Así, nuestro gráfico queda diseñado con 6 nodos. Observar que los enlaces representan las rotaciones que serán realizadas para llegar a una Buena Pose, BP.

Se establece un nodo objetivo (BP), de tal forma que desde cualquier nodo se realicen las rotaciones γ necesarias para llegar a este nodo objetivo. Para establecer este tipo de objetivo, introducimos un valor de recompensa para cada rotación. Las rotaciones que conducen inmediatamente al nodo objetivo, BP, tienen recompensa inmediata de 100 (ver

en flechas rojas en Figura 4.11). Otras rotaciones que no tienen conexión directa con el nodo objetivo, tienen recompensa cero. Finalmente el gráfico utilizado para el algoritmo de entrenamiento de *Q-Learning* recibe el nombre de diagrama de estados.

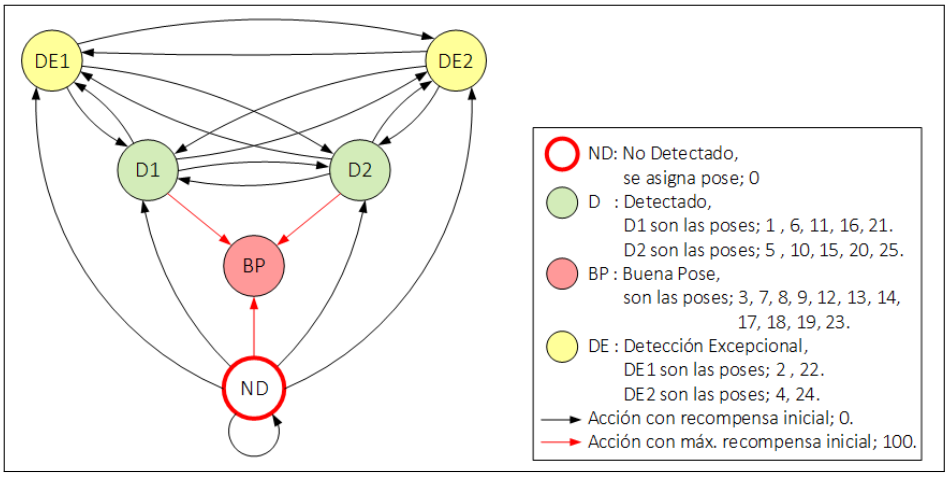


Figura 4.11. Esquema de los estados utilizados en *Q-Learning*.

Este diagrama de estados, permitió el desarrollo del algoritmo de *Q-Learning*, y así utilizar este método como base para la estimación del siguiente movimiento y en definitiva la búsqueda de una buena vista de algún objeto de interés en imágenes de rayos X. El entrenamiento consta básicamente de tres pasos que se describen a continuación:

a) **Definición de poses de manera visual:** A partir de la base de datos de imágenes de rayos X del proceso de entrenamiento (ver Figura 4.7), definimos visualmente y usando conocimiento a priori, las poses del objeto de interés que serán consideradas buenas vistas (o buenas poses BP). El manipulador robótico sólo permite rotación en el eje Z, por lo cual el ángulo estimado por el algoritmo *Q-Learning* será γ. Debido a esta restricción, para cada ángulo α debe existir a lo menos un estado de buena pose BP, aun cuando visualmente no lo sea (pose 3 y 23), de lo contrario el algoritmo entraría en un ciclo sin fin. Las detecciones excepcionales corresponden a malas poses (pose 2, 4, 22 y 24) cuyo ángulo γ coincide con algunas buenas vista. Esto se muestra en la Figura 4.10.

b) **Definición de matriz R:** Se denomina matriz R, al arreglo de datos en el cual se almacenan las recompensas (estímulos positivos) o castigos (estímulos negativos), r, que recibe el agente del algoritmo desde el ambiente. Las recompensas le indican al agente que la acción realizada es correcta, mientras que en el caso contrario las recompensas negativas o castigos, le indican que ha realizado una mala acción. Se definen las recompensas como el valor numérico que recibe el agente al realizar una transición desde un estado a otro ($s \rightarrow s$'). La máxima recompensa de cien ($+100$) se asigna a la matriz R cuando se lograr una transición hacia un estado objetivo de buena pose (BP) y para otras transiciones, se asigna a la matriz R un valor mínimo de cero.

c) **Definición de la matriz Q:** La matriz Q, es el arreglo de datos, en el cual se almacenan los valores correspondientes al par *acción-estado* y se obtiene luego del entrenamiento. En esta matriz se almacenan de forma numérica todos los valores correspondientes a las distintas acciones que se pueden tomar en cada estado. De esta forma, y a partir de estos valores, se puede buscar la ruta óptima siguiendo los valores más altos, para las distintas transiciones desde un estado inicial hacia un estado final o estado objetivo. La matriz Q, inicialmente es una matriz de cero, la cual se va actualizando a través de las iteraciones de *Q-learning*, hasta converger en su valor final. En la Figura 4.12 se aprecian las matrices Q y R utilizadas en nuestro enfoque, además de la ecuación de *Q-Learning* para estados discretos.

La ecuación (2.23) de *Q-Learning* para estados discretos vista en la sección 2.6, se reduce a la ecuación (4.10), debido a que la tasa de aprendizaje α para modelos determinista es igual a 1. Además, como se aprecia en la Figura 4.12, al parámetro η o factor de descuento, le hemos asignado un valor de 0.8 para favorecer la búsqueda de recompensas futuras, privilegiando la exploración sobre la explotación de la información que ya posee (se dice que con un factor de descuento $\eta = 0$ el agente es miope, ya que solo utiliza la información que conoce).

$$Q[s, a] \leftarrow R + \eta \max_{a'} Q[s', a']. \tag{4.10}$$

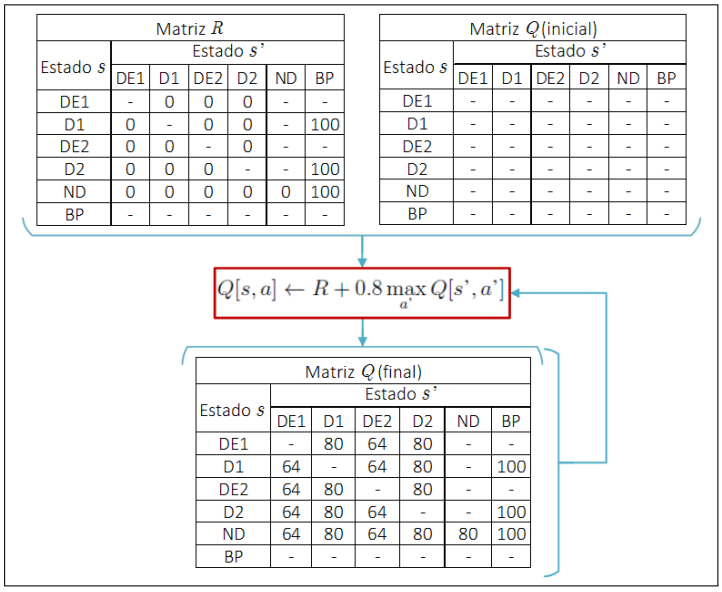


Figura 4.12. Representación algorítmica para obtener la matriz Q final a partir de la matriz R y matriz Q inicial.

Algoritmo 1 Algoritmo de *Q-Learning*

Definiciones: estado = pose, estado objetivo = buena pose
Dado: Diagrama de estados con un estado objetivo (representado por una matriz de recompensa R)
Encontrar: Trayectoria mínima de un estado inicial al estado objetivo (representada por la matriz Q)

Entradas: Factor de descuento η y la matriz de recompensa R;
Salida: Matriz de trayectorias Q;

1: Inicializar en cero la matriz Q;
2: **para** cada estado **hacer**
3: Seleccionar aleatoriamente un estado;
4: **mientras** no se alcance el estado objetivo **hacer**
5: Seleccionar una de entre todas las posibles acciones para el estado actual;
6: Usando esta posible acción, considerar ir al siguiente estado;
7: Obtener el valor máximo de Q de este siguiente estado basado en todas las acciones posibles;
8: Calcular: Q(estado, acción) $= R$(estado, acción) $+ \eta\, max[Q$(siguiente estado, todas las acciones)$]$;
9: Establecer el siguiente estado como el estado actual;
10: **fin mientras**
11: **fin para**

El número de iteraciones necesarias para hacer converger la matriz Q y el tiempo de ejecución necesario para entrenamiento tienen un costo computacional bajo, debido a que nuestro modelo es relativamente pequeño (el modelo propuesto es el mismo tanto para la hoja de afeitar como para el revolver). De esta forma, el número de iteraciones y el tiempo de ejecución en el proceso de estimación de Q no son factores o métrica críticas de medir. Hemos ajustado el número de iteraciones a 1000 como valor máximo, sin embargo, en

nuestros experimentos, el número de iteraciones necesarias para obtener Q nunca supero dicho valor. Debido a la aleatoriedad intrínseca del algoritmo al comenzar desde un estado inicial y llegar a otro estado, siguiendo la máxima recompensa, el número de iteraciones necesarias es variable. El algoritmo que permite obtener Q, se detiene cuando ya no existen variaciones entre los valores presentes y anteriores. El Algoritmo 1 expresado como pseudocódigo muestra el proceso completo de obtención de la matriz Q a partir del factor de descuento η y la matriz de recompensa R.

Una vez entrenado el algoritmo, se asegurará la eficiencia en la búsqueda, a través de la ruta óptima, de forma similar a lo que realizaría un operario humano, sin embargo, como la eficacia del 100 % no está asegurada en el detector, por variadas razones (oclusión, no detección, falsos positivos, entre otros), consideraremos una etapa para la integración del detector y el estimador de movimiento.

C. Estimación de Movimiento

En esta etapa se recibe 'la pose' del objeto como parámetro de entrada que entrega el detector (en caso de ser detectado), o un valor cero en caso de no detección. Si hubo detección, se analiza la pose del objeto, pudiendo darse dos situaciones: *i*) que haya una buena vista, es decir, una pose que permita afirmar visualmente y con certeza que el objeto de interés buscado se encuentra efectivamente en la imagen analizada, terminando de esta forma el proceso de detección, *ii*) que el objeto de interés detectado se encuentra en una mala pose, es decir, no se puede afirmar con certeza que en la imagen se encuentra el objeto de interés buscado. En este caso mediante nuestra propuesta de estimación de movimiento con *Q-Learning*, determinaremos una ruta optima, que se obtiene a partir del entrenamiento de *Q-Learning*, es decir, los grados que debe girar el manipulador robótico para que el objeto contenedor quede en una posición que permita obtener una imagen en la cual el objeto de interés se encuentre en una mejor vista (buena pose, BP). En los casos de imágenes de rayos X donde efectivamente se detectó el objeto de interés buscado, pero éste no se encuentra en una buena pose, es necesario determinar cuánto se debe girar el

objeto para encontrar una posible buena vista. Para esto hacemos uso de nuestro algoritmo basado en *Q-Learning* (ver Algoritmo 2), se determinará a partir de la pose actual del objeto, la ruta óptima para llegar a una pose objetivo. Posteriormente y basado en esta ruta, se calcula el ángulo de giro necesario para obtener una nueva imagen donde es más probable encontrar al objeto de interés en una buena pose, reduciendo de esta manera al mínimo la cantidad de imágenes intermedias. Finalmente, esta nueva imagen será analizada con el detector de objetos, donde se determinará si el objeto de interés está o no en una buena pose.

Algoritmo 2 Algoritmo para utilizar la matriz Q
Definiciones: estado = pose, estado objetivo = buena pose
Entradas: Matriz Q, estado inicial; **Salida:** Secuencia del estado actual desde el estado inicial hasta el estado objetivo;
1: Establecer: estado actual = estado inicial; 2: Desde el estado actual, encontrar acción que produce el máximo valor de Q; 3: Establecer: estado actual = siguiente estado; 4: **ir hacia** 2 **hasta que** estado actual = estado objetivo;

D. Integración de Detector y Algoritmo de Estimación de Movimiento

En esta etapa se integrarán de manera secuencial ambos algoritmos (detector y estimador de movimiento), con el fin de lograr mejores resultados en la detección. Aquí, intentamos resolver las problemáticas típicas de cualquier detector, como son la oclusión parcial, la dificultad para estimar si está el objeto en la escena y la anulación de falsos positivos.

Podemos indicar que *Q-Learning* permite modelar las incertezas producidas por: a) las imperfecciones del detector de objetos de interés al detectar y estimar la pose, lo cual ocurre principalmente por la simetría y/o disposición del objeto de interés al ser caracterizado, provocando imágenes de tipo espejo que dificultan la estimación de la real pose durante la inspección, b) la restricción de movimiento del manipulador robótico, que hará imposible alcanzar algunas buenas poses, y c) las réplicas de los cuadrantes de la base de datos de imágenes de rayos X para el entrenamiento, que hace confundir al estimador de pose, (ver Figura 4.13). Un modelo de estimación de movimiento de tipo heurístico, donde sólo

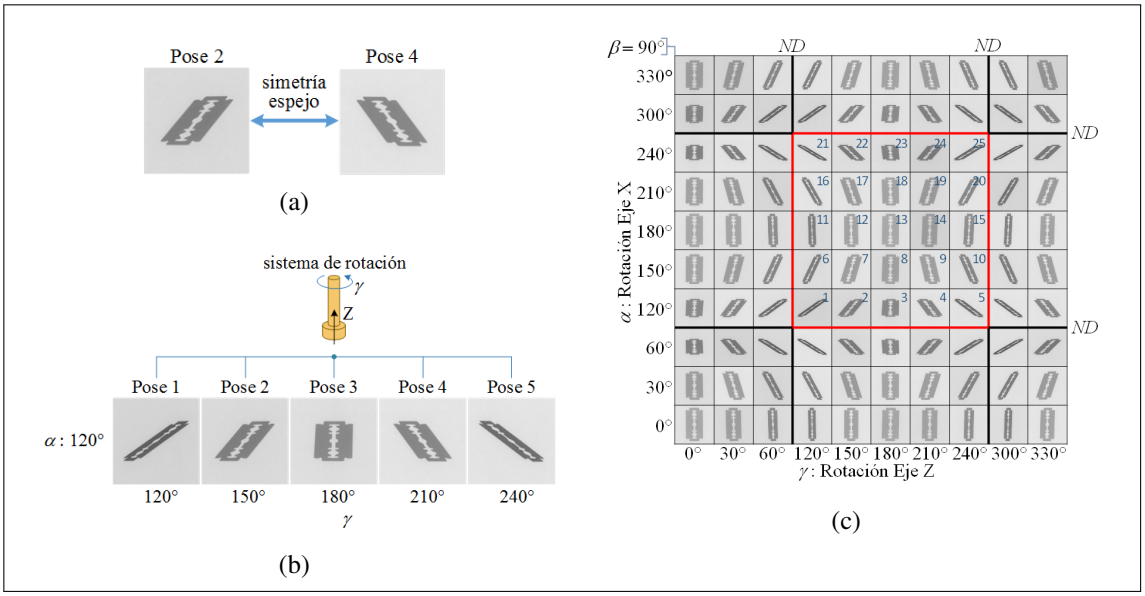


Figura 4.13. Origen de las incertezas que el modelo propuesto de *Q-Learning* incorpora: a) incerteza producida por la simetría espejo que provoca estimaciones erróneas de poses, b) incerteza producida por las limitaciones de movimiento del manipulador robótico que impide alcanzar una buena pose, y c) incerteza producida por las replicas de los cuadrantes de las imágenes de entrenamiento que impide saber la posición real del objeto de interés.

sea necesario un movimiento para alcanzar una buena pose (BP) desde una no detección (ND), una detección en una mala pose (D), o una detección excepcional (DE), no considera las incertezas ni restricciones como lo hace *Q-Learning*, para así evitar algunos niveles de oclusión y estimar el siguiente movimiento que permite encontrar más rápidamente al objeto de interés en una buena pose.

4.2.3. Eliminación de Falsas Alarmas

Una vez aplicado el algoritmo de detección de objetos de interés (objetos amenazantes) sobre una imagen de rayos X, pueden ocurrir algunas de las siguientes situaciones: *i*) el objeto de interés es detectado sin ninguna falsa alarma (falso positivo), *ii*) el objeto de interés es detectado pero aparecen una o más falsas alarmas, *iii*) el objeto de interés no es detectado, pero si aparecen falsas alarmas, y *iv*) el objeto de interés no es detectado y no aparecen falsas alarmas. En las tres primeras situaciones mencionadas es importante validar

la detección y eliminar las falsas alarmas. Las falsas alarmas normalmente corresponden a hechos desafortunados, fortuitos y que no permanecen en una secuencia de imágenes, adquiridas desde diferentes puntos de vista. En un sistema de inspección activa, hay que provocar movimientos en un manipulador robótico para poder lograr la detección del objeto en la mejor pose, dicho movimiento está basado en la en la detección del objeto de interés y en la pose estimada, es por esto que no sería correcto realizar movimientos asumiendo una falsa alarma como una detección válida.

Para validar las detecciones y eliminar las falsas alarmas proponemos aplicar un algoritmo de seguimiento (tracking) de las detecciones (Mery y Filbert, 2002), el cual se basa en la idea que si el objeto de interés aparece en una vista, es probable que aparezca en las siguientes vistas, y por el contrario, las falsas alarmas, por ser hechos fortuitos, probablemente no volverán a aparecer en una secuencia de imágenes, y si lo hacen, será en lugares erráticos e impredecibles. Como nuestro sistema cuenta con el modelo geométrico, es posible aplicar la teoría bifocal y trifocal (Hartley y Zisserman, 2003) para predecir donde debiese aparecer la detección en la siguiente vista.

A continuación se presenta la metodología para realizar un seguimiento de detecciones en una secuencia de imágenes de rayos X. El método propuesto consiste en hacer el seguimiento de las detecciones en una secuencia limitada de imágenes (en nuestro *framework*, para lograr detectar al objeto de interés en una buena pose, hemos propuesto adquirir hasta cuatro imágenes de rayos X). No se sabe a priori cuáles son las detecciones correspondientes, sólo se conoce la posición de la detección y la geometría de la proyecciones de las vistas, es decir se trata de una secuencia calibrada. Es importante señalar que este método se aplica en forma activa, es decir no contamos con la secuencia completa de imágenes, sino más bien, se aplica a medida que avanza la adquisición de imágenes, desde la segunda imagen en adelante.

En cada imagen aparecerán algunas detecciones hipotéticas y es necesario establecer cuáles de ellas son detecciones correctas y cuáles son falsas alarmas. El seguimiento de las detecciones hipotéticas es un buen criterio para la eliminación de las falsas alarmas, de esta

manera se establece que sólo las detecciones hipotéticas que pueden ser seguidas a medida que avanza la secuencia, son consideradas como detecciones correctas.

A. Seguimiento en Dos Vistas

Para llevar a cabo el seguimiento de las detecciones en las dos imágenes, se debe establecer cuáles detecciones son correspondientes. Esta correspondencia se determina evaluando la restricción epipolar explicada en la sección 2.5, la cual dice: para que $\mathbf{m}_i$ y $\mathbf{m}_j$ sean puntos correspondientes, el punto $\mathbf{m}_j$ debe estar en la línea epipolar generada por $\mathbf{m}_i$ (ver Figura 4.14). Siendo esto último una condición necesaria, pero no suficiente. Esta condición 'no suficiente' permite de todas formas desestimar falsas alarmas, las cuales no debiesen estar sobre o próximas a la línea epipolar generada por $\mathbf{m}_i$. En este caso $\mathbf{m}_i$ y $\mathbf{m}_j$ corresponden a los centroides de las detecciones en la imágenes $\mathbf{I}_i$ e $\mathbf{I}_j$ respectivamente.

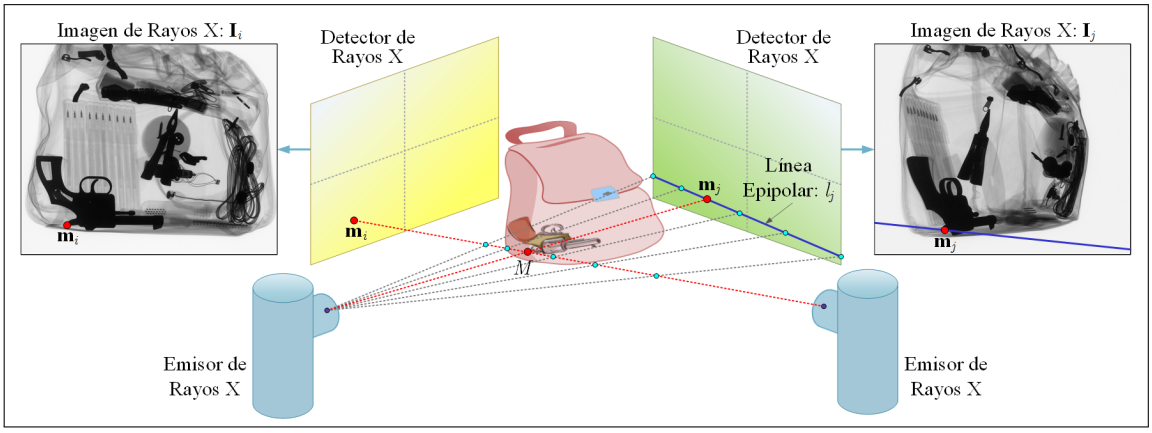


Figura 4.14. Geometría epipolar para establecer correspondencia entre los puntos $\mathbf{m}_i$ y $\mathbf{m}_j$.

Debido a las características del detector, el centroide de la detección no siempre está ubicado en el centro de masa del objeto de interés, y también, debido a errores de calibración; es que $\mathbf{m}_j$ no estará exactamente sobre la línea epipolar l_j, pero si muy cercano a ella. Así el criterio de correspondencia será la restricción epipolar práctica, establecido en la ecuación (2.19), la cual indica que una detección será válida y estará en correspondencia con la detección de la imagen anterior, si la distancia del centroide $\mathbf{m}_j$ de la detección es

menor a un cierto valor de umbral d_0. De esta forma podemos desestimar todos las detecciones hipotéticas que están sobre dicho valor d_0 (mas alejadas de la linea epipolar), y ser consideradas falsas alarmas.

Debido al menor tamaño de la hoja de afeitar y a la ubicación del centroide de las detecciones (centroide de la ventana de detección W_F), hemos aplicado esta metodología sólo en la detección de hojas de afeitar. El detector AISM que hemos propuesto, tiene definido el tamaño de la ventana de detección W_F en 360×200 pixeles, de esta forma, para considerar que una detección sea válida, hemos definido el valor de umbral $d_0 = 100$ pixeles. Quedando la restricción epipolar como:

$$\frac{|\mathbf{m}_j^\top \mathcal{F}_{ij} \mathbf{m}_i|}{\sqrt{a_1^2 + a_2^2}} < 100. \tag{4.11}$$

***B*. Seguimiento en Tres Vistas**

Si ocurre que transcurrida la adquisición y evaluación de dos imágenes, no se ha detectado en una buena pose al objeto de interés, es necesario entonces adquirir una tercera imagen $\mathbf{I}_k$. A partir de esta adquisición y luego de aplicar el algoritmo de detección, se deberá establecer la correspondencia entre los centroides de las detecciones $\mathbf{m}_i$, $\mathbf{m}_j$ y $\mathbf{m}_k$ en las imágenes $\mathbf{I}_i$, $\mathbf{I}_j$ y $\mathbf{I}_k$ respectivamente, y en el caso que exista la correspondencia en la tercera imagen, podríamos desestimar otros posibles centroides de detecciones hipotéticas, que no cumplen con (2.21) y considerarlas falsas alarmas.

Basándose en la geometría epipolar, se puede afirmar que si se calcula la línea epipolar de $\mathbf{m}_i$ y la línea epipolar de $\mathbf{m}_j$ en la tercera imagen $\mathbf{I}_k$, $\mathbf{m}_k$ debe estar en la intersección de ambas líneas epipolares, ya que si $\mathbf{m}_i$ y $\mathbf{m}_k$ son correspondientes $\mathbf{m}_k$ debe estar en la línea epipolar de $\mathbf{m}_i$ en la tercera imagen. La misma deducción se puede hacer para $\mathbf{m}_j$, entonces $\mathbf{m}_k$ debe pertenecer a ambas líneas epipolares, es decir $\mathbf{m}_k$ es el punto de intersección de las líneas epipolares. La Geometría Epipolar en tres imágenes señala entonces, que $\mathbf{m}_i$, $\mathbf{m}_j$ y $\mathbf{m}_k$ son puntos correspondientes si $\mathbf{m}_k$ coincide con el punto de intersección de las líneas epipolares de $\mathbf{m}_i$ y $\mathbf{m}_j$ en la tercera imagen (Faugeras y Robert, 1996). Esta es una condición 'necesaria y suficiente'. Sin embargo, el punto de intersección no está definido

si ambas líneas epipolares son iguales, sumado a la desventaja que la *Geometría Epipolar* no proporciona un método directo para analizar la correspondencia de tres puntos, ya que es necesario calcular dos líneas epipolares y luego su intersección.

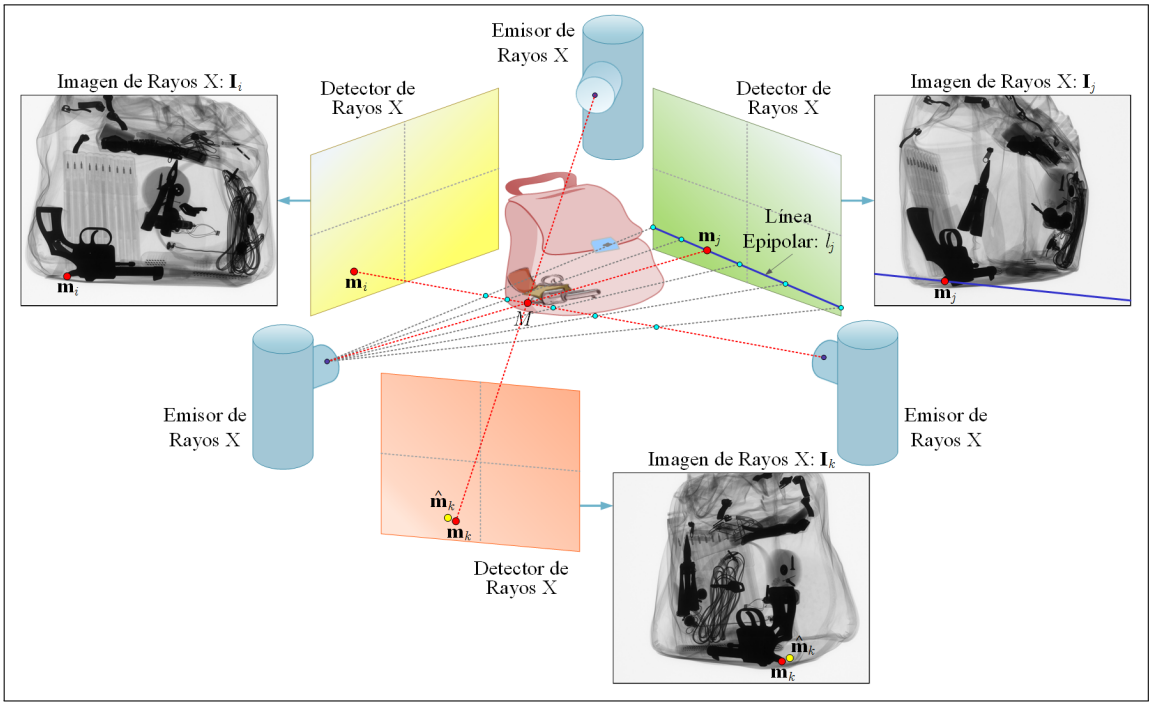


Figura 4.15. Geometría trifocal para establecer correspondencia entre los puntos $\mathbf{m}_i$, $\mathbf{m}_j$ y $\mathbf{m}_k$, a partir de la estimación de $\hat{\mathbf{m}}_k$ (centroide amarillo).

Por esta razón, hemos hecho uso de los *tensores trifocales* definidos en la sección 2.5, los cuales puede evitar la singularidad indicadas anteriormente (líneas epipolares iguales) y obtener una solución directa para la correspondencia en tres vistas. Es decir, se establece la correspondencia entre los centroides de las detecciones $\mathbf{m}_i$ y $\mathbf{m}_j$ en las imágenes $\mathbf{I}_i$ e $\mathbf{I}_j$ respectivamente, usando la *teoría epipolar*, para luego, a partir de estos puntos y los *tensores trifocales* $\mathbf{T}^{jk} = [T_1^{jk}\ T_2^{jk}\ T_3^{jk}]^\top$ para $j, k = 1, 2, 3$, estimar la posición $\hat{\mathbf{m}}_k$ de la detección hipotética en la tercera imagen $\mathbf{I}_k$, tal como se indica en la ecuación (2.20). De esta forma, será considerada como detección válida, a la detección hipotética cuyo centroide $\mathbf{m}_k$ este próximo a $\hat{\mathbf{m}}_k$ (ver Figura 4.15), es decir si cumple con la restricción $\|\hat{\mathbf{m}}_k - \mathbf{m}_k\| < d_1$. En nuestro caso el valor de umbral de distancia $d_1 = 100$. Así, las

demás detecciones hipotéticas que no cumplieron con esta restricción, serán desestimadas y declaradas como falsas alarmas.

Nuestra propuesta de *framework* define que es posible adquirir hasta cuatro imágenes ($C_{max} = 4$), para así detectar al objeto de interés en una buena pose. Para lograr establecer la correspondencia en esta cuarta imagen se utiliza, este último método (tensores trifocales), con los centroides de las detecciones en correspondencia en las imágenes $\mathbf{I}_i$ e $\mathbf{I}_j$ o en las imágenes $\mathbf{I}_j$ e $\mathbf{I}_k$. A modo de ejemplo, ver en Figura 4.16 el seguimiento de una hoja de afeitar detectada en 4 imágenes.

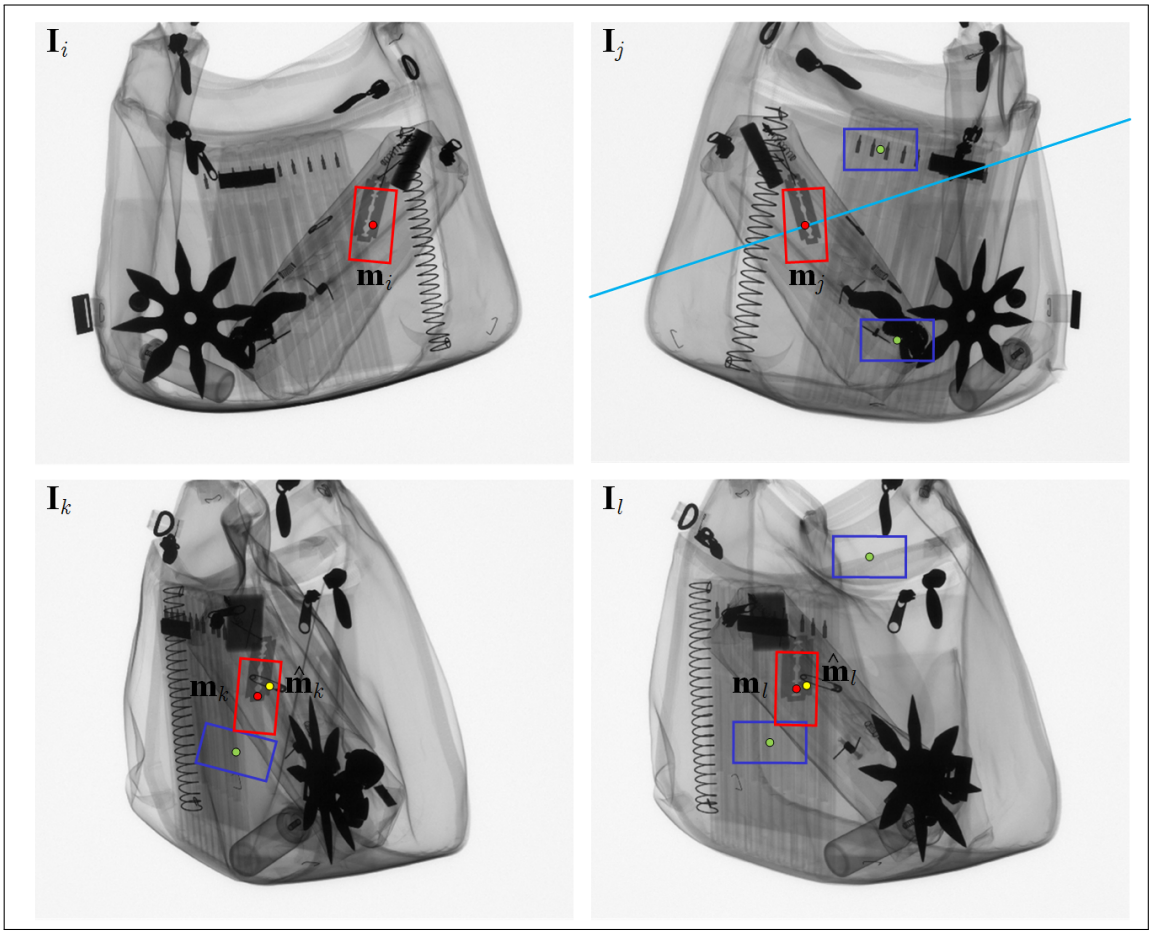


Figura 4.16. Seguimiento de una hoja de afeitar; Centroide rojo con recuadro rojo: Detección válida, Centroide verde con recuadro azul: Detección desestimada y considerada falsa alarma (Falso Positivo, FP), Línea celeste: Línea epipolar, Centroide amarillo: Punto estimado mediante tensores trifocales.

Capítulo 5. RESULTADOS EXPERIMENTALES

La mayoría de los sistemas comerciales de inspección con rayos X que permiten encontrar objetos de interés al interior de algún objeto contenedor, están basados en algoritmos de pseudocolor y algoritmos de mejoramiento de la calidad de imágenes, para que un inspector humano tome la decisión de señalar si se encuentra o no el objeto de interés que buscaba. Por otra parte, los sistemas automáticos o semi-automáticos se encuentran aún en etapas de investigación. Para avanzar en el estado del arte, y debido a que nuestra propuesta de investigación de visión activa con rayos X no tiene precedentes, hemos tenido que diseñar un sistema modelado geométricamente para adquirir imágenes de entrenamiento y pruebas. De esta forma hemos implementado algunas bases de datos de imágenes de rayos X, para la etapa de pruebas, en escenarios realistas, donde se incorporan altos grados de oclusión y desorden. Adicionalmente hemos creado bases de datos de objetos de interés considerados objetos peligrosos: Hoja de afeitar, *shurikens* (estrellas ninja) y revólveres, en donde cada objeto se encuentra aislado y radiografiado desde diferentes puntos de vista. Nuestra metodología hizo necesario establecer una forma de medir el desempeño del sistema en su conjunto, así como también usar las métricas clásicas para la etapa de detección en una vista. Los detalles del hardware, de implementación, y mediciones de desempeño se describen en este capítulo.

5.1. Criterio de Evaluación

Para evaluar el desempeño del sistema de visión activa, que permite la detección de objetos de interés en una buena pose, hemos usado dos estrategias: A) validación visual de la detección, y B) validación automática de la detección. En ambas estrategias hemos usado los indices de desempeño: Precision y Recall, los cuales describiremos posteriormente.

A. Validación Visual de la Detección

En esta estrategia se asume el uso en tiempo real del *framework*, y en la última inspección activa, se decide visualmente si se ha logrado la detección o si corresponde a un

falso positivo. Por ende las imágenes no están etiquetadas, en lo que comúnmente se llama *ground truth*.

B. Validación Automática de la Detección

En esta estrategia se utilizan secuencias de imágenes pre adquiridas, las cuales se anotan manualmente con cuadros delimitadores llamados, "etiquetas" (*ground truth*), que describen la posición en la imagen donde se encuentra el objeto de interés. Del mismo modo el sistema de detección automática, da como resultado de salidas, una o varias detecciones en términos de cuadros delimitadores alrededor de la ubicación de la(s) detección(es) hipotética(s). Los dos conjuntos de cuadros delimitadores, etiquetas y detecciones hipotéticas, (ambos llamados *bounding boxes*, BB_{gt} y BB_{dt}, respectivamente) son comparados en el procedimiento de evaluación de desempeño del algoritmo de detección. Así, la decisión final de considerar imagen por imagen si se se ha logrado la detección o si se trata de un falso positivo (falsa alarma), la determina automáticamente el algoritmo.

Consideramos que ambas estrategias son válidas, ya que en secuencias pre-adquiridas de imágenes de rayos X, realizadas a objetos complejos, donde muchas veces existe un alto grado de oclusión y desorden, a menudo no es posible decidir si un objeto debe ser etiquetado, aun sabiendo a priori que el objeto se encuentra en la imagen. Por lo tanto, decidimos etiquetar todos aquellos objetos de interés que un humano podría detectar claramente, a simple vista y sin tener que recurrir al razonamiento. Como consecuencia, sólo fueron etiquetados los objetos de interés donde al menos alguna parte discriminativa de su forma era visible. Por un lado, esto significa que un buen detector podría ocasionalmente detectar objetos que no están etiquetados, y por otra parte, un número significativo de objetos etiquetados están fuertemente ocluidos que sería poco realista esperar que cualquier algoritmo actual logre tasa de detección del 100 % con sólo un pequeño número de falsos positivos. De esta forma, la validación visual cobra real importancia, ya que permite decidir situaciones de detección donde un algoritmo no sería capaz. Independiente de la

estrategia de validación, para evaluar los resultados de aplicar nuestro *framework*, no sólo nos preocupamos de si ocurrió o no la detección de forma binaria (sí o no), sino que también nos interesa el nivel de superficie detectada, para así definir si una detección ocurre correctamente, y como ya hemos indicado, este análisis puede ser realizado visual o automáticamente.

El desempeño de nuestro método para medir de forma automática la detección en una vista se realiza utilizando el criterio de evaluación de la calidad "PASCAL Visual Object Classes Challenge" (Everingham et al., 2010), el cual mide la cantidad de solapamiento (área de overlap) que tiene la etiqueta BB_{gt} por él cuadro de detección hipotética BB_{dt}, como se muestra en la Figura 5.1, donde una detección se considera correcta si el área normalizada de solapamiento a_o entre el cuadro de detectado BB_{dt} y el cuadro de etiqueta BB_{gt} excede 0,5, donde a_o se define en la ecuación (5.1).

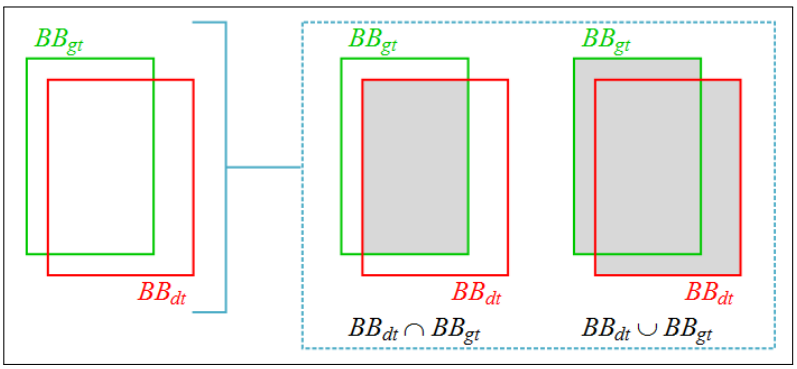


Figura 5.1. Criterio de evaluación que compara los cuadros BB_{gt} y BB_{dt}: Interpretación del criterio de área de solapamiento.

$$a_o = \frac{area(BB_{dt} \cap BB_{gt})}{area(BB_{dt} \cup BB_{gt})}. \tag{5.1}$$

donde, $BB_{dt} \cap BB_{gt}$ es la intersección de los cuadros de detección y etiqueta (ground truth), y $BB_{dt} \cup BB_{gt}$ su unión.

Al comparar la detección hipotética con la etiqueta, de acuerdo al criterio anterior, tenemos que considerar cuatro casos. Por un lado, la etiqueta de un objeto de interés, que se valida o no se valida por la detección hipotética. Y por otro lado, la estructura del fondo,

que puede ser confundida o no, con un objeto de interés. Así, llamaremos a una detección hipotética com una detección *Verdadera Positiva* TP (*True Positive*), y a partir de esto usaremos la terminología mostrada en la Table 5.1. Como de alguna forma ya se ha argumentado, sólo una detección hipotética por objeto se acepta como correcta (la de mayor a_o) y las otras detecciones hipotéticas adicionales sobre el mismo objeto de interés se cuentan como una detección *Falsa Positiva* FP (*False Positive*).

Tabla 5.1. Matriz de confusión para detecciones hipotéticas.

Clase Detectada / Clase Actual	Objeto de Interés	No es Objeto de Interés
Objeto de Interés	Verdadero Positivo (TP)	Falso Negativo (FN)
No es Objeto de Interés	Falso Positivo (FP)	Verdadero Negativo (TN)

Al medir el desempeño de cualquier detector de objetos de interés, las dos cantidades más importantes son el numero de detecciones correctas TP, que se desea maximizar, y el número de detecciones falsas FP, que se se desea minimizar. La mayoría de los algoritmos de detección incluyen un parámetro de umbral que puede variar, para situarse en diferentes puntos de equilibrio entre las correctas y falsas detecciones (Agarwal et al., 2004). Con el fin de cuantificar correctamente y caracterizar el desempeño de un algoritmo de detección de objetos se han propuesto varias medidas en la literatura. A continuación, vamos a describir brevemente las dos medidas más populares: ROC y Precision-Recall.

5.1.1. ROC

La característica de funcionamiento del receptor (*Receiver-Operating-Characteristic*: ROC) tiene su origen en la teoría de detección de señal y ha sido desarrollado para medir el rendimiento de los clasificadores binarios. ROC muestra el equilibrio entre la sensibilidad (tasa de verdaderos positivos, TPR: *True Positive Rate*) y la especificidad (tasa de falsos positivos, FPR: *False Positive Rate*). La tasa de verdaderos positivos y la tasas de falsos

positivos se calculan con las siguientes ecuaciones:

$$\text{tasa de verdaderos positivos, TPR} = \frac{\text{TP}}{N_p} = \frac{\text{TP}}{\text{TP + FN}}, \tag{5.2}$$

$$\text{tasa de falsos positivos, FPR} = \frac{\text{FP}}{N_n} = \frac{\text{FP}}{\text{FP + TN}}. \tag{5.3}$$

donde, TP es el número de verdaderos positivos, FP es el número de falsos positivos, TN es el numero de verdaderos negativos, FN es el número de falsos negativos, N_p es el número total de positivos en base de datos de prueba, y N_n es el número total de negativos en la base de datos de prueba. Al variar el parámetro de umbral de confianza del algoritmo de detección, se obtiene una secuencia de tasas de verdaderos y falsos positivos que se pueden representar en un sistema de coordenadas bidimensional. Típicamente, la tasa de verdaderos positivos se traza a lo largo del eje Y, mientras que el eje X muestra la tasa de falsos positivos. Mientras los denominadores de las ecuaciones (5.2) y (5.3) se mantengan fijas, la curva resultante será siempre monótona creciente en ambos ejes. Por lo tanto, cuanto más cerca la curva esté de la esquina superior izquierda (TPR = 100 % y FPR = 0 %), mejor es el algoritmo de detección. En algunas publicaciones la curva ROC se resume con la medida del área bajo curva ROC (*Area Under ROC Curve*: AUC) (Fawcett, 2003).

5.1.2. Precision-Recall

La recuperación (Recall) mide el número de ejemplos positivos detectados con éxito por el algoritmo. Por otro lado, la precisión de la detección (Precision), mide el porcentaje de hipótesis, que son correctas. Estos valores se calculan con las siguientes ecuaciones:

$$\text{Recall} = \frac{\text{TP}}{N_p}, \tag{5.4}$$

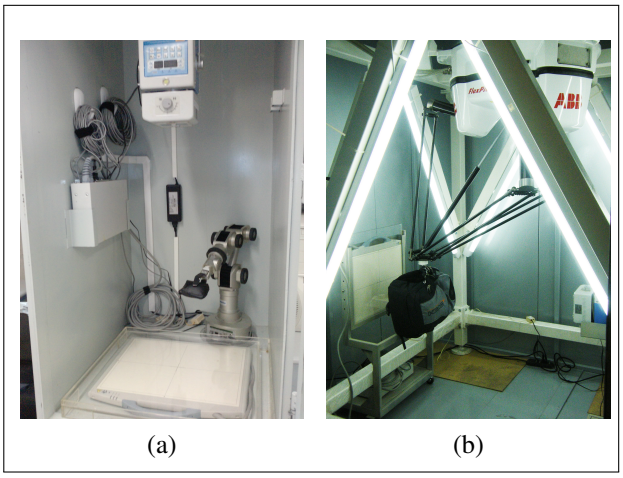
$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}}. \tag{5.5}$$

Un buen algoritmo de detección de objetos, trata de llegar a un Recall de 100 % y una Precision de 100 %. Hay que destacar, que ninguno de los valores en la ecuación (5.4) y

(5.5) dependen de los verdaderos negativos. También se debe considerar, que el denominador de la ecuación (5.5) no es fijo y cambios de acuerdo con el umbral de confianza.

5.2. Equipamiento

Para propósitos experimentales, hemos implementado un sistema de adquisición de imágenes de rayos X, y debido a que en los seres humanos, la exposición prolongada a este tipo de radiación ionizante, puede provocar serios problemas de salud, es que el ambiente de trabajo se circunscribió a una cabina de plomo, que impide la refracción de los rayos X. Así, para la primera propuesta de *framework*, utilizamos una cabina de menor tamaño que la utilizada en la propuesta mejorada, ya que en esta última, se utilizó un manipulador robótico de grandes dimensiones (ver Figuras 5.2a y 5.2b).



(a) (b)

Figura 5.2. Interior de cabinas de plomo, con emisor y detector de rayos X, así como también un manipulador robótico, a) Para propuesta inicial de *framework*, y b) Para propuesta mejorada de *framework*.

En los experimentos hemos usado un detector digital de rayos X (marca Canon, modelo CXDI-50G) que permite adquirir imágenes de 2208 $\times$ 2688 pixeles, tubo emisor de rayos X (marca Poskom, Modelo PXM-20BT), para la propuesta inicial de nuestro *framework* usamos dos sistemas de sujeción de pbjetos; un mecanismo de posicionamiento de objetos semi-automático (giroscopio), un manipulador robótico tipo brazo (marca Neuronics,

modelo Katana 6M), y para la propuesta mejorada del *framework* usamos un manipulador robótico flexible (marca ABB, modelo IRB 340 Flexpicker), equipamiento que se muestra en la Figura 5.3.

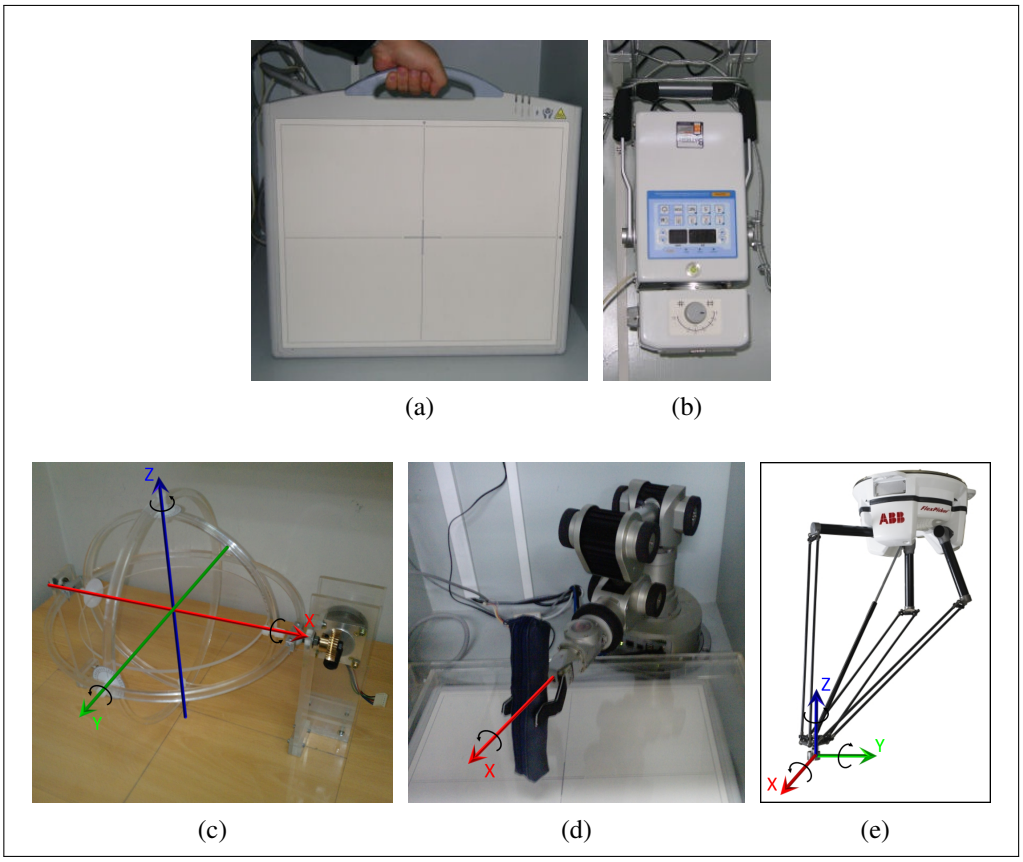


Figura 5.3. Equipamiento necesario para la inspección con rayos X: a) detector de rayos X, b) tubo emisor de rayos X, c) giroscopio semi-automático, d) brazo robótico, e) flexpicker.

Adicionalmente todos los algoritmos fueron implementados en MATLAB R2013b, 64-bit (win64), usamos la librería de código abierto VLFeat (Vedaldi y Fulkerson, 2010) para *K-means* y SIFT. El código fue ejecutado sobre un laptop Intel(R) Core(TM) i7-3537U CPU @ 2.00GHz con 4 nucleos y memoria RAM de 8 GB.

Para la propuesta inicial de *framework*, hemos inspeccionado 9 objetos $Obj_1, ..., Obj_9$, (ver Figura 5.4), cada uno de ellos con una hoja de afeitar; usando el sistema semi-automático

hemos inspeccionado siete de los nueve objetos y usando el brazo robótico hemos inspeccionado seis de los nueve objetos.

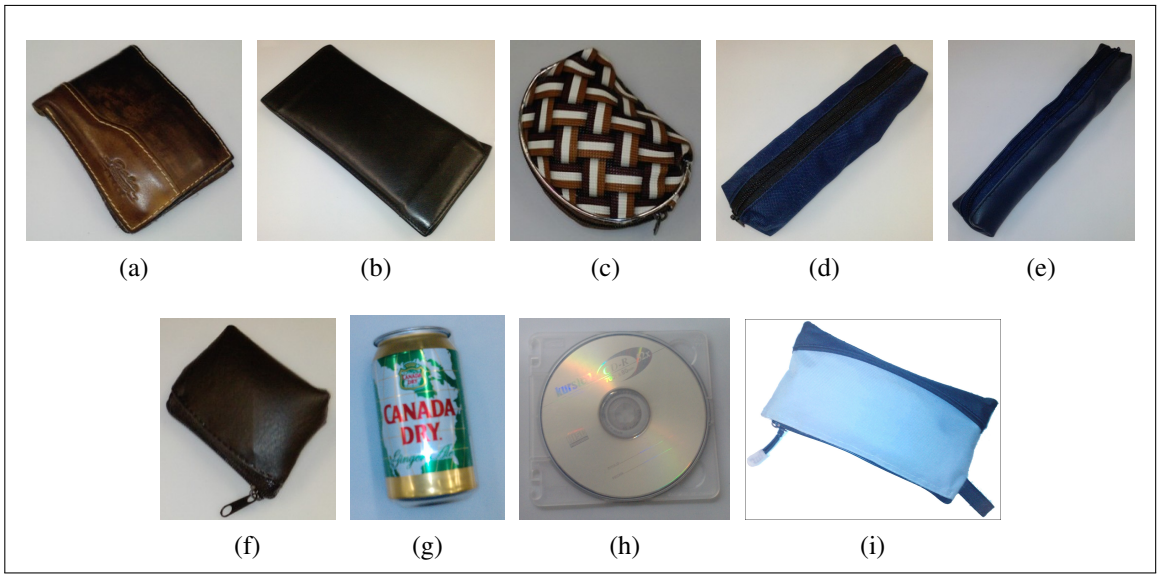


Figura 5.4. Objetos sometidos a inspección radioscópica con propuesta inicial de *framework*: a) Obj_1, b) Obj_2, c) Obj_3, d) Obj_4, e) Obj_5, f) Obj_6, g) Obj_7, h) Obj_8, e i) Obj_9.

La propuesta mejorada de *framework* incluye un detector (AISM) el que hemos evaluado en la detección de tres diferentes objetos amenazantes, que podrían estar presentes en el interior del equipaje de una persona: hojas de afeitar, *shurikens* y revólveres (ver Figura 5.5).

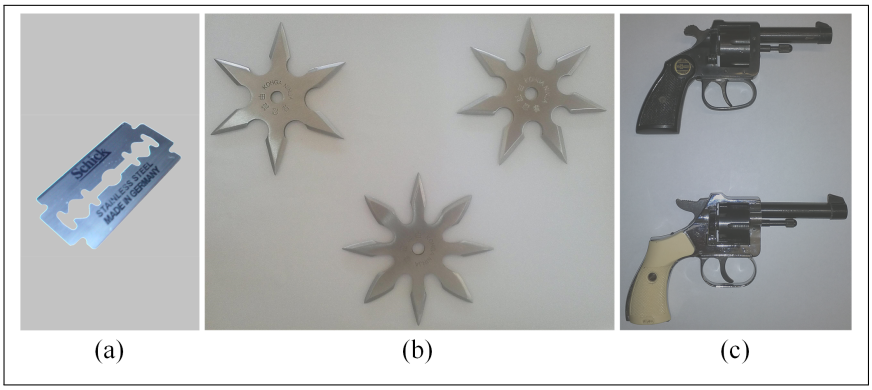


Figura 5.5. Objetos amenazantes en el interior de bolsos; (a) Hoja de afeitar, (b) *Shuriken* con 6, 7 y 8 puntas, y (c) Revolveres.

5.3. Experimentos y Evaluación de Propuesta Inicial de *Framework*

A continuación describiremos los aspectos más relevantes de la experimentación y los resultados obtenidos en nuestra propuesta inicial de *framework* de inspección activa con rayos X (ver sección 4.1).

5.3.1. Caracterización de Objeto de Interés

Como ya mencionamos en la sección 4.1, el objeto de interés evaluado en esta propuesta inicial de *framework* fue una "hoja de afeitar", la cual debió ser caracterizada para las diferentes poses que se han estimado podrían ocurrir en un proceso de inspección. Para esto se ha girado la hoja de afeitar en α y β, como se ve en la Figura 4.3a y 4.3b, situándola en el centro de una esfera de poliestireno expandido EPS (*Expanded PolyStyrene*), tal como se muestra en la Figura 5.6a. Se ha utilizado una esfera de EPS, por la facilidad de movimiento y principalmente por su mínimo coeficiente de absorción de rayos X. En la Figura 5.6b se puede ver la caracterización de una hoja de afeitar mediante los descriptores SIFT, los cuales son mostrados; indicando la magnitud y orientación del descriptor o simplemente como un punto (*keypoint*) dispuesto en una coordenada (x, y).

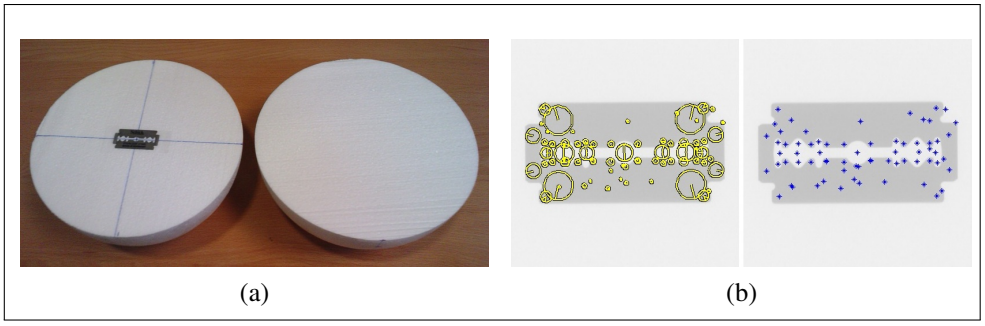


Figura 5.6. Caracterización de una hoja de afeitar; (a) Esfera de EPS para la caracterización, y (b) Hoja de afeitar con dos formas de representación de los descriptores SIFT: magnitud-orientación y *keypoints*.

5.3.2. Secuencias de Inspección

Cada objeto fue inspeccionado diez veces, y la primera imagen de cada secuencia fue tomada en una posición aleatoria. Un resumen de los resultados se encuentra en la Tabla 5.2

para el sistema semi-automático y en la Tabla 5.3 para el brazo robótico. A continuación se muestran en las Figuras 5.7 y 5.8 ejemplos de secuencias de inspección utilizando el sistema semi-automático, y en las Figuras 5.9 y 5.10 secuencias de inspección utilizando un brazo robótico:

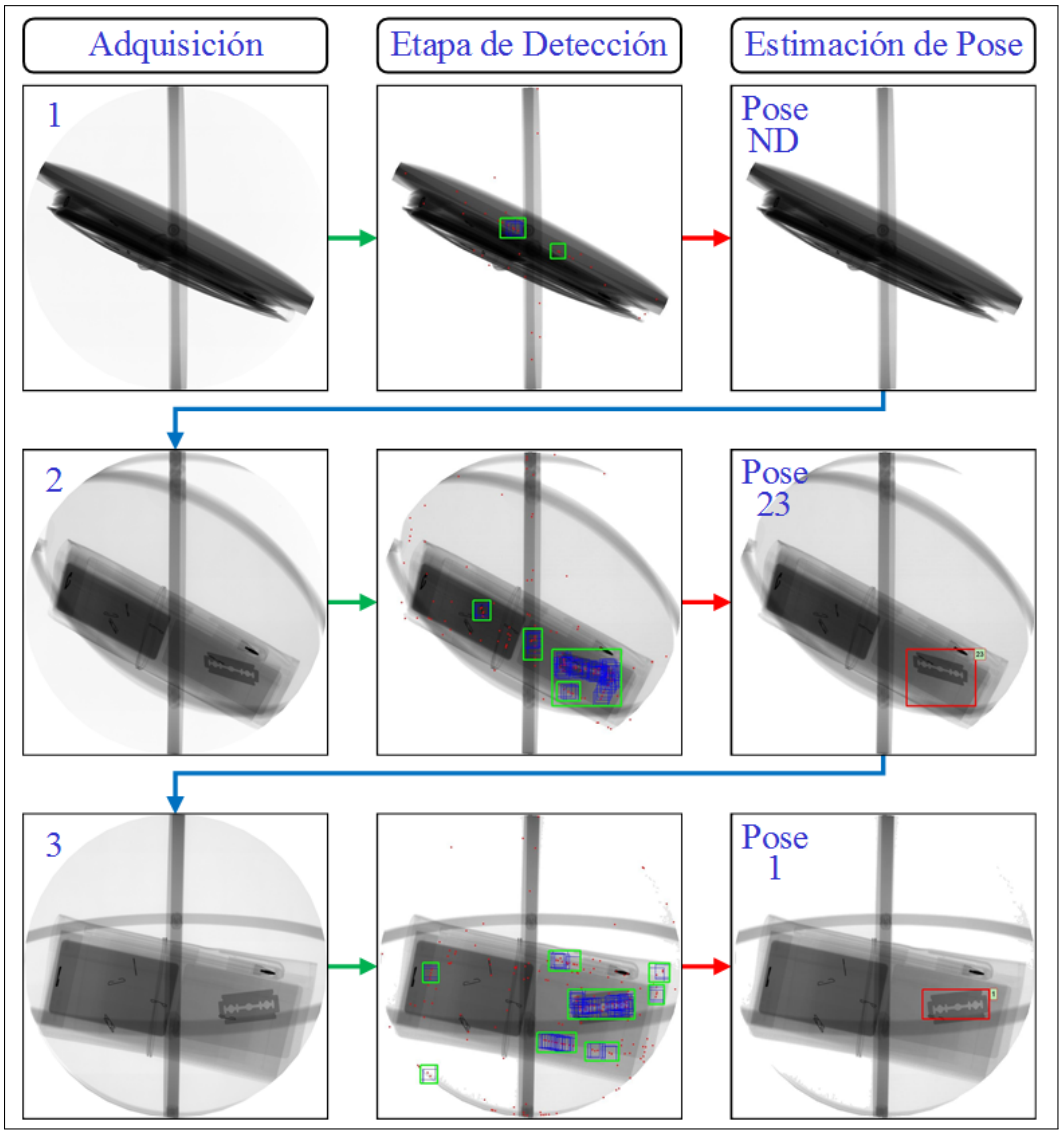


Figura 5.7. Inspección de objeto b (Obj_2), secuencia 10 con sistema semi-automático (ver Tabla 5.2).

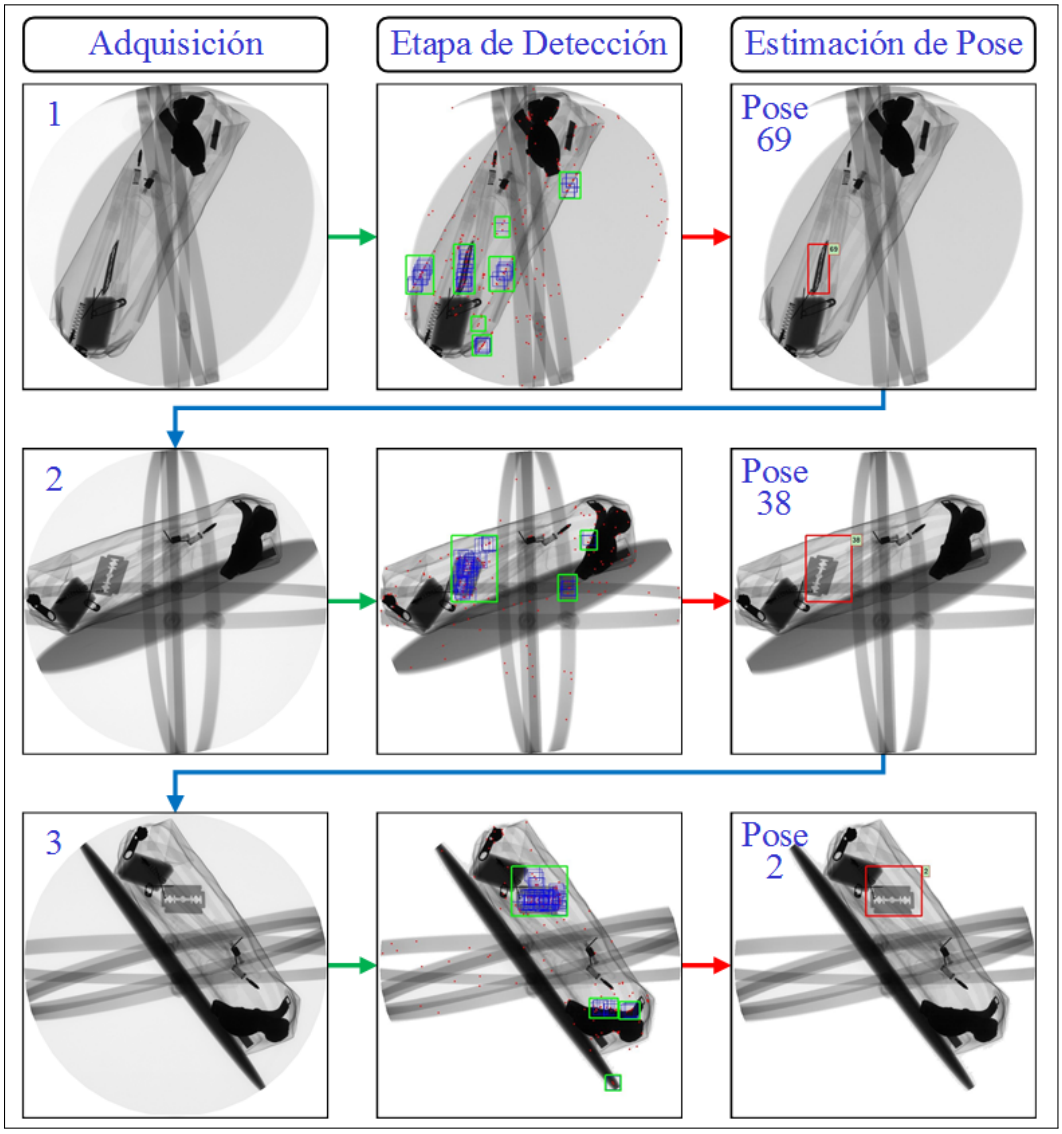


Figura 5.8. Inspección de objeto d (Obj_4), secuencia 5 con sistema semi-automático (ver Tabla 5.2).

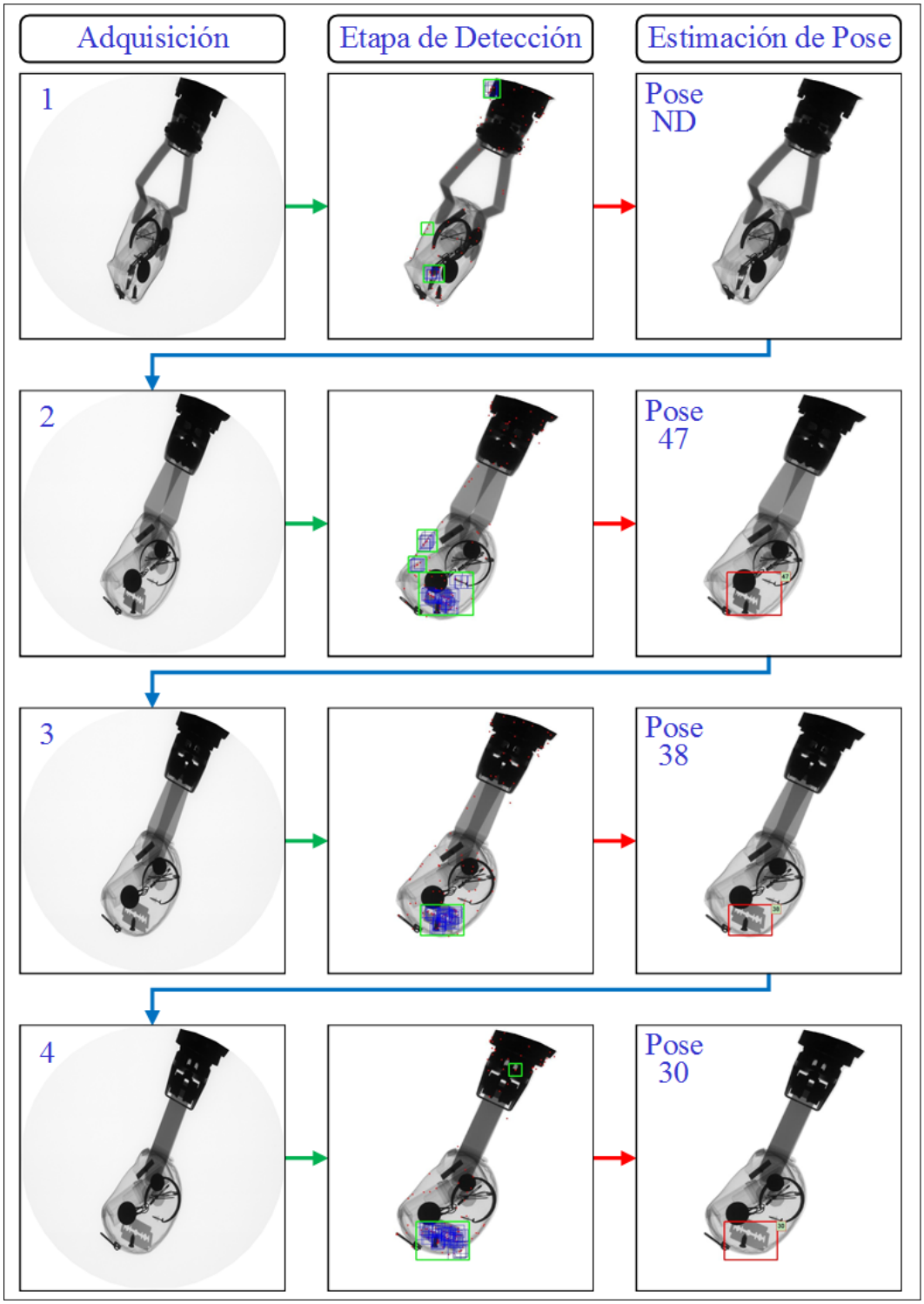


Figura 5.9. Inspección de objeto c (Obj_3), secuencia 7 con brazo robótico (ver Tabla 5.3).

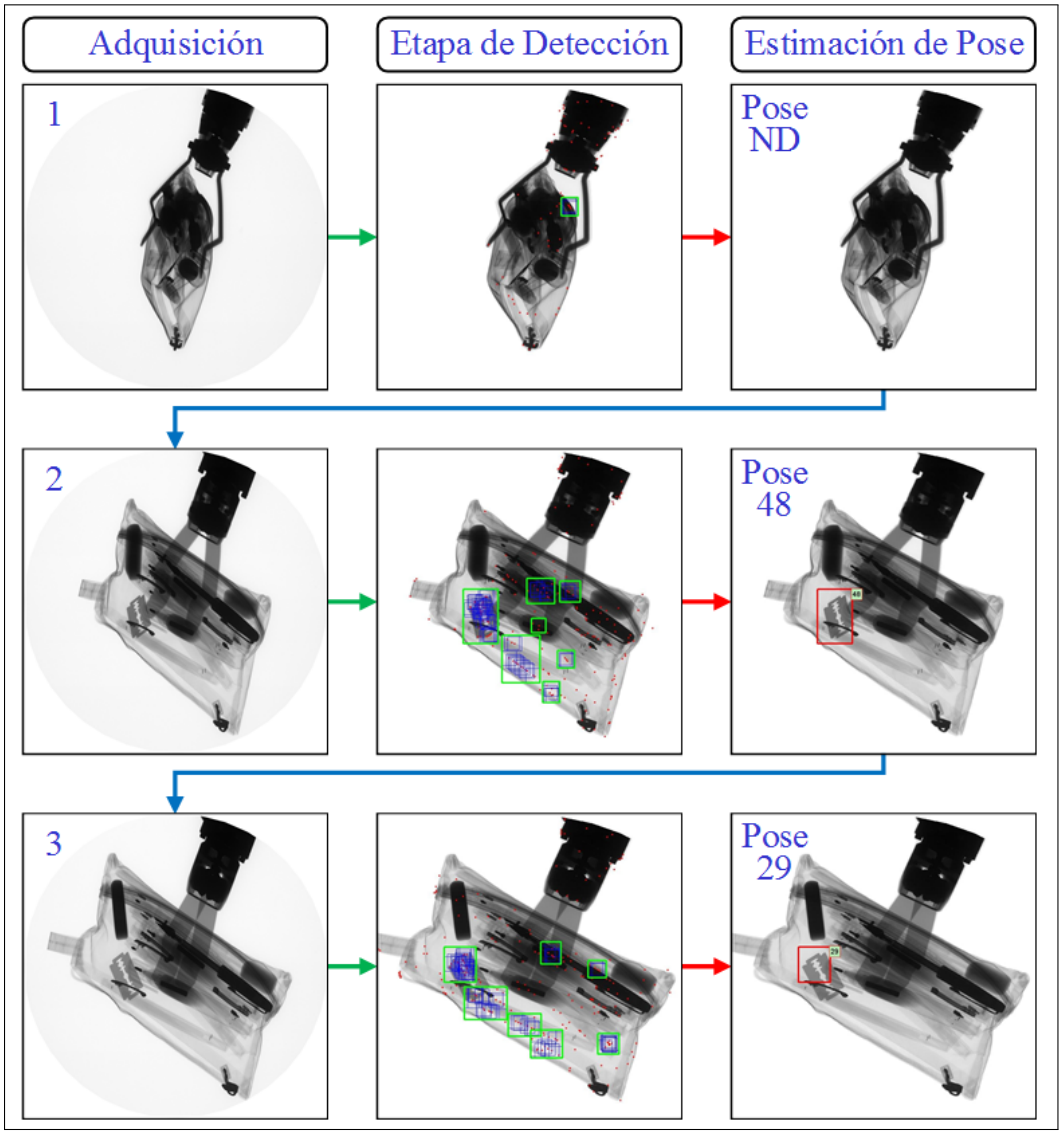


Figura 5.10. Inspección de objeto i (Obj_9), secuencia 2 con brazo robótico (ver Tabla 5.3).

Las Figuras anteriores 5.7, 5.8, 5.9 y 5.10, muestran 4 objetos distintos que han sido inspeccionados; en ellos se aprecian las capacidades de nuestra propuesta para encontrar el objeto de interés en una buena pose, también, podemos ver que en algunos casos existen oclusiones parciales y que de igual manera es posible avanzar en la detección. A partir de los resultados que están representados en forma numérica en la Tabla 5.2 y de las observaciones de estas secuencias, podemos decir que el sistema de sujeción semi-automático (giroscopio) tiene un alto grado de absorción a los rayos X, oscureciendo fuertemente las

imágenes, lo que en ocasiones causó algunas zonas de falsas alarmas, FP, asociadas comúnmente a la pose 2 y debido a la aparición de descriptores SIFT en dichas zonas. Los resultados mostrados en la tabla 5.3, indican que el brazo robótico es una buena alternativa para realizar el movimiento del objeto, sin embargo, y aunque se obtienen las detecciones, estas no siempre están asociadas a una buena pose. Esta situación ocurrió en más oportunidades que con el sistema semi-automático, debido principalmente a que sólo fue posible usar un grado de libertad del brazo robótico, y así no causar auto-oclusiones.

Como se aprecia en las Tablas 5.2 y 5.3, se lograron como máximo secuencias de hasta cuatro imágenes de rayos X, sin embargo, en algunas ocasiones fue suficiente la primera imagen para la detección. En algunos casos se pasó de la no-detección (ND) –en una vista–, a la detección –en la siguiente–, lo cual es muy meritorio, ya que demuestra que el método es capaz de soslayar vistas iniciales inadecuadas y oclusiones. Sin embargo, debido a fuertes oclusiones y posiciones iniciales desfavorables, algunas veces la detección falló.

A partir de los datos registrados en las Tablas 5.2 y 5.3 hemos estimado el desempeño de nuestra propuesta, calculando los índices Recall y Precision, definidos en las ecuaciones (5.4) y (5.5) respectivamente. Así, usando el sistema semi-automático se obtuvo un índices de Recall = 88.6 % y una Precision = 92.5 %, y usando el brazo robótico hemos logrando en este caso índices de Recall = 88.3 % y Precision = 91.4 %. Valores alentadores que nos indican cuánto realmente es capaz de detectar nuestro sistema [1].

[1]Debido a que los experimentos son distintos, los valores de desempeño aquí señalados, no son comparables con la eficiencia de los procesos humanos de inspección visual con rayos X en los aeropuertos, que no supera el 90 % en el mejor de los casos (sólo inspectores humanos jóvenes, entrenados), y que han sido reportados en la literatura (Schwaninger et al., 2005; Hardmeier et al., 2006; Wales et al., 2009).

Tabla 5.2. Secuencias de inspección, usando sistema semi-automático.

Objeto	*Secuencia*[*]	*Pose*	*Rotación*		*Pose*	*Rotación*		*Pose*	*Rotación*		*Pose*	*Desempeño*	
			X	*Z*		*X*	*Z*		*X*	*Z*		TP	FP
Obj_1	1	2	–	–	–	–	–	–	–	–	–	1	0
	2	2	–	–	–	–	–	–	–	–	–	1	0
	3	ND	-40°	–	2	–	–	–	–	–	–	1	0
	4	ND	-40°	–	28	–	–	–	–	–	–	1	0
	5	2	–	–	–	–	–	–	–	–	–	1	0
	6	ND	-40°	–	ND	-40°	–	28	–	–	–	1	0
	7	2	–	–	–	–	–	–	–	–	–	1	0
	8	ND	-40°	–	2	–	–	–	–	–	–	1	0
	9	2FP	–	–	–	–	–	–	–	–	–	0	1
	10	ND	-40°	–	ND	-40°	–	2	–	–	–	1	0
Obj_2	1	1	–	–	–	–	–	–	–	–	–	1	0
	2	3	–	–	–	–	–	–	–	–	–	1	0
	3	ND	-40°	–	10	–	–	–	–	–	–	1	0
	4	65	-10°	-35°	56	-10°	-17°	55	–	–	–	1	0
	5	59	-40°	-31°	2FP	–	–	–	–	–	–	0	1
	6	60	-50°	-69°	2	–	–	–	–	–	–	1	0
	7	16	-60°	0°	10	–	–	–	–	–	–	1	0
	8	7	-60°	20°	19	–	–	–	–	–	–	1	0
	9	8	-70°	23°	10	–	–	–	–	–	–	1	0
	10	ND	-40°	–	23	-40°	24°	1	–	–	–	1	0
Obj_3	1	19	–	–	–	–	–	–	–	–	–	1	0
	2	ND	-40°	–	28	–	–	–	–	–	–	1	0
	3	ND	-40°	–	ND	-40°	–	64	–	–	–	1	0
	4	ND	-40°	–	ND	-40°	–	ND	–	–	–	0	0
	5	ND	-40°	–	ND	-40°	–	3	–	–	–	1	0
	6	ND	-40°	–	2FP	–	–	–	–	–	–	0	1
	7	10	–	–	–	–	–	–	–	–	–	1	0
	8	ND	-40°	–	ND	-40°	–	37	–	–	–	1	0
	9	ND	-40°	–	ND	-40°	–	19	–	–	–	1	0
	10	ND	-40°	–	ND	-40°	–	5	40°	-85°	21	1	0
Obj_4	1	50	-40°	9°	28	–	–	–	–	–	–	1	0
	2	39	-20°	-77°	51	-50°	43°	ND	-40°	–	3	1	0
	3	2FP	–	–	–	–	–	–	–	–	–	0	1
	4	28	–	–	–	–	–	–	–	–	–	1	0
	5	69	-50°	-83°	38	-10°	-77°	2	–	–	–	1	0
	6	4	–	–	–	–	–	–	–	–	–	1	0
	7	6	-50°	27°	ND	-40°	–	32	-40°	15°	20	1	0
	8	30	–	–	–	–	–	–	–	–	–	1	0
	9	48	-20°	-58°	37	–	–	–	–	–	–	1	0
	10	50	-40°	-44°	2	–	–	–	–	–	–	1	0
Obj_5	1	31	–	–	–	–	–	–	–	–	–	1	0
	2	ND	-40°	–	ND	-40°	–	ND	–	–	–	0	0
	3	77	-40°	-41°	30	–	–	–	–	–	–	1	0
	4	ND	-40°	–	ND	-40°	–	46	–	–	–	1	0
	5	13	–	–	–	–	–	–	–	–	–	1	0
	6	ND	-40°	–	2	–	–	–	–	–	–	1	0
	7	ND	-40°	–	46	–	–	–	–	–	–	1	0
	8	ND	-40°	–	ND	-40°	–	2	–	–	–	1	0
	6	ND	-40°	–	2	–	–	–	–	–	–	1	0
	10	59	-40°	83°	ND	-40°	–	2	–	–	–	1	0
Obj_6	1	ND	-40°	–	2	–	–	–	–	–	–	1	0
	2	7	-60°	-4°	7	-60°	-8°	ND	-40°	–	2	1	0
	3	ND	-40°	–	ND	-40°	–	ND	–	–	–	0	0
	4	2	–	–	–	–	–	–	–	–	–	1	0
	5	ND	-40°	–	2	–	–	–	–	–	–	1	0
	6	ND	-40°	–	ND	-40°	–	2	–	–	–	1	0
	7	ND	-40°	–	37	–	–	–	–	–	–	1	0
	8	16	-60°	-1°	2	–	–	–	–	–	–	1	0
	9	2FP	–	–	–	–	–	–	–	–	–	0	1
	10	ND	-40°	–	2	–	–	–	–	–	–	1	0
Obj_7	1	2	–	–	–	–	–	–	–	–	–	1	0
	2	41	-40°	17°	19	–	–	–	–	–	–	1	0
	3	ND	-40°	–	66	-20°	68°	20	–	–	–	1	0
	4	42	-50°	3°	19	–	–	–	–	–	–	1	0
	5	5	-40°	-64°	19	–	–	–	–	–	–	1	0
	6	ND	-40°	–	64	–	–	–	–	–	–	1	0
	7	41	-40°	4°	20	–	–	–	–	–	–	1	0
	8	22	–	–	–	–	–	–	–	–	–	1	0
	9	24	-50°	10°	19	–	–	–	–	–	–	1	0
	10	48	-20°	-4°	38	-10°	-19°	46	–	–	–	1	0
											TOTAL:	62	5
											Recall = TP/N_p:	88.6 %	
											Precision = TP/(TP + FP):	92.5 %	

ND: *Ninguna Detección*, TP: *Verdadero Positivo*, FP: *Falso Positivo*, N_p: *número total de positivos*.

[*]La primera imagen adquirida para cada una de estas secuencias se pueden ver en el anexo C

Tabla 5.3. Secuencias de inspección, usando brazo robótico.

Objeto	*Secuencia**	*Pose*	*Rotación X*	*Pose*	*Rotación X*	*Pose*	*Rotación X*	*Pose*	*Desempeño* TP	FP
	1	2	–	–	–	–	–	–	1	0
	2	ND	-40°	ND	-40°	4	–	–	1	0
	3	ND	-40°	5	-40°	2	–	–	1	0
	4	ND	-40°	2	–	–	–	–	1	0
Obj_1	5	ND	-40°	2	–	–	–	–	1	0
	6	11	–	–	–	–	–	–	1	0
	7	ND	-40°	14	-40°	2	–	–	1	0
	8	ND	-40°	19	–	–	–	–	1	0
	9	ND	-40°	ND	-40°	1	–	–	1	0
	10	24	-50°	2FP	–	–	–	–	0	1
	1	ND	-40°	4	–	–	–	–	1	0
	2	5	-40°	1	–	–	–	–	1	0
	3	ND	-40°	ND	-40°	4	–	–	1	0
	4	ND	-40°	4	–	–	–	–	1	0
Obj_3	5	ND	-40°	10	–	–	–	–	1	0
	6	46	–	–	–	–	–	–	1	0
	7	ND	-40°	47	-10°	38	-10°	30	1	0
	8	55	–	–	–	–	–	–	1	0
	9	52FP	–	–	–	–	–	–	0	1
	10	48	-20°	31	–	–	–	–	1	0
	1	28	–	–	–	–	–	–	1	0
	2	ND	-40°	ND	-40°	29	–	–	1	0
	3	3	–	–	–	–	–	–	1	0
	4	ND	-40°	37	–	–	–	–	1	0
Obj_4	5	ND	-40°	64	–	–	–	–	1	0
	6	55	–	–	–	–	–	–	1	0
	7	ND	-40°	67	-30°	61	-60°	ND	0	0
	8	37	–	–	–	–	–	–	1	0
	9	68	-40°	ND	-40°	2FP	–	–	0	1
	10	ND	-40°	55	–	–	–	–	1	0
	1	28	–	–	–	–	–	–	1	0
	2	20	–	–	–	–	–	–	1	0
	3	55	–	–	–	–	–	–	1	0
	4	46	–	–	–	–	–	–	1	0
Obj_7	5	65	-10°	56	-10°	46	–	–	1	0
	6	65	-10°	57	-20°	39	-20°	20	1	0
	7	ND	-40°	46	–	–	–	–	1	0
	8	ND	-40°	ND	-40°	10	–	–	1	0
	9	5	-40°	1	–	–	–	–	1	0
	10	ND	-40°	52	-60°	3	–	–	1	0
	1	ND	-40°	6	-50°	1	–	–	1	0
	2	48	-20°	29	–	–	–	–	1	0
	3	5	-40°	2	–	–	–	–	1	0
	4	ND	-40°	3	–	–	–	–	1	0
Obj_8	5	55	–	–	–	–	–	–	1	0
	6	ND	-40°	19	–	–	–	–	1	0
	7	2FP	–	–	–	–	–	–	0	1
	8	ND	-40°	6	-50°	11	–	–	1	0
	9	59	-40°	21	–	–	–	–	1	0
	10	6	-50°	10	–	–	–	–	1	0
	1	19	–	–	–	–	–	–	1	0
	2	ND	-40°	48	-20°	29	–	–	1	0
	3	ND	-40°	ND	-40°	37	–	–	1	0
	4	55	–	–	–	–	–	–	1	0
Obj_9	5	ND	-40°	46	–	–	–	–	1	0
	6	ND	-40°	2FP	–	–	–	–	0	1
	7	ND	-40°	ND	-40°	ND	–	–	0	0
	8	ND	-40°	ND	-40°	2	–	–	1	0
	9	55	–	–	–	–	–	–	1	0
	10	41	-40°	56	-10°	55	–	–	1	0
								TOTAL:	53	5
								Recall = TP/N_p:	88.3 %	
								Precision = TP/(TP + FP):	91.4 %	

ND: *Ninguna Detección*, TP: *Verdadero Positivo*, FP: *Falso Positivo*, N_p: *número total de positivos*.

*La primera imagen adquirida para cada una de estas secuencias se pueden ver en el anexo D

5.4. Evaluación de Propuesta de un Detector de Objetos Amenazantes, AISM

En esta sección mostraremos los experimentos realizados para el diseño y evaluación de nuestra propuesta, la cual permite detectar automáticamente objetos de interés (hoja de afeitar, *shuriken* y revolver) en las imágenes de rayos X. Método que hemos denominado, modelo de forma implícita adaptado, AISM.

5.4.1. Imágenes Para Entrenamiento y Pruebas

Como se explicó en la sección 4.2, las imágenes de rayos X de cada objeto de interés son adquiridas en poses representativas (ver Figura 4.6 para un ejemplo de hoja de afeitar). El número de imágenes de entrenamiento utilizado en cada experimento es: 100 para hojas de afeitar, 100 para *shuriken* y 200 para revólveres. Vale la pena señalar que el número de imágenes de entrenamiento utilizados en la detección de armas de fuego es mayor debido a las grandes variaciones intra-clase.

Hemos usado tres conjuntos de imágenes de prueba (una para cada objeto de interés). Cada conjunto de pruebas consta de 200 imágenes de rayos X, con los siguientes subgrupos: a) 150 imágenes de rayos X con sólo un objeto de interés cada una, b) 30 imágenes de rayos X con dos objetos de interés cada una, y las imágenes c) 20 imágenes de rayos X con ningún objeto de interés. En otras palabras, cada conjunto de pruebas contiene $N_p = 150+2\times30 = 210$ objetos de interés, que corresponden a la clase positiva a ser detectada. Las imágenes fueron adquiridas de diferentes bolsos que contienen varios objetos en muchas poses (ver ejemplos en las Figuras 5.14 y 5.15). En promedio, cada imagen de rayos X contiene 18 objetos que no son objetos de interés (bolígrafos, CDs, clips, monedas, tornillos, pinzas, agujas, etc.). Así, cada conjunto de pruebas contiene $N_n = 18 \times 200 = 3600$ objetos que pertenecen a la clase negativa. Adicionalmente, todos los objetos de interés se etiquetaron manualmente con cuadros delimitadores, que corresponden al *ground truth* de nuestros experimentos, denotados por BB_{gt}, los cuales describen las posiciones del objeto de interés en las imágenes de rayos X.

5.4.2. Metodología de Evaluación

El desempeño de nuestro método se mide utilizando los criterios de evaluación de la calidad del 'PASCAL Visual Object Classes Challenge' (Everingham et al., 2010), donde una detección se considera correcta si el área normalizada de solapamiento a_o, entre el cuadro delimitador de una detección hipotética BB_{dt} y el cuadro de etiqueta BB_{gt} excede un valor de umbral θ_a, como se define en la ecuación (5.1) y se ilustra en la Figura 5.1.

Por lo general, una detección se considera correcta si $a_0 \geq \theta_a$ con $\theta_a = 0.5$. Sin embargo, se incluyeron dos casos adicionales: $\theta_a = 0.45$ y $\theta_a = 0.4$. Lo que se tuvo en cuenta debido a que la orientación de BB_{gt} y BB_{dt} puede ser muy diferente en ciertas detecciones, produciendo un área de intersección que es considerablemente menor que el área de intersección, que se podría obtener para orientaciones similares (ver Figura 5.11).

Figura 5.11. Detección de objetos de interés (hoja de afeitar, *shuriken* y revolver) con $\theta_a = 0.4$.

Con el fin de evaluar el desempeño, hemos ejecutado nuestro algoritmo en cada prueba. Las detecciones obtenidas fueron evaluadas de la siguiente manera: Se midió el número total de verdaderos positivos (TP) en todo el conjunto de prueba, es decir, el número de detecciones que cumplieron con $a_0 \geq \theta_a$; y el número total de falsos positivos (FP), es decir, el número de detecciones en las cuales $a_0 < \theta_a$. Idealmente, TP = N_p (el número total de objetos de interés) y FP $= 0$. Como se mencionó en la metodología (sección 4.2), nuestro enfoque incluye un valor de umbral (θ_s), que puede ser sintonizado para lograr el mejor *trade-off* entre correctas y falsas detecciones. Para graficar la curva ROC, se calculan la tasa de verdaderos positivos, TPR, y la tasa de falsos positivos, FPR, utilizando las ecuaciones (5.2) y (5.3), para diferentes valores de θ_s. Idealmente, TPR = 1 y FPR = 0. Para cada

caso se calcula el área bajo la curva ROC (AUC) con el fin de medir el desempeño. AUC idealmente debe ser 1, y entregamos $TPR_{0.05}$, que es el valor de TPR en FPR = 0.05. Además, calculamos el mejor punto de operación (FPR^*, TPR^*), es decir, el punto de la curva ROC cuya distancia al punto de operación ideales (FPR = 0, TPR = 1) es mínima.

5.4.3. Parámetros de Sintonización

El método propuesto tiene seis parámetros que deben ser ajustados para cada categoría de objeto: Umbrales θ_u, θ_B, θ_m, el tamaño del cuadro W_B y el alto y ancho de la ventana de detección W_F. Para fines de sintonización hemos creado un 'conjunto de datos de ajuste' para cada categoría de objeto, es decir, 10 imágenes de rayos X seleccionadas al azar, tomadas de las 200 imágenes de rayos X del correspondiente conjunto de datos de prueba. Los parámetros fueron sintonizados manualmente mediante ensayo y error, maximizando el desempeño en el conjunto de datos de sintonización. Estos valores se muestran en la Tabla 5.4 para cada categoría de objeto. De este modo, obtuvimos el desempeño reportado, usando los parámetros ajustados, en el conjunto de datos de prueba.

Tabla 5.4. Parámetros sintonizados para realizar las pruebas de inspección

Objeto de Interés:	Hoja de Afeitar	Shuriken	Revolver
θ_u	50000	30000	30000
θ_B	5	3	1
θ_m	1	11	4
W_B [pixels]	100×100	150×150	70×70
W_F [pixels]	200×360	820×820	$800 \times 1{,}300$

5.4.4. Resultados

Los verdaderos positivos, falsos positivos y las curvas ROC para la detección de hojas de afeitar, *shurikens* y revólveres se muestran en la Figura 5.12. La Tabla 5.5 resume el desempeño alcanzado en cada caso. Las Figuras 5.14 y 5.15 muestran algunos ejemplos de detección de hojas de afeitar, *shuriken* y revólveres. La etiqueta (*ground truth*), es decir, el objeto de interés que será detectado, se muestra en verde. Las detecciones se muestran en rojo (para verdaderos positivos) y azul (para falsos positivos).

Tabla 5.5. Resumen del desempeño de AISM.

		con $a_o \geq$		
Objeto de Interés	Variable	0.5	0.45	0.4
Hoja de Afeitar	AUC	0.9915	0.9942	0.9954
	$TPR_{0.05}$	0.9972	0.9998	1.0000
	TPR^*	0.9836	0.9870	0.9876
	FPR^*	0.0350	0.0250	0.0200
Shuriken	AUC	0.9821	0.9832	0.9847
	$TPR_{0.05}$	0.9388	0.9487	0.9621
	TPR^*	0.9650	0.9717	0.9727
	FPR^*	0.0600	0.0600	0.0550
Revolver	AUC	0.8715	0.9029	0.9225
	$TPR_{0.05}$	0.3219	0.4023	0.4754
	TPR^*	0.8261	0.8657	0.8884
	FPR^*	0.2250	0.1950	0.1700

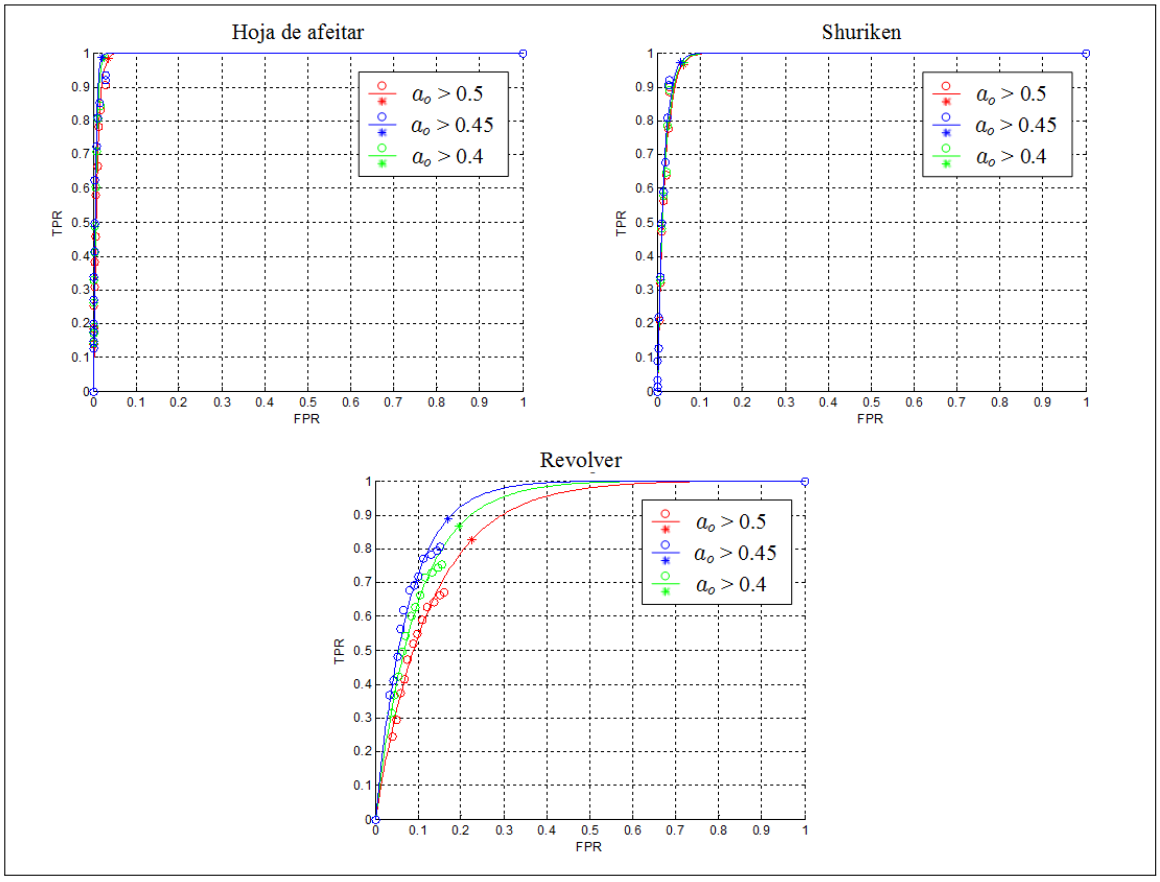


Figura 5.12. Curva ROC para la detección de: hojas de afeitar, *Shurikens* y revólveres para $\theta_a = 0.5, 0.45, 0.4$. En todos los casos, el número de muestras positivas y negativas es $N_p = 210$ y $N_n = 3600$, respectivamente. Los puntos medidos se representan como 'o', los cuales se ajustaron a una curva $y = a(1 - exp(-\gamma x))$. El mejor punto de operación (FPR*, TPR*) se muestra como '*'.

Sobre la base de esta evaluación, la detección de las hojas de afeitar y *shurikens* es claramente muy efectiva. En ambos casos, se obtuvo un alto TPR a muy bajo FPR. Los resultados para la detección de revólveres son algo más bajas (no fue posible obtener una alta TPR en un muy baja FPR). Dado que los objetos asimétricos tienen ocurrencias muy inconexas con respecto al centro real del objeto, en nuestro enfoque se obtienen los mejores resultados para los objetos simétricos como hojas de afeitar y *shurikens*.

5.4.5. Comparación con Otros Métodos

En esta sección, se presentan los resultados que se obtuvieron mediante la comparación de nuestro método con tres métodos conocidos, que se pueden utilizar en la detección de objetos. Los métodos de referencia utilizados aquí son los siguientes:

- SIFT (Lowe, 2004): Usamos el detector en una vista, que aplicamos en nuestra propuesta inicial de *framework* de inspección activa con rayos X (Riffo y Mery, 2012).
- SURF (Bay et al., 2008): Usamos el mismo algoritmo aquí propuesto (AISM), pero con descriptores SURF, en lugar de descriptores SIFT.
- ISM (Leibe et al., 2008): Usamos el método ISM original, que fue desarrollado para detectar categoría de objetos, tales como, automóviles, personas y animales.

Tabla 5.6. Comparación de AISM con otros métodos.

AISM (nuestro)	AUC	0.9917
	$TPR_{0.05}$	0.9975
	TPR^*	0.9849
	FPR^*	0.0350
SIFT	AUC	0.9211
	$TPR_{0.05}$	0.4693
	TPR^*	0.8840
	FPR^*	0.1700
SURF	AUC	0.6162
	$TPR_{0.05}$	0.1276
	TPR^*	0.6564
	FPR^*	0.3700
ISM	AUC	0.9553
	$TPR_{0.05}$	0.6734
	TPR^*	0.9237
	FPR^*	0.1150

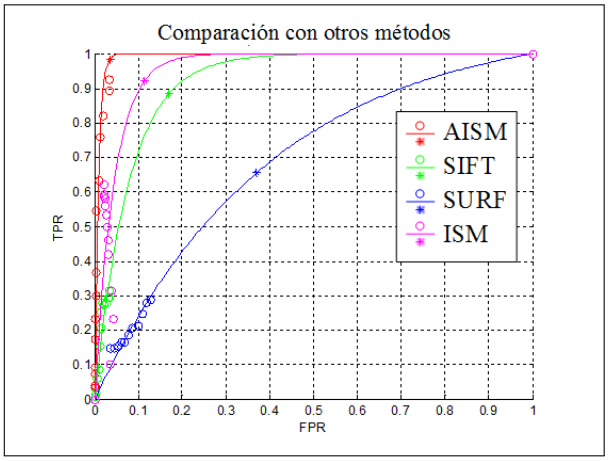


Figura 5.13. Curva ROC de nuestro método AISM en comparación con otros tres métodos conocidos. En todos los casos, el número de muestras positivas y negativas es de $N_p = 150$ y $N_n = 2700$, respectivamente. Los puntos medidos se representan como 'o', los cuales se ajustaron a una curva $y = a(1 - exp(-\gamma x))$. El mejor punto de operación (FPR*, TPR*) se muestra como '*'.

Para comparar los distintos enfoques en condiciones idénticas, utilizamos sólo las primeras 150 imágenes del conjunto de pruebas (subconjunto 'a', como se explica en la sección 5.4.1 con una hoja de afeitar por imagen). Por lo tanto, los métodos fueron sintonizados para detectar sólo un objeto de interés por imagen.

Las distintas curvas ROC que permiten comparan los enfoques se ilustran en la Figura 5.13 y un resumen de los resultados de cada método se muestra en la Tabla 5.6. Para lograr comparar los métodos de referencia y obtener diferentes valores de TPR y FPR, tuvimos que modificar un valor umbral para cada detector: *i*) En nuestro enfoque (AISM), el umbral de activación fue el valor que mantiene en la imagen a los candidatos detectados con las puntuaciones más altas, *ii*) Para nuestra propuesta de algoritmo AISM, utilizando descriptores SURF, el umbral de activación fue el mismo que en nuestra AISM, es decir, el valor que mantiene en la imagen a los candidatos detectados con las puntuaciones más altas, *iii*) En el enfoque SIFT-Lowe, el umbral de activación fue el valor que mantiene sólo los descriptores SIFT útiles $\hat{\mathbf{f}}$ en la imagen de rayos X, y *iv*) En el enfoque ISM original, el umbral de activación fue también el valor que mantiene sólo los descriptores SIFT útiles en la imagen de rayos X.

En la Figura 5.13, comparamos las curvas ROC de los cuatro métodos. Nuestro método tiene el mejor desempeño, logrando TPR = 0.9975 en FPR = 0.05. Se demuestra que la adaptación de ISM con descriptores SIFT aumenta significativamente el rendimiento, ya que los desempeños de reconocimiento obtenidos por nuestro método son mucho más altos que los obtenidos por los otros métodos.

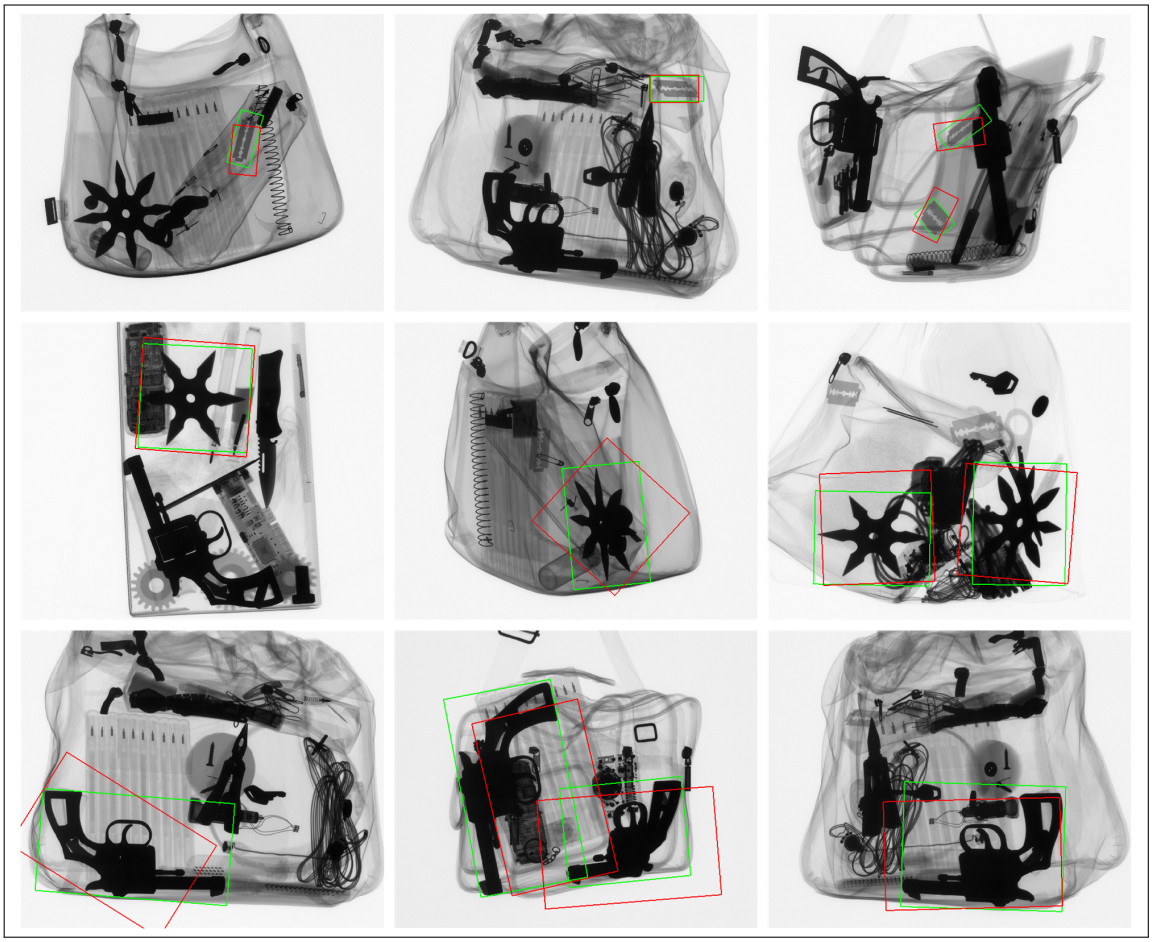


Figura 5.14. Detección de AISM, Verdaderos positivos: Ejemplos de imágenes para las cuales nuestro detector de objetos de interés obtuvo resultados de detección perfectos (restringidas a $a_o \geq 0.5$). Las etiquetas BB_{gt} se muestra en verde y las detecciones BB_{dt} se muestran en rojo. Las imágenes de rayos X son mostradas de la siguiente forma, primera fila: hoja de afeitar, segunda fila: *shuriken* y tercera fila: revolver.

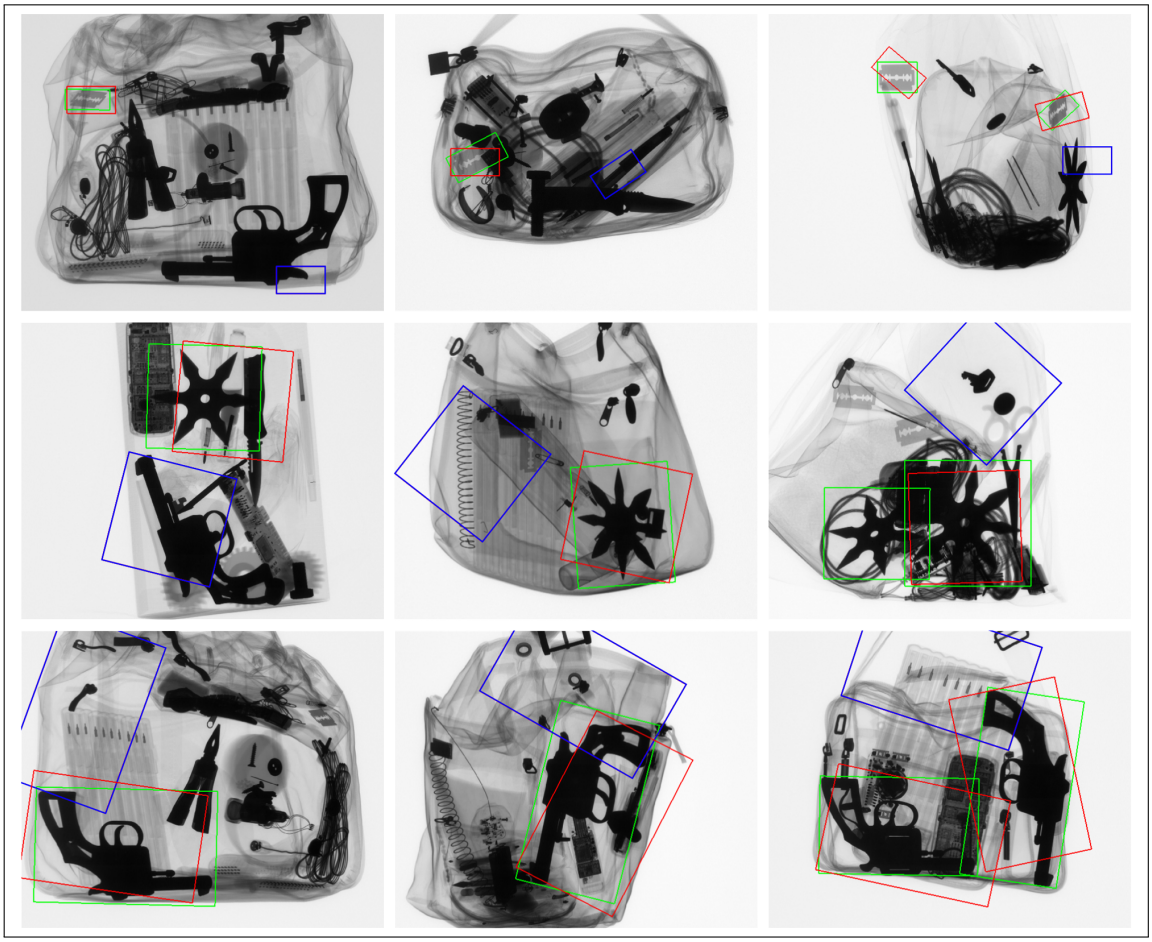


Figura 5.15. Detección de AISM, Verdaderos y falsos positivos: Ejemplos de imágenes para las cuales nuestro detector de objetos de interés obtuvo falsas detecciones. Las etiquetas se muestra en verde. Los verdaderos positivos y falsos positivos se muestran en rojo y azul, respectivamente. Las imágenes de rayos X son mostradas de la siguiente forma, primera fila: hoja de afeitar, segunda fila: *shuriken* y tercera fila: revolver.

5.5. Experimentos y Evaluación de Propuesta de *Framework* Mejorada

Si bien es cierto, en la propuesta inicial de *framework* obtuvimos buenos resultados en la detección de una hoja de afeitar, quisimos ampliar las posibilidades de detección activa con rayos X, introduciendo objetos de interés en ambientes más realistas, es decir, bolsos de mayor tamaño y con mayor variedad de objetos. Adicionalmente, y como describimos en la sección 4.2 decidimos incorporar algoritmos que hicieran más robusta la inspección, tales como, un detector de objetos amenazantes basado en una adaptación de ISM, un estimador

de movimiento del manipulador robótico usando *Q-learning*, y un algoritmo de eliminación de falsas alarmas que usa la restricción epipolar y trifocal.

Para la evaluación del desempeño y atributos de nuestra propuesta mejorada de *framework*, hemos considerado dos objetos amenazantes; revolver y hojas de afeitar. En el caso del revolver y por la particular asimetría en su forma y por la deficiencia del detector AISM de ubicar el centroide del cuadro de detección BB_{dt} lo más cercanamente posible al centro de masa del objeto, es que no se aplicó –para este objeto– la eliminación de falsas alarmas (FP) usando la teoría epipolar y trifocal, ya que la restricción epipolar y trifocal (ver sección 2.5), no es, en la mayoría de los casos coincidente con las detecciones correctas.

5.5.1. Detección Activa de Revolver en Buena Pose

Para la evaluación del desempeño de nuestro *framework* mejorado, aplicado a la detección activa de un revolver, hemos considerado dos secuencias etiquetadas, es decir con anotaciones del *ground truth* BB_{gt}, de dos bolsos, conteniendo un revolver cada uno de ellos y otros objetos que no son de interés. Cada una de estas secuencias consta de 180 imágenes, que han sido adquiridas cada 2 grados de rotación del eje Z del manipulador robótico (Flexpicker).

En este caso hemos considerado la *Validación Automática* del desempeño, donde el seguimiento del objeto de interés se realizó a partir de la primera imagen de rayos X capturada, y hemos aplicado el algoritmo detector AISM, para luego seleccionar automáticamente la detección hipotética (si es que la hubiese) con buena pose. De esta forma se establecen tres situaciones iniciales: *i*) en caso de no existir detección ND, se realiza una rotación del manipulador robótico de $\gamma = 60°$, *ii*) en el caso de tener una detección válida en buena pose, el proceso de inspección se detiene, pues se ha logrado el objetivo de la inspección, y *iii*) en el caso de tener una o varias detecciones hipotéticas en malas poses, se selecciona aleatoriamente la detección y su pose asociada. A partir de esta detección, se realizan de manera automática los movimientos de rotación γ (estimados por el algoritmo de *Q-learning*) y adquisiciones de imágenes de rayos X, para la búsqueda del objeto de interés en una buena pose BP. El proceso de detección y movimientos se realizan automáticamente hasta obtener

un a_o mayor que un cierto valor de umbral θ_a y hasta que la detección sea realizada en una buena pose. La evaluación del desempeño se realizó mediante el análisis del área de overlap a_o, en la última imagen capturada. Cabe señalar que el valor de umbral para el área overlap se determinó empíricamente, y si bien parece bajo, la forma del revólver, la orientación y el tamaño fijo de la ventana de detección, hacen que visualmente dicho valor sea apropiado. Como referencia, la Figura 5.16 muestra una detección con área de overlap $a_o = 0.37$, y claramente vemos que la detección considera casi 2/3 del área visible del revolver.

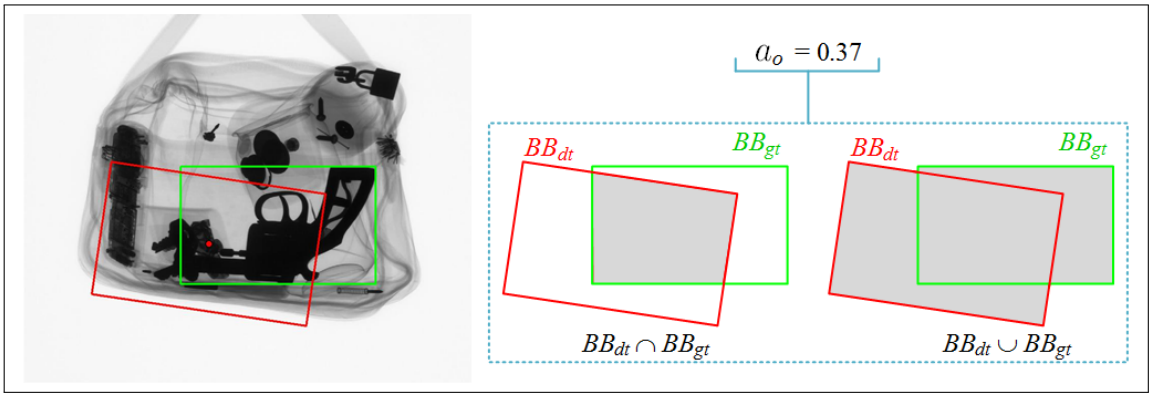


Figura 5.16. Detección de revolver con $a_o = 0.37$.

Para hacer más genérica la evaluación del desempeño de nuestro *framework*, hemos evaluado dos secuencias, considerando que el ángulo inicial γ del manipulador robótico se incrementa cada 2°, en otras palabras, para una mismo bolso hicimos 180 inspecciones y cada inspección usó una imagen inicial distinta. En las tablas 5.7 y 5.8 vemos la evaluación automática considerando distintos valores de umbral: $a_o > 0.3$, $a_o > 0.4$ y $a_o > 0.5$.

Tabla 5.7. Desempeño para la detección activa de un revolver (primera bolso), usando *Q-learning* para estimar el movimiento del manipulador robótico.

Número de inspecciones	Índice de desempeño	Desempeño		
		$a_o > 0.3$	$a_o > 0.4$	$a_o > 0.5$
180	Recall :	63 %	44 %	20 %
	Precision:	80 %	56 %	25 %

Tabla 5.8. Desempeño para la detección activa de un revolver (segundo bolso), usando *Q-learning* para estimar el movimiento del manipulador robótico.

Número de inspecciones	Índice de desempeño	Desempeño		
		$a_o > 0.3$	$a_o > 0.4$	$a_o > 0.5$
180	Recall :	70 %	53 %	8 %
	Precision:	85 %	65 %	10 %

Para el análisis automático de desempeño (Recall y Precision), vemos que los mejores resultados se obtienen con un $a_o > 0.3$, es decir con un umbral de exigencia θ_a mínimo, que hace posible considerar a la detección hipotética como detección válida. Debido a las imprecisiones del detector para este objeto en particular y a la forma asimétrica de este, la exigencia de $a_o > 0.5$, podría ser considerada como excesivamente optimista y en este contexto, una detección casi perfecta, lo cual no es –para este caso– viable.

El Recall obtenido para todas las evaluaciones realizadas, considerando $a_o > 0.3$ indica que esta propuesta de inspección activa, logra que en la mayoría de las inspecciones se llegue a una detección de un objeto de interés en una buena pose, independiente si la inspección comienza con una no detección ND o en una detección con mala pose.

5.5.2. Detección Activa de Hoja de Afeitar en Buena Pose

Para la inspección activa de hojas de afeitar usando la propuesta de *framework* hemos utilizado el algoritmo de eliminación de falsas alarmas (FP), es decir, el algoritmo que utiliza la restricción epipolar y trifocal. En este caso el seguimiento de la hoja de afeitar se realiza a partir de la primera imagen de rayos X capturada. La detección válida en la primera imagen, se configura como el punto de inicio de la inspección activa, y a partir de esta detección, se realizaron de manera automática los movimientos de rotación (definidos por el algoritmo de *Q-learning*) y adquisiciones de imágenes de rayos X, para la búsqueda del objeto de interés en una buena pose BP.

Para el análisis de desempeño, hemos realizado la inspección de dos bolsos, y considerado 180 inspecciones a cada uno de ello, es decir se incrementa cada 2°el ángulo de inicio

del manipulador robótico γ. De esta forma y tal como en el caso del revolver, consideramos en la evaluación automática distintos valores de umbral: $a_o > 0.3$, $a_o > 0.4$ y $a_o > 0.5$.

Tabla 5.9. Desempeño para la detección activa de una hoja de afeitar (primera bolso), usando *Q-learning* para estimar el movimiento del manipulador robótico y la restricción epipolar y trifocal para eliminar las falsas alarmas.

Número de inspecciones	Índice de desempeño	Desempeño		
		$a_o > 0.3$	$a_o > 0.4$	$a_o > 0.5$
180	Recall :	76 %	76 %	72 %
	Precision:	100 %	100 %	95 %

Vemos que para la tabla 5.9 los mejores índices de Recall se dan para $a_o > 0.3$ y $a_o > 0.4$, lo que indica que la detección no ocurre perfectamente centrada, y el tamaño y orientación del cuadro de detección BB_{dt} no siempre tienen el mismo tamaño y orientación que el cuadro de etiqueta BB_{gt} (*ground truth*), sin embargo, en términos cualitativos la detección final se aprecia muy cercana a una detección ideal y esta se encuentra en una buena pose BP. Los valores de Precision de 100 % indican que el algoritmo de eliminación de falsas alarmas (FP) que utiliza la restricción epipolar y trifocal funciona correctamente.

Tabla 5.10. Desempeño para la detección activa de una hoja de afeitar (segundo bolso), usando *Q-learning* para estimar el movimiento del manipulador robótico y la restricción epipolar y trifocal para eliminar las falsas alarmas.

Número de inspecciones	Índice de desempeño	Desempeño		
		$a_o > 0.3$	$a_o > 0.4$	$a_o > 0.5$
180	Recall :	87 %	87 %	82 %
	Precision:	100 %	100 %	95 %

En la tabla 5.10 vemos que para todos los valores de umbral ($a_o > 0.3$, $a_o > 0.4$ y $a_o > 0.5$) los indices de Recall superan el 80 % y al igual que en la evaluación mostrada en la tabla 5.9 los mejores valores se dan cuando se considera $a_o > 0.3$ y $a_o > 0.4$. Nuevamente fue demostrado el correcto funcionamiento del algoritmo para eliminar las falsas alarmas (FP), con indices de Precision de 100 % y 95 %.

A continuación mostramos algunas secuencias que demuestran la efectividad del framework mejorado, aquí vemos como desde una no detección (ND) o una mala pose, es posible llegar a una buena pose (BP), a partir de una segunda, tercera y hasta una cuarta adquisición de imagen de rayos X. Los movimientos del manipulador robótico, estimados por el algoritmo de *Q-learning*, resultan ser en su gran mayoría los adecuados, y las mayores complicaciones se deben al desorden interno de los bolsos inspeccionados, lo cual influye negativamente en el desempeño del detector AISM, provocando falsas alarmas difíciles de eliminar, y en consecuencia una disminución global del desempeño del *framework*.

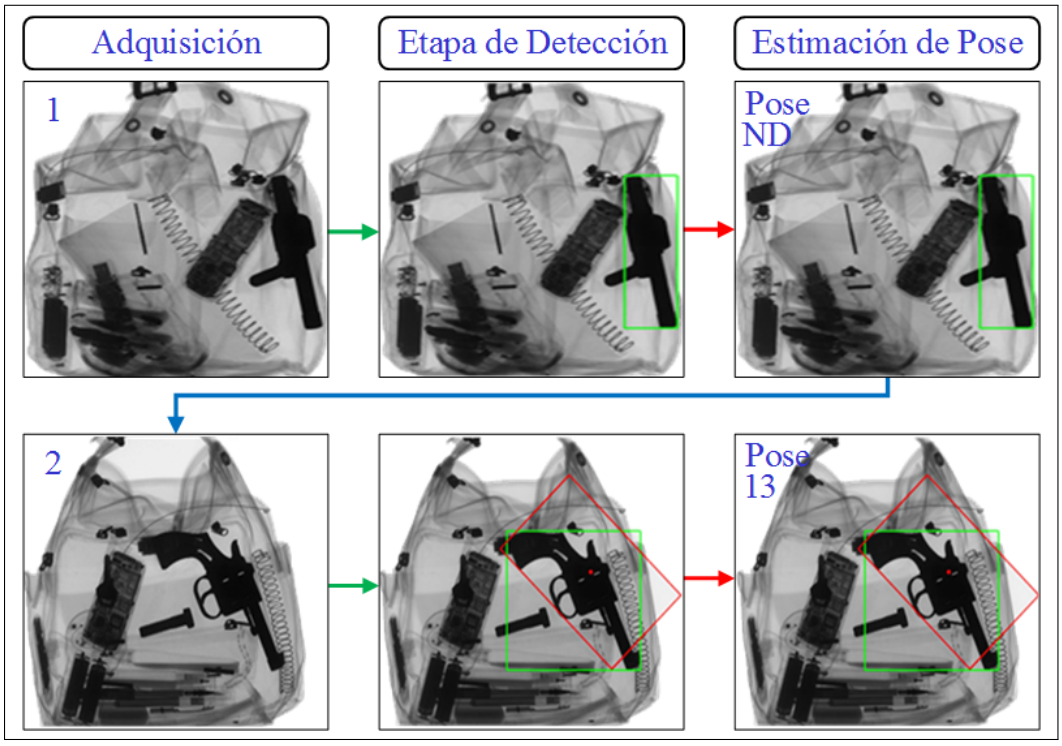


Figura 5.17. Inspección de revolver: desde una no detección (ND), se logra una buena pose (13).

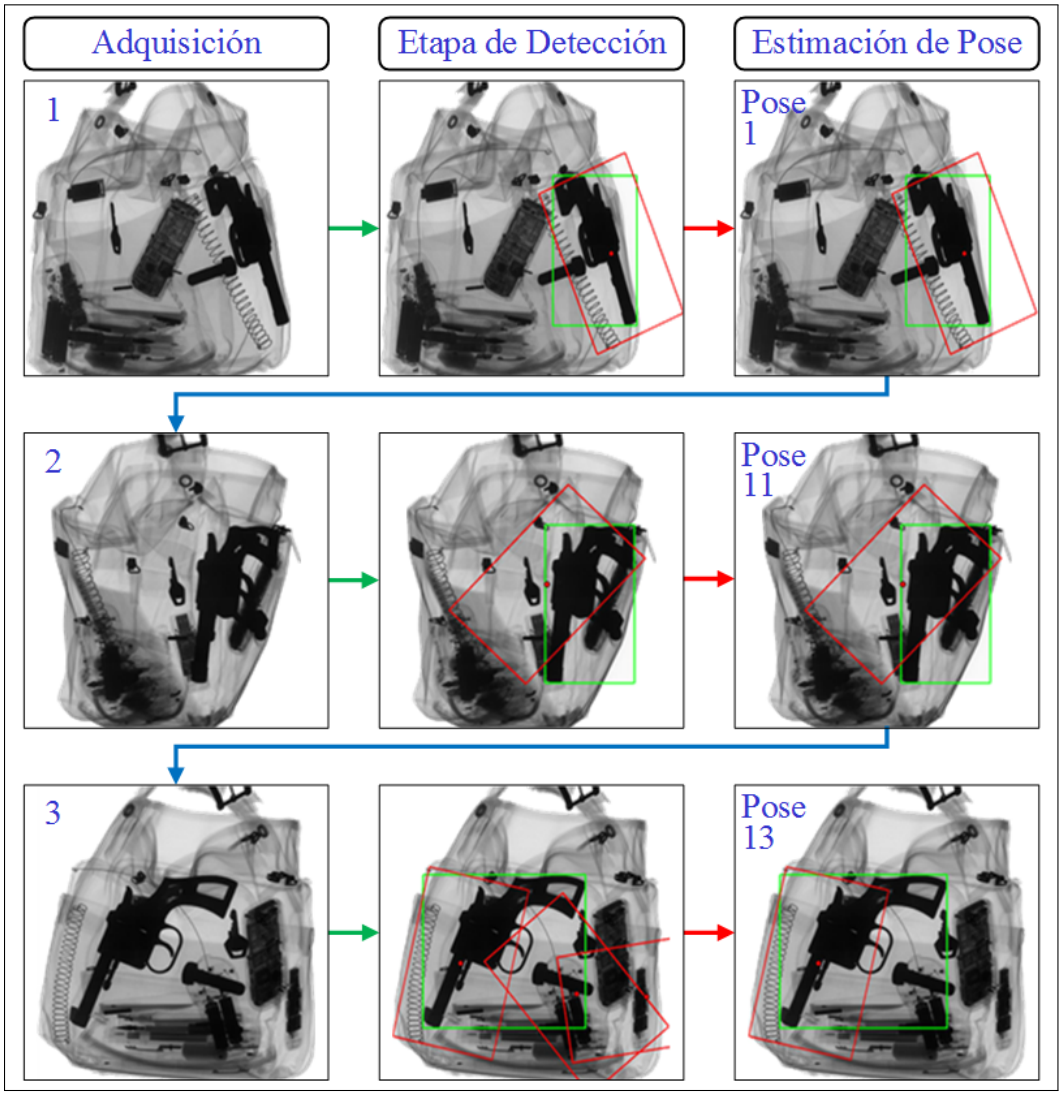


Figura 5.18. Inspección de revolver: desde una mala pose (1), se logra una buena pose (13).

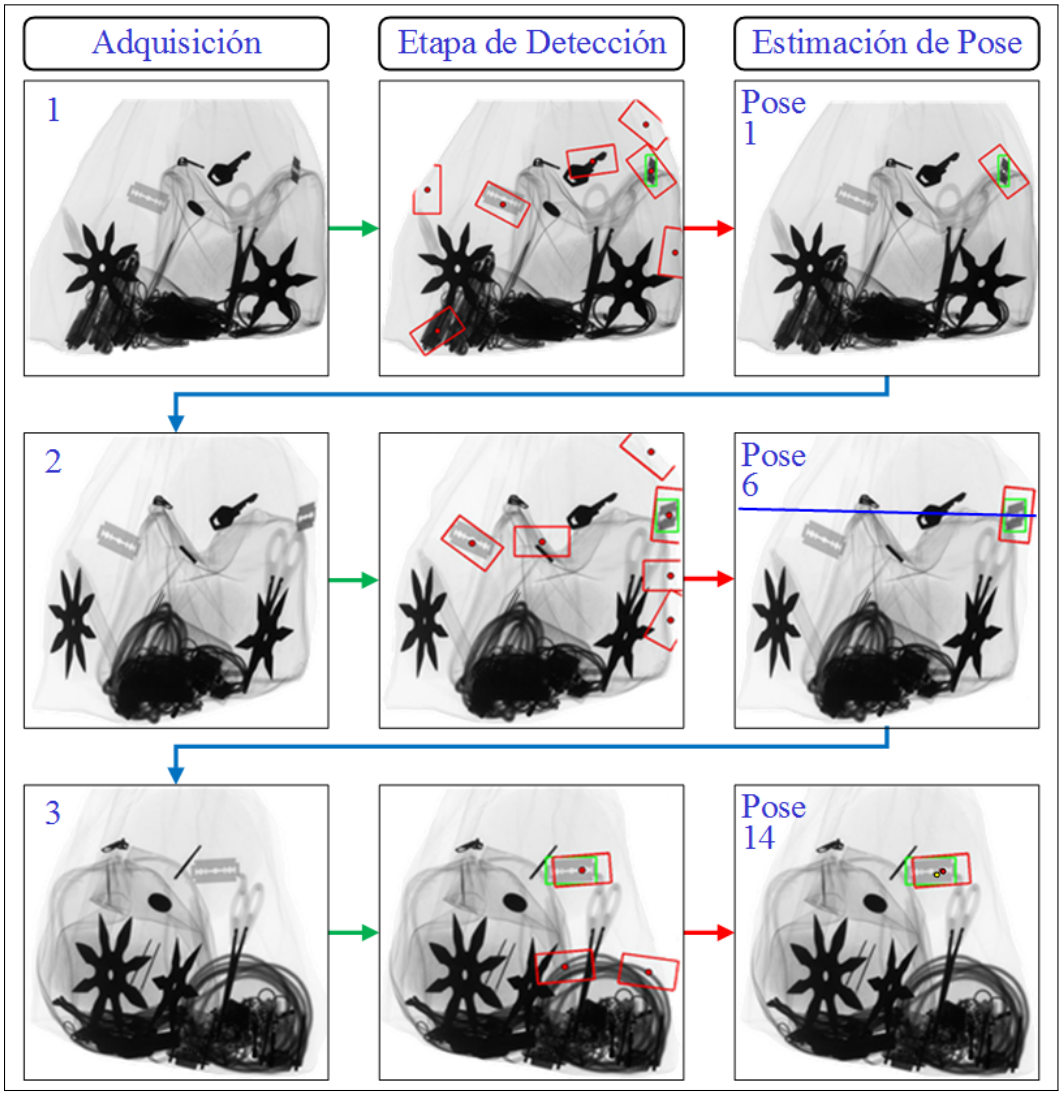


Figura 5.19. Inspección de una hojas de afeitar: desde una mala pose (1), se logra una buena pose (14). Aquí, se utiliza la restricción epipolar a contar de la segunda adquisición y luego la restricción trifocal a contar de la tercera adquisición, para eliminar falsas alarmas.

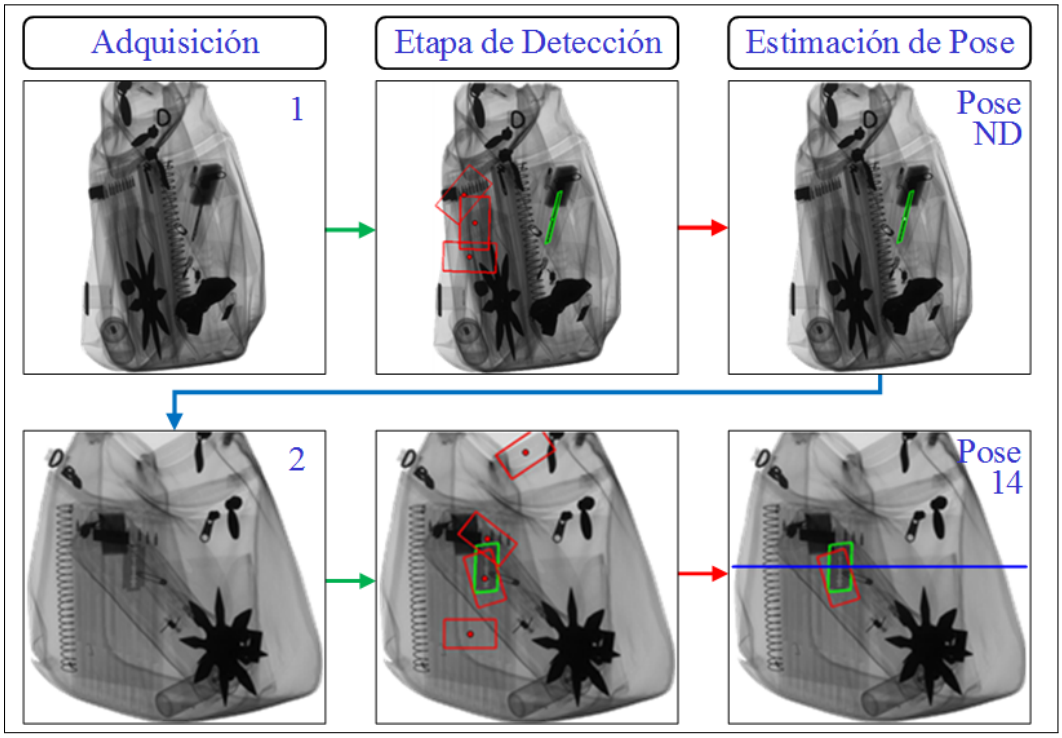


Figura 5.20. Inspección de hoja de afeitar: desde una no detección (ND), se logra una buena pose (14). Aquí, se utiliza la restricción epipolar a contar de la segunda adquisición para eliminar falsas alarmas.

5.5.3. *Q-learning* Para Estimar el Movimiento

En esta propuesta mejorada hemos utilizado el algoritmo de aprendizaje por refuerzo *Q-learning* para estimar el movimiento que debe realizar el manipulador robótico. Nuestros experimentos no están enfocados en medir el desempeño de *Q-learning* en forma individual, sino que el desempeño de la propuesta en su globalidad. De todas formas, en esta sección hacemos un análisis de las bondades del algoritmo estimador de movimiento con *Q-learning* usado en esta propuesta mejorada y lo comparamos con el algoritmo estimador de movimiento de tipo heurístico visto en la sección 4.1.5 y que hemos utilizado en la propuesta inicial de *framework*. Los aspectos considerados en este análisis son los siguientes:

a) Funcionalidad, y

b) Ventajas comparativas.

A. Funcionalidad

Tanto en la propuesta inicial como en la mejorada hemos limitado a cuatro el número de adquisiciones de imágenes de rayos X, por lo cual sólo fue necesario estimar tres movimientos con *Q-learning* en el peor de los casos, es decir, cuando no fue posible en la 1a, 2a ni 3a adquisición detectar el objeto de interés en una buena pose, lo cual se puede ver en la Tabla 5.11. Los resultados de inspección realizados con la propuesta inicial se pueden ver en las Tablas 5.2 y 5.3, mostrando que en la mayoría de los casos bastó uno o dos movimientos para lograr encontrar el objeto de interés en una buena pose, sin embargo, los resultados de la propuesta mejorada señalados en la Tabla 5.11 muestran que en la mayoría de los casos se alcanzó una buena pose después de dos o tres estimaciones de movimiento con *Q-learning*. Esto se debe a que la propuesta mejorada permitió inspecciones más complejas (bolsos con mayor tamaño, número de objetos y desórdenes internos) y desde el punto de vista funcional, *Q-learning* logró hacer las estimaciones de movimientos necesarios para encontrar al objeto de interés en una buena pose, en situaciones muy desfavorables.

Tabla 5.11. Número de detecciones realizadas para cada una de las adquisiciones en la propuesta mejorada.

Objeto	$a_o >$	Adquisición N°:				FP	ND
		1	2	3	4		
Hoja de	0.3	49	47	36	5	0	43
afeitar	0.4	49	47	36	5	0	43
(1er bolso)	0.5	47	43	35	5	7	43
Hoja de	0.3	58	47	31	20	0	24
afeitar	0.4	58	47	31	20	0	24
(2do bolso)	0.5	56	43	29	20	8	24
Revólver	0.3	43	29	17	25	28	38
	0.4	31	20	11	18	62	38
(1er bolso)	0.5	12	10	5	9	106	38
Revólver	0.3	54	39	20	13	23	31
	0.4	44	31	16	6	52	31
(2do bolso)	0.5	10	4	1	0	134	31

En la Figura 5.21 evidenciamos de mejor forma que *Q-learning* en la mayoría de los casos, después de la segunda estimación de movimiento, logra la detección de un objeto de

interés en una buena pose. La Figura 5.21 es un diagrama en cascada que muestra las 180 inspecciones realizadas a un bolso (2do bolso) para detectar una hoja de afeitar, el número total de adquisiciones de imágenes de rayos X (adquisición 1, 2, 3 y 4) en que se alcanzó la detección en una buena pose y el número de estimaciones de movimiento necesarias, considerando $a_o > 0.3$.

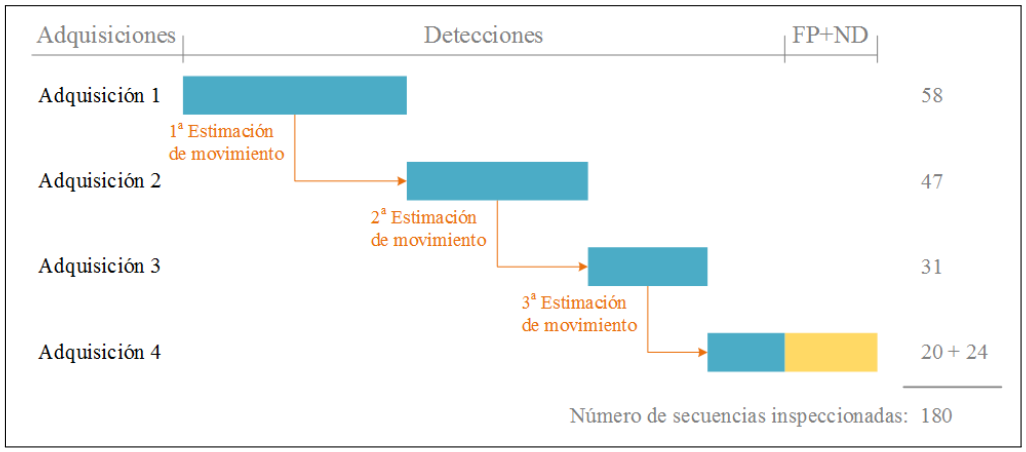


Figura 5.21. Detalle del número de adquisiciones y estimaciones de movimiento con *Q-learning* necesarias para alcanzar la detección de una hoja de afeitar en una buena pose, dispuesta al interior de un bolso (2do bolso) y considerando $a_o > 0.3$.

B. Ventajas Comparativas

Tal como señalamos en la parte final de la sección 4.2.2 *Q-learning* permite modelar las incertezas propias de entorno más complejos, como en los que fueron desarrollados los experimentos de esta propuesta mejorada. A diferencia del método de tipo heurístico de la propuesta inicial, *Q-learning* si incorpora en su modelo las incertezas, haciendo más robusta la estimación de movimiento.

Una de las incertezas que *Q-learning* incorpora en su modelo, es la producida por las limitaciones de movimiento del manipulador robótico que impide alcanzar una buena pose, y en este sentido *Q-learning* es muy superior a la propuesta de tipo heurística, que para hacer detecciones de objetos de interés (en este caso hojas de afeitar) que se encuentren dispuestos de forma horizontal y vertical, necesita de un sistema de sujeción con dos grados de libertad, donde cada grado de libertad tiene asociado una heurística distinta (ver sección 4.1.5), que pueden ser mejor comprendidas al visualizar las Figuras 4.3 y 4.5, que señalan

la heurística de rotación en torno al eje X y en torno al eje Z respectivamente, haciendo necesario el uso de un sistema de sujeción semi-automático con dos grados de libertad. Cuando usamos un manipulador robótico con un modelo de estimación de movimiento de tipo heurístico, sólo podemos detectar objetos de interés que estén paralelos al eje de rotación, y aprovechar subjetivamente la invarianza a la rotación de los descriptores que SIFT que permiten la caracterización de los objetos de interés.

En los experimentos que permitieron evaluar la propuesta mejorada, utilizamos un manipulador robótico que con sólo un grado de libertad nos permitió hacer inspecciones en entornos complejos, facilitado por el modelo *Q-learning* de estimación de movimiento, que incorpora la incerteza originada por las limitaciones de movimiento del manipulador robótico.

Otra situación en que comparativamente el modelo de estimación de movimiento con Q-learning supera al modelo de tipo heurístico, es que *Q-learning* considera las incertezas provocadas por simetrías de tipo espejo, permitiendo alcanzar una detección del objeto de interés en una buena pose en el número de movimientos que estipula el modelo, es decir, en uno o dos, según el detector haya estimado correctamente o no la pose asociada a la detección. En cambio el modelo de estimación de movimiento de tipo heurístico sólo considera las poses predefinidas, para posteriormente estimar el movimiento, a partir de una pose que pudiese ser incorrecta y no contemplada en el modelo (ver Figura 4.3).

La situación descrita en el párrafo anterior se puede ver en el ejemplo de inspección para la detección de un revólver en una buena pose y que es mostrado en la Figura 5.22, considerando el modelo propuesto en la Figura 4.10 y 4.11: a partir de la pose 24 *Q-learning* estima realizar dos movimientos, uno que le permita llegar a la pose 25 y luego otro hacia la pose 23 (excepcionalmente buena pose), sin embargo, bastó sólo un movimiento hacia la pose 23, lo cual indica que la pose inicial 24 en realidad correspondía a la pose 22, es decir una incerteza originada por la simetría de tipo espejo.

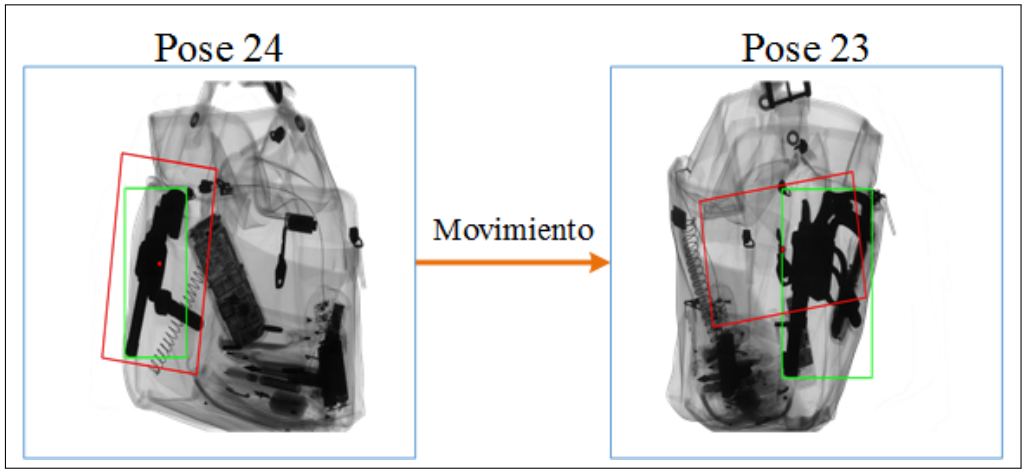


Figura 5.22. Movimiento estimado por *Q-learning* para llegar desde una pose 24 hacia una pose 23.

Capítulo 6. DISCUSIÓN Y CONCLUSIONES

En este capítulo presentamos una discusión acerca de nuestra investigación, las contribuciones y conclusiones de esta tesis. En particular centraremos este capítulo en mostrar las ventajas de nuestra propuesta de *framework* de inspección activa con rayos X y las limitaciones de uso. Finalmente, mostramos algunas ideas de trabajos futuros que pueden ser abordadas y que podrían mejorar nuestra propuesta en algunas de sus etapas.

6.1. Discusión

Para simplificar la lectura y facilitar el entendimiento, separamos la discusión según los contenidos abordados. En particular, hardware, algoritmos y desempeño.

6.1.1. Hardware

Nuestra propuesta de inspección activa con rayos X, tiene como objetivo la detección automática de algún objeto de interés que se encuentre en oclusión y/o en una mala pose al interior de un bolso u otro compartimiento. Para lograr este propósito es necesario contar con un sistema de inspección que tiene una infraestructura de hardware de elevado costo: *i*) emisor y detector de rayos X, *ii*) sistema de sujeción (manipulador robótico), *iii*) una cabina de plomo para aislar al sistema y evitar la refracción de radiación ionizante[1].

Como mencionamos en el capítulo 1, existen tres configuraciones que podrían ser utilizadas para la adquisición de las imágenes de rayos X y así realizar una inspección activa: 1) movimiento del emisor, 2) movimiento simultáneo del emisor y detector , y 3) movimiento del objeto. Estas configuraciones son mostradas en la Figura 1.3. Por razones técnicas y económicas, en nuestra propuesta hemos decidido mover el objeto que será inspeccionado, lo cual tiene dos restricciones que acotan la utilización de nuestra propuesta:

a) sólo es posible inspeccionar objetos rígidos, o realizar movimientos que impidan que el contenido del objeto inspeccionado tenga desplazamientos internos (objetos en estado de reposo),

[1]Para nuestros experimentos, el equipamiento tiene un precio aproximado de US$100.000 (cien mil dólares).

b) dependiendo del tipo de manipulador robótico o sistema de sujeción utilizado, existirán movimientos prohibidos, tales como aquellos que provoquen que el manipulador o sistema de sujeción causen oclusiones.

Estas restricciones se eliminarían totalmente si el objeto inspeccionado estuviese estático y pudiésemos mover simultáneamente el emisor y detector de rayos X, es decir, utilizar un "sistema robotizado de arco en C". El sistema de arco en C se usa frecuentemente en el área médica para la detección de fracturas y otros análisis médicos en humanos.

6.1.2. Algoritmos

La visión activa no es un concepto nuevo en el área de visión por computador, ya que ha sido usada principalmente en las aplicaciones de robótica móvil, sin embargo, con nuestra propuesta hemos dado los primeros pasos de inspección activa con rayos X aplicada a la inspección de equipaje.

La propuesta de visión activa para la inspección con rayos X que hacemos en esta tesis, involucra una serie de algoritmos, y uno de los principales es el algoritmo de detección en una vista. Diseñamos un algoritmo para la detección de objetos que trabaja directamente sobre las imágenes adquiridas (en escala de grises, con un sólo nivel de energía, sin algoritmos de segmentación basado en pseudocolor). Así, el algoritmo de detección, ha sido concebido para realizar el proceso de detección de forma automática, y si se amplía la cantidad de objetos a detectar, y se aumenta la velocidad de ejecución, el detector podría automatizar (o servir de apoyo) a los actuales procesos de inspección realizados en los aeropuertos o centros aduaneros.

Cada objeto de interés que fue detectado, debió previamente haber sido caracterizado. En esta investigación, utilizamos tres objetos amenazantes: hoja de afeitar, estrellas ninja (*shuriken*) y revólveres. El proceso de caracterización es parte de la etapa de entrenamiento y dos de las contribuciones asociadas a esta etapa de la tesis, son las siguiente: a) la metodología, que permite realizar la caracterización de los objetos de interés, y b) la base de datos con imágenes de rayos X de objetos amenazantes, en diferentes poses.

En la etapa de entrenamiento hemos almacenado los descriptores SIFT que caracterizan al objeto de interés, además hemos asociado la pose del objeto a cada imagen de entrenamiento. El algoritmo de estimación de pose es aplicado después de la detección y está basada en la votación de descriptores SIFT contenidos en la ventana de detección, y que son similares a los descriptores SIFT resultantes de la caracterización del objeto de interés y que se encuentran almacenados en una base de datos. Este algoritmo resulta muy simple de implementar y los errores en la estimación de pose ocurren cuando la ventana de detección no queda del todo alineada con el objeto que fue detectado, es decir, descriptores que no forman parte del objeto de interés y que son asociados erróneamente a una pose.

El algoritmo de estimación de movimientos que debe realizar el sistema automático o semi-automático para lograr una siguiente mejor vista, es realizada de dos formas: a) en la propuesta inicial desarrollada en esta tesis, usamos una estimación de movimiento que podríamos denominar de tipo heurístico, y b) en la propuesta mejorada usamos un modelo de Q-learning. En ambos casos los resultados son cualitativamente buenos, ya que al observar las secuencias de inspección, vemos que la pose estimada en la siguiente vista, una vez provocado el movimiento, resultó ser una buena pose o una más próxima a ella. Sin embargo, el algoritmo de *Q-learning* permitió estimar el movimiento de un manipulador robótico con un solo grado de libertad, usado en entornos más complejos de inspección radioscópica y con variados niveles de incertezas. Las incertezas que incorporamos a *Q-learning* son producidas por: a) las imperfecciones del detector de objetos de interés al detectar y estimar la pose, lo cual ocurre principalmente por la simetría y/o disposición del objeto de interés al ser caracterizado, provocando imágenes de tipo espejo que dificultan la estimación de la real pose durante la inspección, b) la restricción de movimiento del manipulador robótico, que hará imposible alcanzar algunas buenas poses, y c) las réplicas de los cuadrantes de la base de datos de imágenes de rayos X para el entrenamiento, que hace confundir al estimador de pose. Así, el estimador de movimientos basado en *Q-learning* es capaz de encontrar la política óptima de movimientos que minimizan la cantidad de imágenes necesarias para encontrar al objeto de interés en una buena pose.

Un algoritmo complementario a nuestra propuesta y que hemos utilizado en la detección activa de hojas de afeitar, es el de eliminación de falsas alarmas, que hace uso de la restricción epipolar y trifocal. Este algoritmo demostró ser muy efectivo, pues en la mayoría de nuestros experimentos logró la eliminación total de falsas alarmas, incrementando de esta forma el índice de Precision de la propuesta *framework* de inspección activa. Para poder usar la restricción epipolar y trifocal se debió modelar geométricamente el sistema de inspección.

En la propuesta mejorada de inspección activa con rayos X hemos usando una adaptación del detector ISM, en el cual podrían existir tres causales de sobre ajuste o sobre entrenamiento: *i*) utilizar las mismas imágenes para la etapa de entrenamiento y para la etapa de detección, *ii*) utilizar en el proceso de entrenamiento un tamaño inadecuado para el vocabulario visual y para las muestras de *keypoints*, y *iii*) utilizar todas las imágenes de prueba para hacer los ajustes de parámetros y valores de umbral del detector. Estos inconvenientes, se resolvieron de la siguiente forma:

i) Si bien los objetos utilizados en el entrenamiento y la detección son los mismos, las imágenes son totalmente distintas. Para la creación de los tres vocabularios visuales (de las hojas de afeitar, *shurikens* y revólveres) hemos utilizado imágenes de rayos X de los objetos aislados, dispuestos en una esfera de poliestireno expandido (EPS), la cual tiene casi nulo coeficiente de absorción lineal, es decir, condiciones consideradas ideales y adquiridas exclusivamente para este proceso. No se recortaron imágenes de los objetos de interés dispuestos en el ambiente real de detección (interior de un bolso). Adicionalmente dichos objetos fueron dispuestos en ángulos que permiten obtener las poses más representativas, como se muestra en la Figura 4.7 4.7. De ninguna manera se consideraron para crear el vocabulario visual las infinitas poses que pudiese tener el objeto de interés durante el proceso de detección.

ii) En Guo et al. (2013) y Nowak et al. (2006) se reporta que para cada detector hay una mejora sustancial en el desempeño a medida que aumenta el tamaño del vocabulario visual (*codebook*), pero se produce un sobre ajuste para vocabularios visuales de gran tamaño,

en nuestro caso el vocabulario visual para cada objeto de interés está formado por 400 elementos, lo cual es un número razonable e inferior al tamaño que podría causar sobre ajuste (sobre 3.000). En los métodos basados en *keypoints* también hay evidencia que para un gran número de muestras puede existir sobre ajuste, en nuestro caso fueron muestreados como máximo 500 descriptores SIFT, número que está por debajo de los valores que podría causar sobre ajuste (sobre 1.000).

iii) Para hacer los ajustes de los parámetros y valores de umbral del detector AISM se utilizó un conjunto de imágenes distinto a las imágenes de prueba.

6.1.3. Desempeño

En la primera propuesta de inspección activa desarrollamos un detector de hojas de afeitar que usa el enfoque de similitud de descriptores SIFT (similitud de descriptores que aparecen en la imagen inspeccionada con los que caracterizan al objeto de interés almacenados en una base de datos). En este caso no se evalúa el desempeño del detector, sino más bien el desempeño global de la propuesta (Precision y Recall), y los resultados globales fueron muy auspiciosos.

La segunda propuesta de inspección activa usa una adaptación al modelo de forma implícita (ISM), al cual hemos denominado AISM (*Adapted Implicit Shape Model*), en este caso se midieron los indices de desempeño TPR y FPR, sobre base de datos propias, para detectar hojas de afeitar, *shurikens* y revólveres. Los resultados de desempeño de este detector (AISM), son comparativamente mejores que con otro tipo de detectores conocidos, tales como: el detector SIFT basado en la propuesta de Lowe (2004), el detector original ISM propuesto por Leibe et al. (2008) y nuestra propuesta de detector AISM usando descriptores SURF propuestos por Bay et al. (2008). Con nuestra propuesta de detector AISM hacemos un doble aporte al estado del arte; a) tres bases de datos con imágenes de rayos X, conteniendo objetos de interés etiquetados (*ground truth*) en entornos reales, y b) una propuesta de detector de objetos amenazantes (objetos de interés) que usa una adaptación al modelo de forma implícita.

Para evaluar el desempeño del detector AISM y de la propuesta mejorada de inspección activa, hemos usado el método PASCAL, es decir, que el área de overlap supere un cierto valor de umbral. Debido a las imperfecciones al estimar la orientación de la ventana de detección, es que los mejores resultados se obtuvieron con áreas de overlap superiores a 0.3, que pareciese ser un valor pequeño, pero que visualmente se ajusta bien a la detección real. Debemos considerar que en el algoritmo de detección en una vista, el objeto de interés puede estar en cualquier pose, por lo tanto, la etiqueta (*ground truth*) tiene siempre un tamaño menor que la ventana de detección, es por eso que los valores pequeños del umbral para el área de overlap, parecen ser adecuados y ser útiles en el proceso de detección activa.

Hemos observado que existen situaciones de inspección que son imposibles tanto para un inspector humano como para cualquier sistema automático, es decir, altos niveles de oclusión y desórdenes internos, en estos casos la aplicación de nuestra propuesta se ve desfavorecida y no recomendada. Un ejemplo que ilustra esta esta discusión se ve en los aeropuertos, donde los inspectores humanos se han dado cuenta que los computadores portátiles generan altos niveles de oclusión y por ende, obligan a sus propietarios a sacarlos de sus bolsos y mostrarlos.

6.2. Conclusiones

En esta tesis, proponemos un *framework* que trabaja con múltiples vistas de imágenes de rayos X y que es capaz de inspeccionar eficazmente objetos complejos usando visión activa. Hemos demostrado que es posible automatizar el proceso de detección de objetos de interés (en esta tesis, objetos amenazantes) en sólo una imagen de rayos X y que al incorporar un sistema automático que permita mover activamente a los objetos que son inspeccionados, desde una pose difícil (en la que la detección es prácticamente imposible) hasta una más fácil (en la que el reconocimiento es más simple). De esta forma la inspección radioscópica activa en múltiples vistas, realizada de manera automática es capaz de superar problemas propios de la inspección en una vista, como lo es la no detección y algunos niveles de oclusión, aumentando así la certeza de la detección. Con esta investigación

hemos dado los primeros pasos en la inspección radioscópica activa, con resultados auspiciosos, que nos llevan a pensar que un sistema de estas características dispuesto en lugares específicos podría aumentar los niveles de seguridad de las personas.

La propuesta inicial, fue desarrollada para la detección de una hoja de afeitar, que en lugares como aeropuertos y centros de aduana se considera un objeto amenazante. Aunque la detección de una hoja de afeitar puede parecer simple, la evidencia experimental nos mostró la complejidad de resolver este problema, debido a la simetría en todos los cuadrantes, la delgadez, la pequeñez, y el bajo coeficiente de absorción de rayos X.

Inicialmente en el desarrollo de esta tesis hemos usando un brazo robótico y un sistema manipulador semi-automático, la robustez y la fiabilidad del método se han verificado en la detección automática de hojas de afeitar ubicadas dentro de nueve objetos diferentes, mostrando resultados que son prometedores: en 130 experimentos hemos podido detectar 115 veces la hoja de afeitar con 10 falsas alarmas, logrando un Recall de 89 % y Precision de 92 %. Podemos señalar que nuestra primera propuesta de inspección activa resultó ser prometedora y nos dio el marco referencial y metodológico para ampliar nuestro enfoque y así poder realizar inspecciones más realistas.

En nuestra investigación implementamos un detector en una vista, para lo cual hemos usado el conocido modelo de forma implícita ISM y que hemos adaptado para su uso en imágenes de rayos X, a esta adaptación la hemos denominado AISM (*Adapted* ISM). Las principales adaptaciones son las siguientes: *i*) cuando se extraen *keypoints* de la imagen de prueba, sólo los *keypoints* útiles son considerados por AISM, y *ii*) en ISM, sólo la ocurrencia con la puntuación de similitud más alta se considera que es una detección válida, mientras AISM fusiona las ocurrencias con las puntuaciones de similitud cuyos valores exceden un cierto umbral.

Los bolsos utilizados en los experimentos que permitieron validar AISM contenían aproximadamente 20 objetos, uno o dos de los cuales eran objetos de interés. Los experimentos se llevaron a cabo en tres diferentes conjuntos de datos (hojas de afeitar, *shuriken* y revólveres) con 200 imágenes de rayos X cada uno. Mostramos la solidez del enfoque para

los objetos de forma regular (hoja de afeitar y *shuriken*) con una alta tasa de verdaderos positivos (TPR $>$ 0.95) con una baja tasa de falsos positivos (FPR = 0.05), para ambos objetos. En el caso de revólveres, se obtuvieron tasas aceptables (TPR = 0.89 con FPR = 0.18) debido a la forma irregular del objeto, lo que causó detecciones dispersas dentro y fuera del objeto. Estos resultados preliminares son prometedores porque establecen un método para la detección de las categorías de objetos en imágenes individuales de rayos X. Además, comparamos nuestro algoritmo con otros tres métodos que pueden utilizarse para detectar objetos, tales como los ya señalados en la sección 6.1.3: el detector SIFT basado en la propuesta de Lowe (2004), el detector original ISM propuesto por Leibe et al. (2008) y nuestra propuesta de detector AISM usando descriptores SURF propuestos por Bay et al. (2008). Como las curvas ROC demuestran, nuestro método supera a todos estos métodos en términos de las tasas de verdaderos y falsos positivos positivos.

Nuestro enfoque AISM fue diseñado para detectar sólo una categoría objeto a la vez, pero repetimos la estrategia de entrenamiento y pruebas en tres categorías de objetos diferentes, con el fin de mostrar la eficacia del método propuesto. Por lo tanto, varias categorías de objetos pueden ser detectadas simultáneamente. Usando otros conjuntos de datos de entrenamiento representativos, nuestra metodología podría ser utilizada en la detección de otros objetos amenazantes (por ejemplo; cuchillos). Creemos que nuestro algoritmo podría ser una herramienta útil para los inspectores humanos.

Los resultados obtenidos en la primera etapa del desarrollo de esta tesis nos motivaron a implementar una propuesta mejorada, para la cual incorporamos un nuevo detector y un estimador de movimiento basado en *Q-learning*, de esta forma realizamos la inspección de bolsos para encontrar hojas de afeitar y revólveres. El algoritmo de *Q-learning* propuesto en esta tesis es capaz de superar algunos niveles de oclusión e incertezas, lo cual fue incorporado al modelo estado-acción. El algoritmo de *Q-learning* incorpora en su modelo tres tipos de incertezas: a) originada por la simetría espejo, b) originada por las limitaciones de movimiento del manipulador robótico, y c) originadas por las replicas de los cuadrantes de las imágenes de entrenamiento, haciéndolo más robusto y fácil de utilizar con un

manipulador robótico con un sólo grado de libertad, ya que contempla estimar el movimiento cuando el objeto de interés se encuentra en forma horizontal o vertical respecto al eje de rotación del manipulador robótico, superando de esta forma al estimador de movimiento de tipo heurístico que se ve restringido a asumir poses incorrectas ya que las poses reales no están contempladas en el modelo. Para superar esta restricción el modelo de tipo heurístico requiere la utilización de un sistema de sujeción de objetos con dos grados de libertad (en nuestro caso un sistema semi-automático) y adoptar una heurística adicional de movimiento.

Para medir el desempeño de la propuesta mejorada de inspección activa de revólveres usamos dos bolsos, los cuales fueron inspeccionados 180 veces cada uno, obteniendo un Recall 63 % y 70 % con una Precision de 80 % y 84 %, para cada bolso, respectivamente y relajando la condición del área de *overlap* ($a_o > 0.3$). Para la inspección activa de hojas de afeitar, a nuestra propuesta de *framework* mejorada le incorporamos además la eliminación de falsas alarmas con la restricción epipolar y trifocal, y en este caso usamos dos bolsos, los cuales fueron inspeccionados 180 veces cada uno, usando cada vez un punto de vista inicial distinto, para así obteniendo un Recall de 76 % y 87 % con una Precision de 100 % y 100 %, para cada bolso, respectivamente y con $a_o > 0.3$. Hacemos notar que el valor de Precision de 100 % para la detección activa de una hoja de afeitar, indica la efectividad del algoritmo de eliminación de falsas alarmas. Estos resultados son buenos y nos llevan a decir que la inspección activa con rayos X, es aplicable como un apoyo a procesos de inspección que son realizados por inspectores humanos.

La propuesta mejorada de *framework*, fue desarrollada para la detección de una hoja de afeitar y revólveres, situados en contenedores (bolsos), con altos grados de complejidad (desorden interno, cantidad de objetos y mayores niveles de oclusión), es decir, situaciones más realistas que las consideradas en la propuesta inicial. A partir de los resultados obtenidos en ambas propuestas de *framework* (inicial y mejorada), podemos concluir que la inspección activa permite la detección de objetos amenazantes que se encuentran en poses intrincadas o poco representativas y esta detección se facilita, pues permite eficazmente

soslayar algunos niveles de oclusión y desórdenes internos que son difíciles o imposibles de resolver en inspecciones tradicionales de una vista.

6.3. Trabajo Futuro

Durante todo el desarrollo de este trabajo de investigación, nos hemos propuesto dar los primeros pasos en la inspección de imágenes de rayos X con visión activa, buscando que nuestra propuesta sea eficaz y superen algunas restricciones de hardware y algorítmicas. Los resultados que hemos obtenido en todas las etapas de esta investigación distan de la perfección, pero son promisorios, quedando una serie de temas que pueden seguir siendo investigados. Como ideas de trabajo futuro proponemos:

a) Ampliar las posibilidades a inspección de objetos no rígidos, para lo cual se debe contar con un sistema que permita mover simultaneamente al emisor y detector de rayos X, en torno al objeto que se necesita inspeccionar (arco en C).
b) Extender las capacidades de inspección a otras categorías de objeto, con formas regulares e irregulares. Lo que implica el desarrollo de otros algoritmos de detección.
c) Creemos que el desempeño global de nuestra propuesta podría mejorar, si se mejora el algoritmo de detección en una vista. En este ámbito, una propuesta de trabajo futuro sería mejorar el proceso de caracterización de objetos del detector AISM, extrayendo los descriptores SIFT sólo en las regiones del objetos que son discriminativas. Por lo tanto, podríamos evitar, la caracterización de las regiones muy oscuras de los objetos con alta absorción de rayos X y tener en cuenta sólo las regiones de los objetos más cercanas a los contornos (gradientes altos).
d) Incorporar a la visión activa algoritmos que actualmente son utilizados por los inspectores humanos en los aeropuertos y que han demostrado ser de mucha utilidad, como pseudoclor a partir de emisiones de rayos X con dos niveles de energía (*dual-energy*). De esta forma podríamos facilitar y orientar los procesos de detección según el tipo de material y sustancia.
e) Sería interesante evaluar el desempeño de los operadores humanos con y sin la tecnología que en esta tesis hemos propuesto, de tal forma de incorporar modificaciones que

a los humanos les resulten más amigables, como por ejemplo, transformar etapas que son totalmente automáticas en etapas semi-automáticas, para permitir adaptaciones en tiempo real.

f) Creemos que el marco conceptual y experimental de la inspección activa con rayos X que en esta tesis hemos desarrollado para la detección de objetos amenazantes contenidos en bolsos, puede ser expandido a los actuales procesos de inspección que se realiza a los humanos.

g) En la propuesta inicial y la propuesta mejorada desarrolladas en este trabajo, específicamente en la etapa de estimación del siguiente movimiento, hemos definido arbitrariamente las *Buenas Poses*, considerando el criterio de que una Buena Pose es aquella que permite a una persona reconocer fácilmente y a simple vista al objeto de interés. Proponemos como trabajo futuro, la implementación de un algoritmo, que indique automáticamente las Buenas Poses de un objeto de interés, basado en una métrica de superficie visible, es decir, que la similitud de superficie de la imagen registrada en el plano de proyección sea de a lo menos un porcentaje elevado de la mayor superficie visible (por ejemplo: sobre un 70 % de la mayor superficie visible). De esta forma podremos decir que una buena pose de un objeto de interés, será aquella que un clasificador que trabaja con una sola vista, podrá clasificar correctamente en un porcentaje cercano al 100 %.

BIBLIOGRAFÍA

Abidi, B., Zheng, Y., Gribok, A., y Abidi, M. (2006). Improving weapon detection in single energy X-ray images through pseudocoloring. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *36*(6), 784-796.

Abusaeeda, O., Evans, J., D., D., y Chan, J. (2011). View synthesis of KDEX imagery for 3D security X-ray imaging. En *Proc. 4th international conference on imaging for crime detection and prevention (ICDP-2011).*

Agarwal, S., Awan, A., y Roth, D. (2004). Learning to detect objects in images via a sparse, part-based representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *26*(11), 1475-1490.

Als-Neielsen, J., y McMorrow, D. (2011). *Elements of modern X-ray physics* (Second ed.). Willey.

Ballard, D. (1981). Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, *13*(2), 111 - 122.

Baştan, M., Byeon, W., y Breuel, T. (2013). Object recognition in multi-view dual energy X-ray images. En *Proceedings of the british machine vision conference* (pp. 130.1–130.11). BMVA Press.

Baştan, M., Yousefi, M., y Breuel, T. (2011). Visual words on baggage X-ray images. En *Computer analysis of images and patterns* (Vol. 6854, pp. 360–368). Springer Berlin Heidelberg.

Bay, H., Ess, A., Tuytelaars, T., y Gool, L. V. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346 - 359. (Similarity Matching in Computer Vision and Multimedia)

Bolfing, A., Halbherr, T., y Schwaninger, A. (2008). How image based factors and human factors contribute to threat detection performance in X-ray aviation security screening. En *HCI and usability for education and work* (Vol. 5298, p. 419-438). Springer Berlin Heidelberg.

Carrasco, M., y Mery, D. (2006). Robust algorithm for nondestructive testing of weld seams. *Ultrasonic and Advanced Methods for Nondestructive Testing and Material Characterization*, 635–658.

Carrasco, M., y Mery, D. (2011). Automatic multiple view inspection using geometrical tracking and feature analysis in aluminum wheels. *Machine Vision and Applications*, *22*(1), 157-170.

Chan, J., Evans, P., y Wang, X. (2010). Enhanced color coding scheme for kinetic depth effect X-ray (KDEX) imaging. En *Security technology (ICCST), 2010 IEEE international carnahan conference on* (p. 155-160).

Chan, J., Omar, A., Evans, J. P. O., Downes, D., Wang, X., y Liu, Y. (2009). Feasibility of SIFT to synthesise KDEX imagery for aviation luggage security screening. En *Crime detection and prevention (ICDP 2009), 3rd international conference on* (p. 1-6).

Chen, Z., Zheng, Y., Abidi, B. R., Page, D. L., y Abidi, M. A. (2005). A combinational approach to the fusion, de-noising and enhancement of dual-energy X-ray luggage images. En *IEEE conference on computer vision and pattern recognition workshops (CVPRW).*

Cheng, Y. (1995). Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *17*(8), 790-799.

Comaniciu, D., Ramesh, V., y Meer, P. (2001). The variable bandwidth mean shift and data-driven scale selection. En *Computer vision, 2001. ICCV 2001. proceedings. eighth IEEE international conference on* (Vol. 1, p. 438-445 vol.1).

Ding, J., Li, Y., Xu, X., y Wang, L. (2006). X-ray image segmentation by attribute relational graph matching. En *8th IEEE international conference on signal processing* (Vol. 2).

Duan, X., Cheng, J., Zhang, L., Xing, Y., Chen, Z., y Zhao, Z. (2009). X-ray cargo container inspection system with few-view projection imaging. *Nuclear Instruments and Methods in Physics Research A*, *598*, 439-444.

Everingham, M., Gool, L. V., Williams, C. K. I., Winn, J., y Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, *88*(2), 303 - 338.

Faugeras, O., Luong, Q.-T., y Papadopoulo, T. (2001). *The geometry of multiple images:*

The laws that govern the formation of multiple images of a scene and some of their applications. Cambridge MA, London: The MIT Press.

Faugeras, O., y Robert, L. (1996). What can two images tell us about a third one? *International Journal of Computer Vision*, *18*(1), 5-19.

Fawcett, T. (2003). ROC graphs: notes and practical considerations for data mining researchers. *Intelligent Enterprise Technologies Laboratory, HP Laboratories Palo Alto*.

Franzel, T., Schmidt, U., y Roth, S. (2012). Object detection in multi-view X-ray images. En *Pattern recognition* (Vol. 7476, p. 144-154). Springer Berlin Heidelberg.

Frosio, I., Borghese, N., Lissandrello, F., Venturino, G., y Rotondo, G. (2011). Optimized acquisition geometry for X-ray inspection. En *Instrumentation and measurement technology conference (I2MTC), 2011 IEEE* (p. 1-6).

Guo, J., Qiu, Z., y Gurrin, C. (2013). Exploring the optimal visual vocabulary sizes for semantic concept detection. En *Content-based multimedia indexing (CBMI), 2013 11th international workshop on* (pp. 109–114).

Haff, R., y Slaughter, D. (2004). Real-time X-ray inspection of wheat for infestation by the granary weevil, sitophilus granarius (L.). *Transactions of the American Society of Agricultural Engineers*, *47*, 531-537.

Haff, R., y Toyofuku, N. (2008). X-ray detection of defects and contaminants in the food industry. *Sensing and Instrumentation for Food Quality and Safety*, *2*(4), 262-273.

Hardmeier, D., Hofer, F., y Schwaninger, A. (2006). The role of recurrent cbt for increasing aviation security screeners' visual knowledge and abilities needed in X-ray screening. , 338–342.

Hartley, R., y Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge Univ Pr.

Heitz, G., y Chechik, G. (2010). Object separation in X-ray image sets. En *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on* (p. 2093-2100).

Hellier, C. (2013). *Handbook of nondestructive evaluation* (Second ed.). McGraw Hill.

Hough, P. (1962). Methods and means for recognizing complex patterns. *U.S. Patent 3,069,654*.

Jiang, J.-A., Chang, H.-Y., Wu, K.-H., Ouyang, C.-S., Yang, M.-M., Yang, E.-C., ... a, T.-T. L. (2008). An adaptive image segmentation algorithm for X-ray quarantine inspection of selected fruits. *Computers and Electronics in Agriculture*, *60*, 190-200.

Kase, K. (2002). Effective use of color in X-ray image enhancement for luggage inspection. *Master's Degree thesis*.

Kwon, J.-S., Lee, J.-M., y Kim, W.-Y. (2008). Real-time detection of foreign objects using X-ray imaging for dry food manufacturing line. En *Proceedings of IEEE international symposium on consumer electronics (ISCE 2008)* (p. 1-4).

Leibe, B., Leonardis, A., y Schiele, B. (2004). Combined object categorization and segmentation with an implicit shape model. En *Workshop on statistical learning in computer vision (eccv)* (Vol. 2).

Leibe, B., Leonardis, A., y Schiele, B. (2008). Robust object detection with interleaved categorization and segmentation. *International Journal of Computer Vision*, *77*(1-3), 259-289.

Leibe, B., y Schiele, B. (2003). Interleaved object categorization and segmentation. En *Proceedings of the british machine vision conference* (p. 78.1-78.10). BMVA Press.

Li, X., Tso, S. K., , Guan, X.-P., y Huang, Q. (2006). Improving automatic detection of defects in castings by applying wavelet technique. *IEEE Transactions on Industrial Electronics*, *53*(6), 1927-1934.

Liao, T. W. (2008). Classification of weld flaws with imbalanced class data. *Expert Systems with Applications*, *35*(3), 1041 - 1052.

Liao, T. W. (2009). Improving the accuracy of computer-aided radiographic weld inspection by feature selection. *NDT & E International*, *42*(4), 229 - 239.

Lowe, D. (1999). Object recognition from local scale-invariant features. En *Computer vision, 1999. the proceedings of the seventh IEEE international conference on* (Vol. 2, p. 1150-1157 vol.2).

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91-110.

Lu, Q., y Conners, R. (2006). Using image processing methods to improve the explosive

detection accuracy. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *36*(6), 750-760.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. En *Proceedings of the fifth berkeley symposium on mathematical statistics and probability, volume 1: Statistics* (pp. 281–297). Berkeley, Calif.: University of California Press.

Mansoor, M., y Rajashankari, R. (2012). Detection of concealed weapons in X-ray images using fuzzy K-NN. *International Journal of Computer Science, Engineering and Information Technology*, *2*(2).

Megherbi, N., Breckon, T. P., y Flitton, G. T. (2013). Investigating existing medical CT segmentation techniques within automated baggage and package inspection. En *SPIE security + defence* (pp. 89010L–89010L).

Mery, D. (2003). Explicit geometric model of a radioscopic imaging system. *NDT & E International*, *36*(8), 587-599.

Mery, D. (2006). Automated radioscopic testing of aluminum die castings. *Materials Evaluation*, *64*(2), 135-143.

Mery, D. (2011a). Automated detection in complex objects using a tracking algorithm in multiple X-ray views. En *Proceedings of the 8th IEEE workshop on object tracking and classification beyond the visible spectrum (OTCBVS 2011), in conjunction with cvpr 2011, colorado springs* (p. 41-48).

Mery, D. (2011b). Automated detection of welding discontinuities without segmentation. *Materials evaluation*, *69*(6), 656–663.

Mery, D. (2014). Computer vision technology for X-ray testing. *Insight-Non-Destructive Testing and Condition Monitoring*, *56*(3), 147–155.

Mery, D. (2015). Inspection of complex objects using multiple-X-ray views. *Mechatronics, IEEE/ASME Transactions on*, *20*(1), 338-347.

Mery, D., y Berti, M. A. (2003). Automatic detection of welding defects using texture features. *Insight-Non-Destructive Testing and Condition Monitoring*, *45*(10), 676–681.

Mery, D., y Filbert, D. (2002). Automated flaw detection in aluminum castings based on

the tracking of potential defects in a radioscopic image sequence. *IEEE Transactions on Robotics and Automation*, *18*(6), 890–901.

Mery, D., Lillo, I., Riffo, V., Soto, A., Cipriano, A., y Aguilera, J. (2011). Automated fish bone detection using X-ray testing. *Journal of Food Engineering*, *2011*(105), 485-492.

Mery, D., Mondragon, G., Riffo, V., y Zuccar, I. (2013). Detection of regular objects in baggage using multiple X-ray views. *Insight: Non-Destructive Testing & Condition Monitoring*, *55*(1), 16 - 20.

Mery, D., Riffo, V., Zuccar, I., y Pieringer, C. (2013). Automated X-ray object recognition using an efficient search algorithm in multiple views. En *Computer vision and pattern recognition workshops (CVPRW), 2013 IEEE conference on* (p. 368-374).

Michel, S., Koller, S. M., y Schwaninger, A. (2008). Relationship between level of detection performance and amount of recurrent computer-based training. En *Security technology, 2008. ICCST 2008. 42nd annual IEEE international carnahan conference on* (pp. 299–304).

Newman, T., y Jain, A. (1995). A survey of automated visual inspection. *Computer Vision and Image Understanding*, *61*(2), 231-262.

Nowak, E., Jurie, F., y Triggs, B. (2006). Computer vision – ECCV 2006: 9th european conference on computer vision, graz, austria, may 7-13, 2006, proceedings, part iv. En (pp. 490–503). Springer Berlin Heidelberg.

Ogawa, Y., Kondo, N., y Shibusawa, S. (2003). Inside quality evaluation of fruit by X-ray image. En (Vol. 2, p. 1360-1365 vol.2).

Pieringer, C., y Mery, D. (2010). Flaw detection in aluminium die castings using simultaneous combination of multiple views. *Insight*, *52*(10), 548–552.

Pizarro, L., Mery, D., Delpiano, R., y Carrasco, M. (2008). Robust automated multiple view inspection. *Pattern Analysis and Applications*, *11*(1), 21-32.

Poole, D. L., y Mackworth, A. K. (2010). *Artificial intelligence: foundations of computational agents*. Cambridge University Press.

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Quinn, R., Sigl, C., y Company, E. K. (1980). *Radiography in modern industry* (Vol. 4). Eastman Kodak (Rochester, NY).

Riffo, V., y Mery, D. (2012). Active X-ray testing of complex objects. *Insight: Non-Destructive Testing & Condition Monitoring*, *54*(1), 28 - 35.

Riffo, V., y Mery, D. (2016). Automated detection of threat objects using adapted implicit shape model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *46*(4), 472-482.

Röntgen, W. C. (1895). Über nine neue art von strahlen (vorläufige mitteilung). *Sitzungsberichte der Würzburger Physikalisch und Medizinischen Gesellschaft*, 132.

Rowlands, J. (2002). The physics of computed radiography. *Physics in medicine and biology*, *47*(23), R123.

Schmidt-Hackenberg, L., Yousefi, M. R., y Breuel, T. M. (2012). Visual cortex inspired features for object detection in X-ray images. En *Pattern recognition (ICPR), 2012 21st international conference on* (pp. 2573–2576).

Schwaninger, A. (2003). *Detection systems: Screener evaluation and selection.* AIRPORT.

Schwaninger, A., Bolfing, A., Halbherr, T., Helman, S., Belyavin, A., y Hay, L. (2008). The impact of image based factors and training on threat detection performance in X-ray screening. En *Proceedings of the 3rd international conference on research in air transportation, ICRAT 2008* (pp. 317–324).

Schwaninger, A., Hardmeler, D., y Hofer, F. (2005). Aviation security screeners visual abilities visual knowledge measurement. *Aerospace and Electronic Systems Magazine, IEEE*, *20*(6), 29 - 35.

Shi, D.-H., Gang, T., Yang, S.-Y., y Yuan, Y. (2007). Research on segmentation and distribution features of small defects in precision weldments with complex structure. *NDT & E International*, *40*(5), 397 - 404.

Silva, R., y Mery, D. (2007a). State-of-the-art of weld seam inspection using X-ray testing: Part I – image processing. *Materials Evaluation*, *65*(6), 643-647.

Silva, R., y Mery, D. (2007b). State-of-the-art of weld seam inspection using X-ray testing: Part II – pattern recognition. *Materials Evaluation*, *65*(9), 833-838.

Singh, S., y Singh, M. (2003). Explosives detection systems (eds) for aviation security. *Signal Processing*, *83*(1), 31–55.

Sutton, R. S., y Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 28). MIT press.

Tang, Y., Zhang, X., Li, X., y Guan, X. (2009). Application of a new image segmentation method to detection of defects in castings. *The International Journal of Advanced Manufacturing Technology*, *43*(5-6), 431–439.

Turcsany, D., Mouton, A., y Breckon, T. (2013). Improving feature-based object recognition for X-ray baggage security screening using primed visualwords. En *Industrial technology (ICIT), 2013 IEEE international conference on* (p. 1140-1145).

Vedaldi, A., y Fulkerson, B. (2010). VLFeat: An open and portable library of computer vision algorithms. En *International ACM conference on multimedia (ICM)* (pp. 1469–1472).

Vilar, R., Zapata, J., y Ruiz, R. (2009). An automatic system of classification of weld defects in radiographic images. *NDT & E International*, *42*(5), 467–476.

von Bastian, C., Schwaninger, A., y Michel, S. (2010). *Do multi-view X-ray systems improve X-ray image interpretation in airport security screening* (Vol. 52). GRIN Verlag.

Wales, A., Anderson, C., Jones, K., Schwaninger, A., y Horne, J. (2009). Evaluating the two-component inspection model in a simplified luggage search task. *Behavior research methods*, *41*(3), 937.

Wang, Y., Sun, Y., Lv, P., y Wang, H. (2008). Detection of line weld defects based on multiple thresholds and support vector machine. *NDT & E International*, *41*(7), 517 - 524.

Watkins, C. J., y Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3-4), 279-292.

Watkins, C. J. C. H. (1989). *Learning from delayed rewards* (Tesis Doctoral no publicada). University of Cambridge England.

Wells, K., y Bradley, D. (2012). A review of X-ray explosives detection techniques for checked baggage. *Applied Radiation and Isotopes*.

Witkin, A. (1984). Scale-space filtering: A new approach to multi-scale description. , *9*,

150–153.

Zentai, G. (2008). X-ray imaging for homeland security. *Imaging Systems and Techniques, 2008. IST 2008. IEEE International Workshop on*, 1-6.

Zhigang, Z., Li, Z., y Jiayan, L. (2005). 3D measurements in cargo inspection with a gamma-ray linear pushbroom stereo system. En *Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05).*

Zhigang, Z., Yu-Chi, H., y Li, Z. (2010). Gamma/X-ray linear pushbroom stereo for 3D cargo inspection. *Machine Vision & Applications*, *21*(4), 413 - 425.

Zschornack, G. (2007). *Handbook of X-ray data*. Springer Verlag.

ANEXOS

ANEXO A. PUBLICACIONES REALIZADAS COMO PRIMER AUTOR

Riffo, V., y Mery, D. (2016). Automated detection of threat objects using adapted implicit shape model. *Systems, Man, and Cybernetics: Systems, IEEE Transactions on*, *46*(4), 472 - 482.

Riffo, V., y Mery, D. (2012). Active X-ray testing of complex objects. *Insight: Non- Destructive Testing & Condition Monitoring*, *54*(1), 28 - 35. (*The Ron Halmshaw Award; Best paper on radiography published in Insight in the 2012*).

Automated Detection of Threat Objects Using Adapted Implicit Shape Model

Vladimir Riffo and Domingo Mery

Abstract

Baggage inspection using X-ray screening is a priority task that reduces the risk of crime and terrorist attacks. Manual detection of threat items is tedious because very few bags actually contain threat items and the process requires a high degree of concentration. An automated solution would be a welcome development in this field. We propose a methodology for automatic detection of threat objects using single X-ray images. Our approach is an adaptation of a methodology originally created for recognizing objects in photographs based on Implicit Shape Models. Our detection method uses a visual vocabulary and an occurrence structure generated from a training dataset that contains representative X-ray images of the threat object to be detected. Our method can be applied to single views of grayscale X-ray images obtained using a single energy acquisition system. We tested the effectiveness of our method for the detection of three different threat objects: razor blades, *shuriken* (ninja stars) and handguns. The testing dataset for each threat object consisted of 200 X-ray images of bags. The true positive and false positive rates (TPR, FPR) are: (0.99, 0.02) for razor blades, (0.97, 0.06) for *shuriken* and (0.89, 0.18) for handguns. If other representative training datasets were utilized, we believe that our methodology could aid in the detection of other kinds of threat objects.

Active X-ray Testing of Complex Objects

Vladimir Riffo and Domingo Mery

Abstract

X-ray testing of complex objects –such as luggage screening at airports– is usually performed manually. This is not always effective, since it depends strongly on the pose of the objects of interest (target objects), occlusion and human capabilities as well. Additionally, certain target objects are difficult to be detected using only one viewpoint. For this reason, we developed an active X-ray testing framework that is able to adequate the viewpoint of the target object in order to obtain better X-ray images to analyze. The key idea of our method is to adapt automatically the viewpoint of the X-ray images in order to project the target object in poses where the detection performance should be higher. Thus, the detection inside of complex objects can be performed in a more effective way. Using a robotic arm and a semi-automatic manipulator system, the robustness and reliability of the method have been verified in the automated detection of razor blades located inside of nine different objects showing promising preliminary results: in 130 experiments we were able to detect 115 times the razor blade with 10 false alarms, achieving recall of 89 % and precision of 92 %.

ANEXO B. PUBLICACIONES REALIZADAS COMO CO-AUTOR

Mery, D., Riffo, V., Zscherpel, U., Mondragón G., Lillo, I., Zuccar, I., Lobel, H., y Carrasco, M. (2015). GDXray: The Database of X-ray Images for Nondestructive Testing, *Journal of Nondestructive Evaluation*, *34*(4), 1-12.

Mery, D., y Riffo, V., (2014). Automated Object Recognition Using Multiple X-ray Views, *Materials Evaluation*, *72*(11), 1362-1372.

Mery, D., Mondragón, G., Riffo, V., y Zuccar, I. (2013). Detection of regular objects in baggage using multiple X-ray views. *Insight: Non-Destructive Testing & Condition Monitoring*, *55*(1), 16-20. (*The John Grimwade Medal; Best paper award published in Insight in the 2013*).

Mery, D., Lillo, I., Riffo, V., Soto, A., Cipriano, A., y Aguilera, J. (2011). Automated fish bone detection using x-ray testing. *Journal of Food Engineering*, *2011*(105), 485-492.

GDXray: The Database of X-ray Images for Nondestructive Testing

Domingo Mery, Vladimir Riffo, Uwe Zscherpel, German Mondragón, Iván Lillo, Irene Zuccar, Hans Lobel, Miguel Carrasco

Abstract

In this paper, we present a new dataset consisting of 19,407 X-ray images. The images are organized in a public database called 𝔾𝔻𝕏ray that can be used free of charge, but for research and educational purposes only. The database includes five groups of X-ray images: castings, welds, baggage, natural objects and settings. Each group has several series, and each series several X-ray images. Most of the series are annotated or labeled. In such cases, the coordinates of the bounding boxes of the objects of interest or the labels of the images are available in standard text files. The size of 𝔾𝔻𝕏ray is 3.5 GB and it can be downloaded from our website. We believe that 𝔾𝔻𝕏ray represents a relevant contribution to the X-ray testing community. On the one hand, students, researchers and engineers can use these X-ray images to develop, test and evaluate image analysis and computer vision algorithms without purchasing expensive X-ray equipment. On the other hand, these images can be used as a benchmark in order to test and compare the performance of different approaches on the same data. Moreover, the database can be used in the training programs of human inspectors.

Automated Object Recognition Using Multiple X-ray Views

Domingo Mery and Vladimir Riffo

Abstract

In order to reduce the security risk of a commercial aircraft, passengers are not allowed to take certain items in their carry-on baggage. For this reason, human operators are trained to detect prohibited items using a manually controlled baggage screening process. The inspection process, however, is highly complex as hazardous items are very difficult to detect when placed in close packed bags, superimposed by other objects, and/or rotated showing an unrecognizable profile. In this paper, we review certain advances achieved by our research group in this field. Our methodology is based on multiple view analysis, because it can be a powerful tool for examining complex objects in cases in which uncertainty can lead to misinterpretation. In our approach, multiple views (taken from fixed points of view, or using an active vision approach in which the best views are automated selected) are analyzed in the detection of regular objects. In order to illustrate the effectiveness of the proposed method, experimental results on recognizing guns, razor blades, pins, clips and springs in baggage inspection are presented achieving around 90 % accuracy. We believe that it would be possible to design an automated aid in a target detection task using the proposed algorithm.

Detection of Regular Objects in Baggages Using Multiple X-ray Views

Domingo Mery, Germán Mondragón, Vladimir Riffo, Irene Zuccar

Abstract

In order to reduce the security risk of a commercial aircraft, passengers are not allowed to take certain items in carry-on baggage. For this reason, human operators are trained to detect prohibited items using a manually controlled baggage screening process. In this paper, we propose the use of a method based on multiple X-ray views to detect some regular prohibited items with very defined shapes and sizes. The method consists of two steps: 'structure estimation', to obtain a geometric model of the multiple views from the object to be inspected (a baggage), and 'parts detection', to detect the parts of interest (prohibited items). The geometric model is estimated using a structure from motion algorithm. The detection of the parts of interest is performed by an *ad–hoc* segmentation algorithm (object dependent) followed by a general tracking algorithm based on geometric and appearance constraints. In order to illustrate the effectiveness of the proposed method, experimental results on detecting regular objects –razor blades and guns– are shown yielding promising results.

Automated Fish Bone Detection Using X-ray Imaging

Domingo Mery, Iván Lillo, Hans Loebel, Vladimir Riffo, Alvaro Soto, Aldo Cipriano, José Miguel Aguilera

Abstract

In countries where fish is often consumed, fish bones are some of the most frequently ingested foreign bodies encountered in foods. In the production of fish fillets, fish bone detection is performed by human inspection using their sense of touch and vision which can lead to misclassification. Effective detection of fish bones in the quality control process would help avoid this problem. For this reason, an X-ray machine vision approach to automatically detect fish bones in fish fillets was developed. This paper describes our approach and the corresponding experiments with salmon and trout fillets. In the experiments, salmon X-ray images using 10×10 pixels detection windows and 24 intensity features (selected from 279 features) were analyzed. The methodology was validated using representative fish bones and trouts provided by a salmon industry and yielded a detection performance of 99 %. We believe that the proposed approach opens new possibilities in the field of automated visual inspection of salmon, trout and other similar fish.

ANEXO C. IMÁGENES DE PROPUESTA INICIAL DE *FRAMEWORK*, USANDO SISTEMA SEMI-AUTOMÁTICO DE SUJECIÓN Y ROTACIÓN

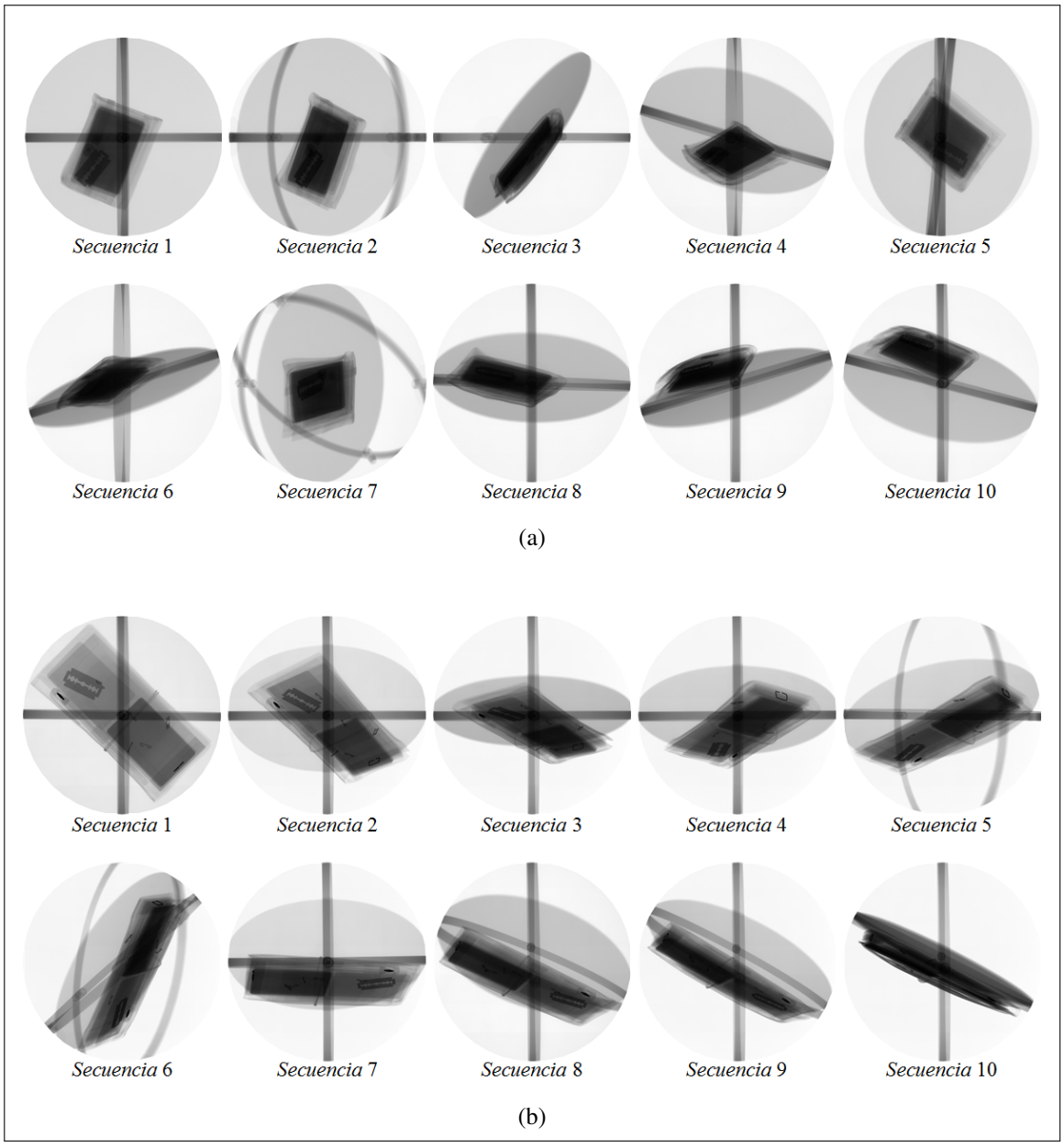


Figura C.1. Primera imagen adquirida para cada secuencia de inspección, usando sistema semi-automático de sujeción y rotación, a) Inspección de objeto Obj_1, y b) Inspección de objeto Obj_2.

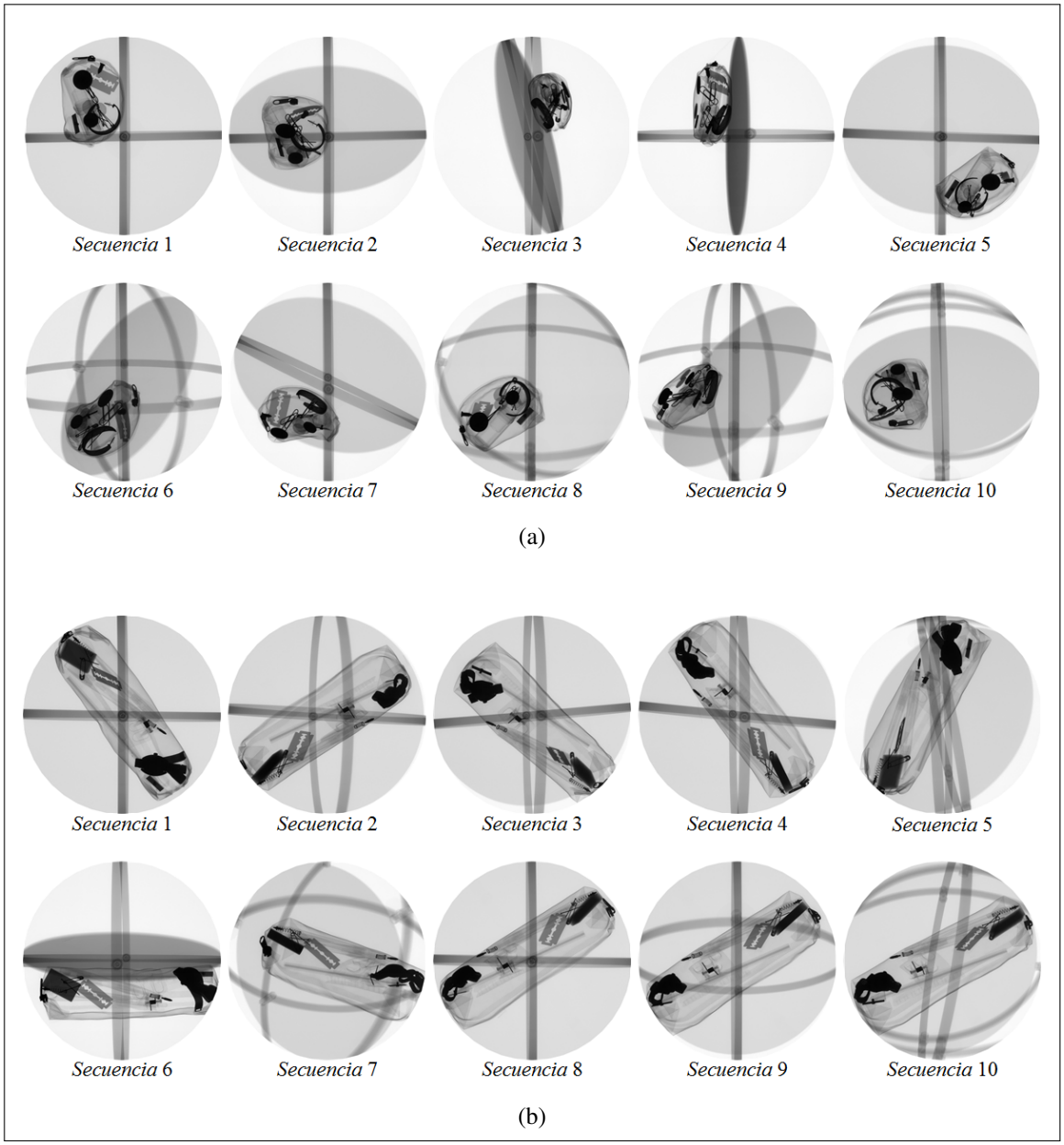


Figura C.2. Primera imagen adquirida para cada secuencia de inspección, usando sistema semi-automático de sujeción y rotación, a) Inspección de objeto Obj_3, y b) Inspección de objeto Obj_4.

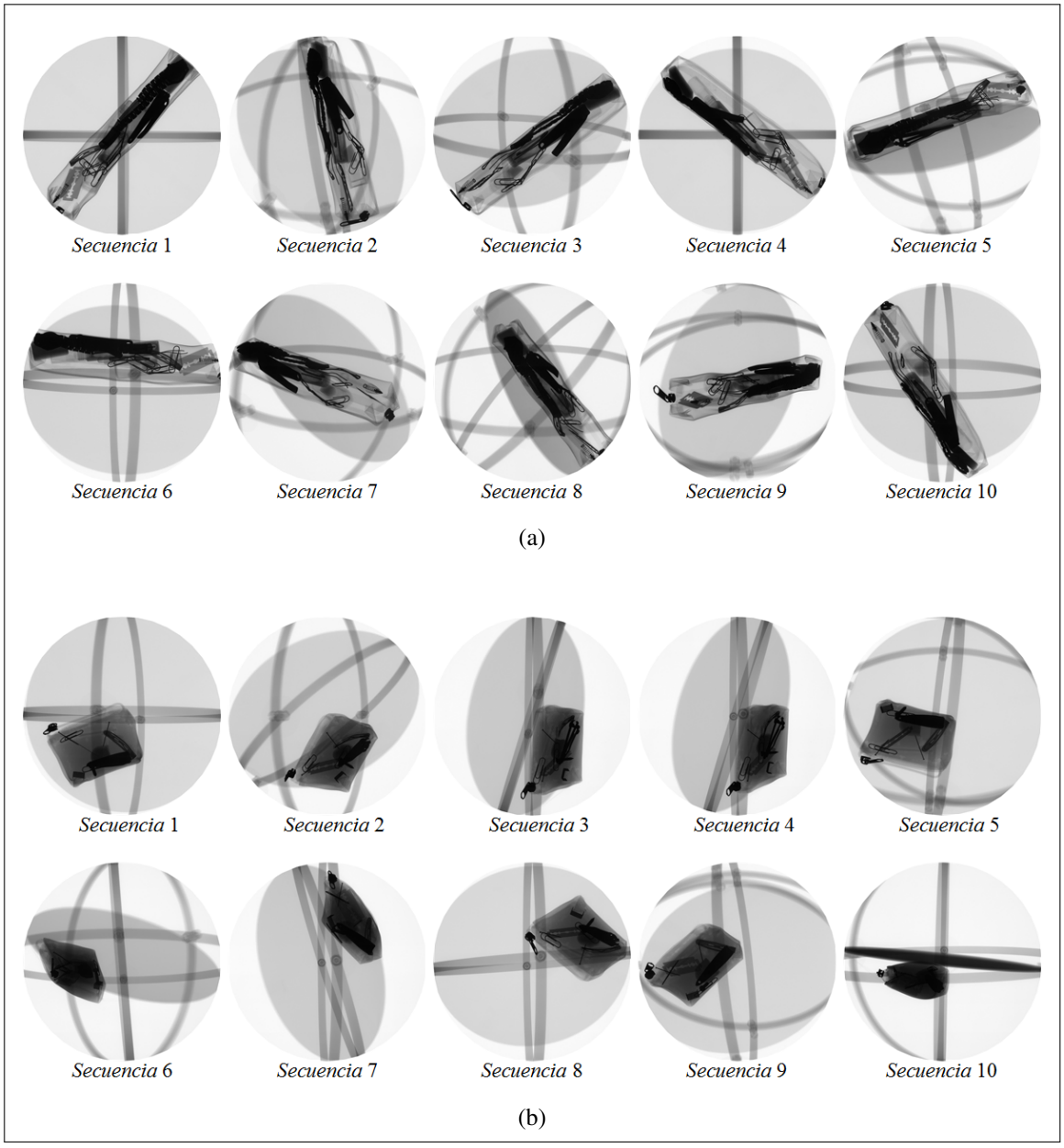


Figura C.3. Primera imagen adquirida para cada secuencia de inspección, usando sistema semi-automático de sujeción y rotación, a) Inspección de objeto Obj_5, y b) Inspección de objeto Obj_6.

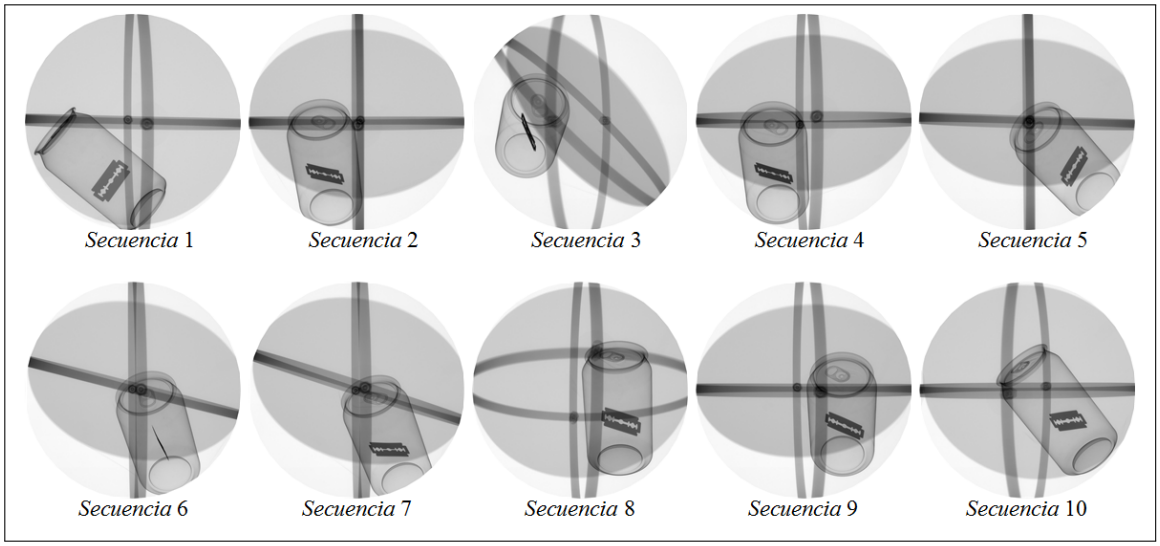


Figura C.4. Primera imagen adquirida para cada secuencia de inspección, usando sistema semi-automático de sujeción y rotación; Inspección de objeto Obj_7.

ANEXO D. IMÁGENES DE PROPUESTA INICIAL DE *FRAMEWORK*, USANDO SISTEMA AUTOMÁTICO DE SUJECIÓN Y ROTACIÓN

Figura D.1. Primera imagen adquirida para cada secuencia de inspección, usando sistema automático de sujeción y rotación, a) Inspección de objeto Obj_1, y b) Inspección de objeto Obj_3.

Figura D.2. Primera imagen adquirida para cada secuencia de inspección, usando sistema automático de sujeción y rotación, a) Inspección de objeto Obj_4, y b) Inspección de objeto Obj_7.

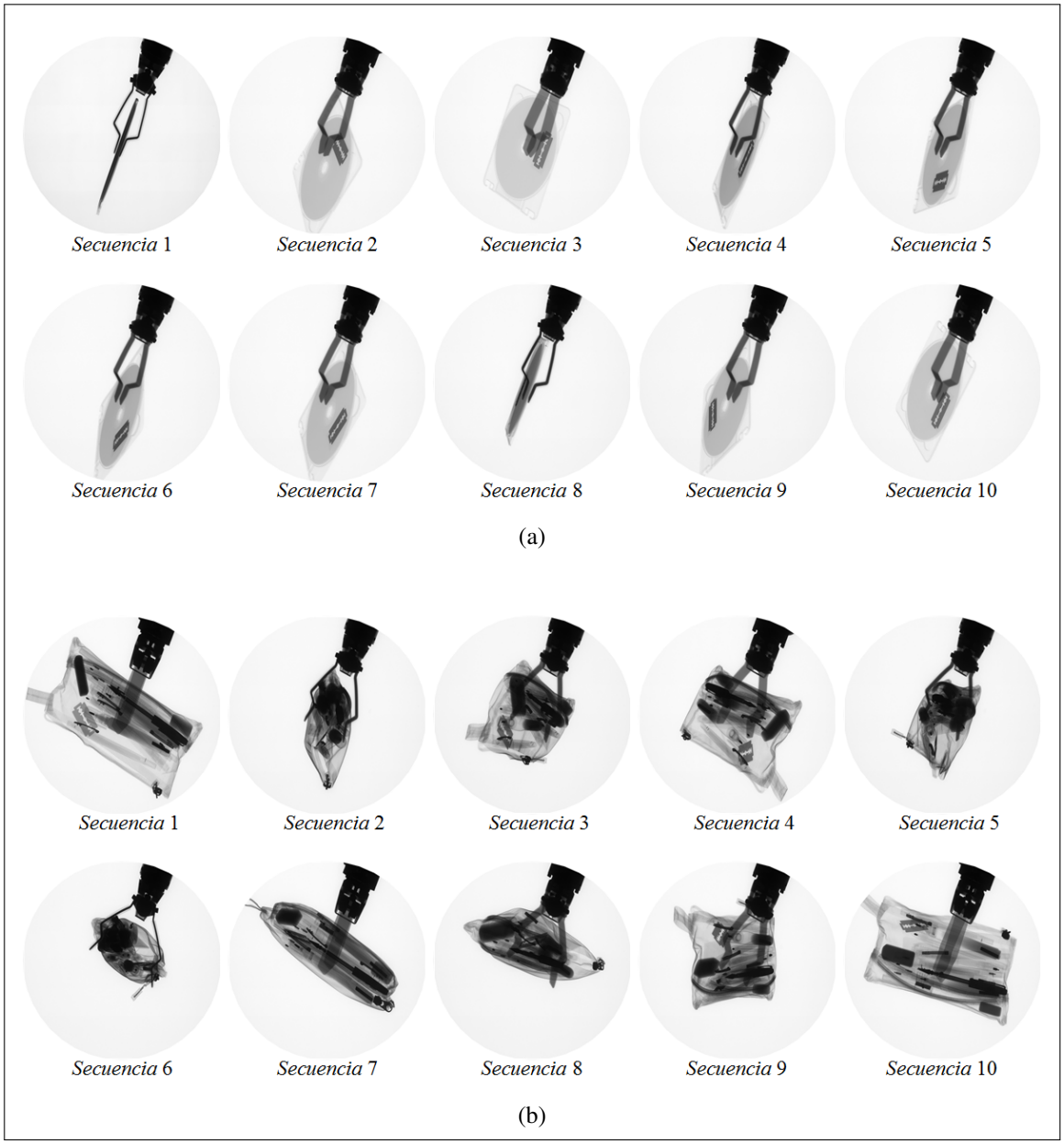


Figura D.3. Primera imagen adquirida para cada secuencia de inspección, usando sistema automático de sujeción y rotación, a) Inspección de objeto Obj_8, y b) Inspección de objeto Obj_9.