



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

METODOLOGÍA PARA EL DESARROLLO DE UN CLASIFICADOR DE MÚLTIPLES VISTAS DE RADIOGRAFÍAS:

*APLICACIÓN PARA LA BÚSQUEDA DE ARMAS PEQUEÑAS
DENTRO DE EQUIPAJES.*

GERMÁN MONDRAGÓN GONZÁLEZ

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

Profesor Supervisor:
DOMINGO MERY

Santiago de Chile, Julio 2012

© MMXII, GERMÁN MONDRAGÓN GONZÁLEZ



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

METODOLOGÍA PARA EL DESARROLLO DE UN CLASIFICADOR DE MÚLTIPLES VISTAS DE RADIOGRAFÍAS:

*APLICACIÓN PARA LA BÚSQUEDA DE ARMAS PEQUEÑAS
DENTRO DE EQUIPAJES.*

GERMÁN MONDRAGÓN GONZÁLEZ

Tesis presentada a la Comisión integrada por los profesores:

DOMINGO MERY

MIGUEL TORRES

MIGUEL CARRASCO

PABLO IRARRÁZVAL

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

Santiago de Chile, Julio 2012

*A mi familia: Lucia, Paloma,
Emiliano, Amaité.*

AGRADECIMIENTOS

Quería agradecer a quienes me han acompañado a lo largo de este desafío:

Agradecer a Patricia Castillo, mujer hermosa, por darme ánimo y compañía.

Agradecer a Lucia González, por darme aliento y soporte en todos mis desafíos.

Agradecer a Domingo Mery por su paciencia e inteligencia. Siempre aportó con ideas e iluminó el camino en momentos oscuros.

Agradecer a Gabriel Leiva, compañero de trabajos y de ideas.

Agradecer a Cristobal Arrieta, compañero destellante de ideas y proyectos.

Agradecer a mis amigos quienes siempre están.

Agradecer a Jorge Soto Huenún quien facilitó armas para realizar el estudio.

Por último agradecer a Fondecyt No 1100830: Automated X-ray testing of complex structures using efficient view search, view planning and active learning .

INDICE GENERAL

AGRADECIMIENTOS	IV
INDICE DE FIGURAS	VIII
INDICE DE TABLAS	X
RESUMEN	XI
ABSTRACT	XII
1. Introducción	1
1.1. Descripción del Problema	1
1.2. Hipótesis y Objetivos	3
1.3. Actualidad y Antecedentes Previos	3
1.4. Motivación	5
1.5. Alcances y Contribuciones	7
1.6. Solución Propuesta	8
1.7. Organización de la Tesis	10
2. Conceptos Básicos	12
2.1. Reconocimiento de Patrones	12
2.1.1. Discriminante Linear de Fisher <i>Fisher's linear discriminant</i>	14
2.1.2. Características	16
2.1.3. Selección de Características, <i>Sequential Forward Selection</i>	20
2.1.4. Evaluación del Clasificador	20
2.2. Geometría de las Múltiples Vistas	22
2.2.1. Geometría de dos vistas	22
2.2.2. Geometría de tres vistas	24
2.2.3. Auto-calibración con <i>Bundle Adjustment</i>	26
2.2.4. Método RANSAC para Estimar Puntos Inliers	28

3. Método Propuesto	29
3.1. Detector	30
3.1.1. Adquisición de Imágenes	30
3.1.2. Procesamiento de las Imágenes	31
3.1.3. Búsqueda en la Imagen <i>Sliding Windows</i>	34
3.1.4. Segmentación	35
3.1.5. Características	37
3.1.6. Clasificación	38
3.2. Estimación Geométrica	41
3.3. Clasificador en Múltiples Vistas	43
3.3.1. Correspondencia y Re proyecciones	44
3.3.2. Reducción de Múltiples Detecciones y Clasificación	47
4. Experimentos y Resultados	49
4.1. Secuencias	50
4.2. Segmentación	53
4.3. Selección de Características	55
4.4. Clasificador	57
4.5. Estimación de Matrices de Proyección	58
4.6. Reducción de puntos	59
4.7. Configuración de Clasificador en Múltiples Vistas.	59
4.8. Comparación de detecciones con diferentes niveles de dificultad.	61
4.8.1. Comparación con Método Alternativo de Tracking entre Vistas	63
4.8.2. Múltiples Detecciones	64
4.8.3. Mediciones de Tiempo por Etapas	66
5. Conclusiones	67
5.1. Trabajos Futuros	70
BIBLIOGRAFIA	71
Anexos	79

Clasificación de cada secuencia utilizando la configuración <i>LDA-LDA</i>	79
Detalle visual de las detecciones sobre las secuencias	91

INDICE DE FIGURAS

1.1. Descripción general del sistema	9
2.1. Clasificación de dos clases mediante Lineas Discriminative Analysis, LDA. . . .	14
2.2. Líneas epipolares	23
3.1. Diagrama de bloques del clasificador múltiples-vistas	29
3.2. Diseño del detector	30
3.3. Imagen de referencia para normalización	31
3.4. Histograma original, transformado y función de transformación LUT	34
3.5. Determinación de umbral mediante otsu	36
3.6. Etapas para el diseño del clasificador	38
3.7. Imágenes de pistolas, gatillos , no-gatillos	39
3.8. Comparación de clasificadores en fase de entrenamiento.	40
3.9. Estimación geométrica	42
3.10. Obtención puntos correspondientes	42
3.11. Clasificador en múltiples vistas	43
3.12. Diagrama de bloques para establecer correspondencia entre dos vistas.	44
3.13. Correspondencia entre dos imágenes	45
3.14. Puntos proyectados en todas las vistas	46
3.15. MeanShift para múltiples detecciones	47
3.16. Proyecciones del punto M_r	48
4.1. Modelos de pistolas	49
4.2. Categorías de dificultades en las secuencias	50

4.3. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia con dificultad baja	51
4.4. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia con dificultad media, imágenes con cables y chips	51
4.5. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia con dificultad alta, imágenes superpuestas sobre un laptop	52
4.6. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia de arma sola	52
4.7. Imagen generada para comprobación de la segmentación	54
4.8. Desempeño de clasificador v/s cantidad de características, con rojo la cantidad de características seleccionada	56
4.9. Muestra de gatillos	57
4.10. Comparaciones de cada configuración, con color azul el porcentaje de desempeño, con color naranja el porcentaje de falsos positivos	60
4.11. Detección en secuencia dificultada baja	62
4.12. Detección en secuencia dificultada media	63
4.13. Gatillo ocluido por una cámara	64
4.14. Gatillo ocluido con un laptop	65
4.15. Cantidad de puntos seleccionados por etapa.	65
4.16. Tiempos de ejecución por etapas.	66

INDICE DE TABLAS

4.1. Distribución de la muestra de secuencias por nivel de dificultad	53
4.2. Cantidad de identificaciones de gatillos: Métodos de segmentación y λ	55
4.3. Características seleccionadas	56
4.4. Matrices de confusión y performance	58
4.5. Desviación de puntos re-proyectados en pixeles	58
4.6. Clasificaciones en múltiples-vistas por nivel de dificultad, verdaderos positivos . .	61
4.7. Falsos positivos por nivel de dificultad	61
4.8. Detalle del desempeño método (Mery, 2011a)	63
5.1. Rendimiento y falsos positivos de las secuencias de fotografías de control	79
5.2. Rendimiento y falsos positivos de las secuencias con dificultad baja	80
5.3. Rendimiento y falsos positivos de las secuencias con dificultad media (artefactos electrónicos)	83
5.4. Rendimiento y falsos positivos de las secuencias con dificultad media	85
5.5. Rendimiento y falsos positivos de las secuencias dificultad alta (cámara)	86
5.6. Rendimiento y falsos positivos de las secuencias con dificultad alta (laptop) . . .	87
5.7. Rendimiento y falsos positivos de las secuencia con problemas de estimación modelo geométrico	90
5.8. Detecciones de las secuencias	91

RESUMEN

Actualmente los inspectores en los aeropuertos o aduanas logran detectar como máximo el 80 % -90 % de los objetos peligrosos. Para mejorar estos números, es necesario estudiar diferentes tecnologías no invasivas y rápidas, tales como imágenes de rayos X, que pueden optimizar el rendimiento de detección.

Esta investigación presenta un método para diseñar un clasificador de múltiples vistas que trabaje en situaciones no controladas, con altos niveles de ruido, bajos niveles de contraste y diferentes niveles de oclusión utilizando imágenes radioscópicas para la detección de pistolas dentro de bolsos. El método propuesto incorpora información geométrica e información cromática para el análisis en múltiples vistas y sus pasos son: 1) Búsqueda mono ocular de potenciales gatillos en todas las imágenes, 2) Estimación del espacio geométrico de la secuencia, 3) Seguimiento entre vistas y correspondencia de potenciales gatillos, 4) Eliminación de múltiples detecciones y falsos positivos.

Se validó el método con 27 secuencias de imágenes de rayos X y 3 secuencias de fotografías. Las secuencias contaron con movimientos libres en traslación y rotación sobre los ejes X , Y , Z entre las imágenes. Las imágenes contaron con diferentes niveles de oclusión debido a la superposición de objetos como cables, radios, monedas, cámara fotográfica, laptop, por lo que se dividieron en tres niveles de dificultad (baja, media, alta) simulando una situación real. El rendimiento fue de un 100 % para secuencias con dificultad baja, un 93 % para secuencias con dificultad media, un 73,4 % para secuencias con dificultad alta.

Este estudio demostró la capacidad de detección de múltiples vistas para ser implementado por la seguridad de los aeropuertos, sin embargo, con el fin de mejorar la validación se necesita una muestra mayor de armas.

Palabras Claves: múltiples-vistas, rayos X, detección pistolas

ABSTRACT

Actually airport inspectors and officials detect only a maximum of 80 % -90 % of dangerous objects inside personal bags. To improve this numbers, it is necessary the study of how different non-invasive and quick technologies such as X-ray imaging can be optimized to enhance this detection performance.

The aim of this study was to develop a methodology to design multi-view classifiers able to work in uncontrolled situations with high noise, low contrast and different occlusion levels using radiosopic images for the detection of guns inside the bags. Our method incorporated geometric and chromatic information for subsequently be analysed in multiple views including the follow steps: 1) Search monkey eye of potential triggers in all images, 2) Estimation of geometrical space sequence 3) Tracing between correspondence and views of potential triggers 4) Elimination of the false positives and multiple detections. The method was validated through 27 sequences X-ray images and 3 photograph sequences. The sequences counted with free movements in translation and rotation over the axes X, Y, Z between the images. The images had different levels of occlusion trough the overlapping objects such as cables, radios, coins, cameras, laptops and were divided in three difficulty levels (low, medium, high) simulating a real situation. The efficiency was 100 % for sequences with low difficulty, 93 % for sequences with moderate difficulty, 73.4 % for sequences with high difficulty.

This study demonstrated the capability of multi-view detection to be implemented in safety airport detection of bags, However, further improvements are necessary in order to improve validation with a larger sample of weapons.

Keywords: multi-view, x-ray, detection guns

1. INTRODUCCIÓN

El mundo globalizado se ha caracterizado, entre otras cosas, por desarrollar un constante aumento en el transporte de personas y productos. Esta situación ha ido incorporando la necesidad de realizar inspecciones más rápidas y menos invasivas. Por estas razones, se ha intensificado el uso de rayos X. El proceso de inspección con rayos X es dificultoso ya que los objetos de amenaza son difíciles de detectar, pues podrían estar superpuestos, rotados, o en una vista irreconocible (Zentai, 2008). La tasa de fallas y el tiempo involucrado por inspectores humanos está directamente relacionado con la dificultad del objeto a inspeccionar (Wales, Halbherr, y Schwaninger, 2009). Las técnicas de visión por computador han aportado a aumentar la eficacia en la inspección, brindando un apoyo a los inspectores humanos, incrementando el grado de automatización.

Esta tesis propone una metodología para buscar objetos utilizando imágenes de rayos X y múltiples vistas. Si bien existen métodos basados en el uso de imágenes de rayos X, en su mayoría utilizan una sola vista, afectando el rendimiento en la detección ya que las imágenes presentan un menor contraste y mayor ruido. Incorporando el uso de técnicas de visión por computador y la utilización de múltiples vistas se puede disminuir el ruido en la imágenes y los posibles falsos positivos, aumentando así el rendimiento.

La implementación de esta metodología en áreas relacionadas con seguridad, tendría como función asistir al personal encargado de las labores de inspección de objetos peligrosos. También podría ser un aporte para distintos rubros de la producción agrícola e industrial, con aplicaciones para la detección de objetos internos, productos dañados o defectuosos.

1.1. Descripción del Problema

Existen diversos tipos de objetos peligrosos posibles de ser transportados: armas de fuego, explosivos y armas blancas. Por lo general una pistola es pequeña y letal, de fácil transporte, fácil de ocultar y puede tener una presentación en innumerables modelos. Amnistía Internacional (2008) plantea que un 60 % de las violaciones de los derechos humanos son perpetrados con el uso de armas pequeñas. En promedio un 73 % del total de las armas de fuego

pequeñas están en manos de civiles (Berman, 2011). La declaración de Ginebra (2011) plantea que un 75 % de las muertes son homicidios intencionales sin conflicto bélico y un 10.4 % son consecuencia de conflictos bélicos. Existe una especial preocupación por los Estados de Francia y Bélgica en normar y aumentar los controles en la aviación para evitar el transporte de armas. Si bien todos los Estados adoptan leyes que rigen las exportaciones de armas, leyes de aduanas y reglamentos de aviación civil, sigue siendo necesario aumentar el control tanto en el punto de partida como en el de llegada. También se requieren mayores niveles de coordinación entre los países, situación que no ha quedado completamente resuelta en los acuerdos de Wassenaar Arrangement (2012).

No hay duda que las armas pequeñas son peligrosas, pues las posibilidades de quedar herido de muerte son muy altas si es que el amenazado no se somete a la voluntad de quien la posea. Es por ello que los Estados se preocupan por controlar su posesión y transporte, sin embargo, la cantidad de armas de fuego con uso civil es de alrededor 650,000,000 en el mundo (Berman, 2011), lo que plantea una dificultad real para su control.

Actualmente la responsabilidad de la revisión de bolsos, con el fin de detectar objetos peligrosos, es de los inspectores que trabajan en las aduanas. El escaso tiempo que tienen para revisar un gran volumen de bolsos con una importante cantidad de objetos diversos en su interior hacen que los inspectores humanos logren detectar como máximo el 80 % -90 % de los objetos peligrosos (Michel y cols., 2007a). Para aumentar la cantidad de detecciones se necesita ayudar a los inspectores humanos en la toma de decisiones acerca de la posible presencia de objetos peligrosos. Con este objetivo las técnicas de visión por computador han dado un buen soporte a la inspección asistida mediante la utilización de imágenes de rayos X. Sin embargo, las imágenes de rayos X presentan una serie de dificultades dado el nivel de oclusiones producidas por objetos superpuestos, es por ello que el análisis de múltiples vistas da mayores opciones de reconocimiento a partir de distintos puntos de vistas.

1.2. Hipótesis y Objetivos

La hipótesis central de la investigación es que mediante la utilización de un clasificador de múltiples-vistas en imágenes de rayos X, es posible detectar pistolas al interior de un bolso, con y sin oclusión parcial.

El objetivo principal de esta investigación es diseñar una metodología para el desarrollo de un clasificador basado en múltiples vistas, que opere en un ambiente no calibrado y que logre la detección de fallas u objetos al interior de otros objetos. Para esta investigación el objetivo se acotó a la detección de gatillos de pistolas, al interior de bolsos.

Como objetivos específicos presentamos además:

- Diseñar un detector de pistolas basado en un clasificador de gatillos mono ocular.
- Implementar un sistema de calibración automática para imágenes de rayos X.
- Diseñar un clasificador de múltiples vistas que considere información tanto geométrica como información de intensidad.

Para enfrentar el problema como si fuera un caso real se evaluará el clasificador múltiples-vistas para 3 niveles de dificultad:

- ▷ La detección y clasificación de objetos internos sin oclusión.
- ▷ La detección y clasificación de objetos internos con poca oclusión.
- ▷ La detección y clasificación de objetos internos con alta oclusión y ruido.

1.3. Actualidad y Antecedentes Previos

En la década de los setenta y los ochenta la mayor amenaza para la aviación civil eran los secuestros de aviones, lo que implicó que la industria de seguridad se enfocara en la detección de objetos metálicos que portasen los pasajeros. En la década del 80, la explosión del vuelo 103 de Pan Am terminó por imponer la necesidad de incorporar la amenaza de los explosivos en la industria de la aviación. Los diferentes Estados comenzaron profundizar las regulaciones de los mecanismos en prevención del transporte. Así es como se fuerza un debate acerca de los métodos más fiables en la detección de explosivos. Hasta la fecha, el método aceptado por

las FAA (Federal Aviation Authority) trabajaba en base a la detección de hidrógeno con una técnica que utiliza Thermal Neutron Activation (TNA) (Murphy, 1989), sin embargo, debido a su costo, a la alta tasa de falsos positivos, como al tiempo requerido para la detección, no es aceptado como método de detección en la industria de aviación civil.

En los primeros análisis de equipaje con imágenes de rayos X los inspectores jugaban un rol fundamental ya que para la interpretación de las imágenes no existía ayuda. Al inicio, los trabajos que aportaron información estaban orientados a la detección de explosivos y su desarrollo se orientó en determinar el material del equipaje (Murray y Riordan, 1995). Esta técnica consiste en el cálculo del número atómico (Z_{eff}) del material mediante la utilización de energía dual (Murphy, 1989). Con dos ondas de energía distinta se mide la diferencia de absorción para aproximar el Z_{eff} . Con el número atómico se puede determinar si los materiales son orgánicos, inorgánicos, metálicos. Pueden identificarse al menos tres defectos en los experimentos que utilizan energía dual, el primero, es que este método detecta el número atómico integrado del objeto sin considerar información del volumen, el segundo defecto, es que cuando existen materiales superpuestos el cálculo del Z_{eff} dependerá de la integración de todos los Z_{eff} de los objetos superpuestos, el tercero, es que el Z_{eff} de los explosivos plásticos es similar a los plásticos en general, confundiendo objetos inocuos con explosivos. Las técnicas que operan con rayos de dispersión aportan información de la densidad (Strecker, 1998; Hussein E.M.A. Waller, 1998).

Evans, Robinson, y Godber (1998) proponen un método que logra extraer información 3D de una serie de imágenes estereoscópicas de rayos X. Su método busca puntos correspondientes y va realizando interpolación entre la información adquirida con cada par de imágenes 2D, mientras el objeto se mueve por una cinta transportadora. La distancia entre las imágenes 2D es conocida como Z_{min} , con lo cual logra obtener información por capas. Esta metodología requiere de un hardware específico que opere con un doble sensor para la captura de imágenes estéreos, lo cual lo hace bastante costoso.

La utilización de imágenes de rayos X se muestra como mecanismo eficiente y no invasivo de asistir al inspector por lo que las investigaciones se profundizan en esta línea. Debido a las

dificultades en el análisis de las imágenes de rayos X producto del bajo contraste, las oclusiones y del ruido, es que la mayoría de los métodos trabajan fuertemente la segmentación o detección de bordes (M. Singh y Singh, 2005; Nercessian, Panetta, y Agaian, 2008; Zentai, 2008).

Una línea de investigación productiva es la que sigue combinando información obtenida con energía dual e información 3D, un trabajo reciente es el de Chan, Evans, y Wang (2010) quienes desarrollan técnicas de segmentación por tipo de material. Otra línea de investigación se enfoca en buscar mejores técnicas de clasificación como el método para la detección de explosivos con un clasificador difuso de nivel múltiple Liu y Wang (2008). Las investigaciones que trabajan con una sola vista orientan su preocupación en la segmentación como en el clasificador, por ejemplo, el método propuesto para la detección de una pistola Colt Beretta (Oertel y Bock, 2006) que trabaja con segmentación y con características de textura, sin considerar casos de oclusión. Sin embargo, las líneas de investigación más interesantes tienen un énfasis en la identificación del tipo de material (Z_{eff}), aún cuando no se considera la oclusión ni la información espacial del objeto. En este sentido resulta prometedora la línea de investigación propuesta en (Mery, 2011a) donde la combinación de información cromática y geométrica es utilizada para realizar detecciones automáticas. Lo interesante de esta detección automática es que se orienta al desarrollo de un modelo de detección con múltiples vistas sin acotar la problemática a una detección específica, de esta forma el modelo se puede aplicar a múltiples problemas de detección en cualquier rama industrial, no solo la seguridad. Además, este modelo al trabajar con múltiples vistas, analiza la escena desde diferentes ángulos por lo que se puede evitar posibles oclusiones en ciertas posiciones, o buscar mediante visión activa la mejor vista (Riffo y Mery, 2012). Su aplicación en el área de seguridad podría ser de gran interés ya que la tarea de detectar objetos peligrosos, como detectar objetos ocluidos por otros objetos al interior de un equipaje es un tema aún no resuelto (Zentai, 2008).

1.4. Motivación

El principal interés de esta investigación es aportar al estudio y reflexión para el desarrollo de metodologías en el diseño de clasificadores con múltiples-vistas que trabajen en situaciones no controladas, con altos niveles de ruido, bajos niveles de contraste y con diferentes niveles

de oclusión. Profundizar estas metodologías es de interés para el desarrollo de detectores cada vez más precisos que aumenten los niveles de percepción para la automatización en áreas industriales tanto de la robótica, como de la seguridad.

Los sistemas de detección en el área de transporte hoy en día presentan un énfasis que está orientado a la detección de explosivos. Sin embargo, la gran cantidad de armas pequeñas en manos de civiles y la cantidad de muertes producidas a partir de ellas, nos presentan una real dimensión de su peligrosidad y de su necesario control por parte de los Estados en su transporte.

El volumen y la velocidad de las inspecciones hechas por los inspectores humanos diariamente producen altos niveles de estrés, por lo que es necesario aumentar la asistencia para la inspección fortaleciendo la prevención en el transporte de objetos peligroso.

La investigación para detectar objetos peligrosos ha sido impulsada de los Estados, de las empresas relacionadas con el transporte, y a la comunidad científica mundial. A pesar de los avances que se han logrado realizar, las aplicaciones comerciales no han sido capaces de dejar el problema resuelto. El mayor énfasis sigue puesto en la detección de explosivos, por lo que la mayor cantidad de enfoques se orientan en la clasificación del tipo de material.

El problema de la seguridad en la sociedad civil no solo se ve amenazado por posibles ataques, la cantidad de muertes producidas por armas pequeñas, así como la cantidad de ellas en manos de civiles, deja a la armas pequeñas como la mayor amenaza mundial. Tendría que existir una mayor preocupación por su control por parte de los Estados y mayor capacidad para su detección desde la industria de seguridad y del transporte.

Los inspectores humanos no logran detectar el total de objetos peligrosos (Zentai, 2008) transportados. Las jornadas de trabajo son extensas con remuneraciones bajas. Esta situación plantea que existe una debilidad latente en los puntos de control del transporte. Si bien es cierto que con mayor entrenamiento y mayor experiencia las detecciones aumentan (Michel y cols., 2007b), la cantidad de controles que se tienen que realizar en pocos segundos aumentan a diario.

La propuesta metodológica desarrollada en la tesis es un modelo que utiliza múltiples vistas el cual podría aplicarse a cualquier área industrial, sin embargo, si se aplica al área de seguridad del transporte podría ser de gran utilidad como apoyo a los inspectores en las aduanas.

1.5. Alcances y Contribuciones

Esta investigación propone una metodología para el diseño de un clasificador múltiples vistas que opere con imágenes de rayos X para realizar detecciones de objetos como de anomalías internas.

Se diseñó e implementó un clasificador múltiples-vistas para la detección de armas pequeñas al interior de equipajes.

Se realizó una validación del método propuesto para detección de armas pequeñas utilizando imágenes de rayos X con diferentes niveles de oclusión y ruido sobre las imágenes. Se dividió la muestra en tres niveles de oclusión: sin oclusión, con oclusión parcial, con altos niveles de oclusión. Se logró obtener un desempeño de un 100 % de detecciones en secuencias sin oclusión. 98,3 % con oclusión parcial y un 73,4 % fuertemente ocluido.

Se realizó una validación del método de clasificación propuesto con múltiples vistas en general. Esta validación se realizó mediante una clasificación de armas con secuencias de fotografías, presentando un 100 % de detecciones.

El resultado de esta investigación aportará al desarrollo de metodologías para el diseño de clasificadores múltiples vistas que operen en la detección de objetos interiores con oclusión parcial y ruido.

Esta tecnología se puede incorporar positivamente a distintas áreas de la robótica, de la industria productiva y del campo de la seguridad.

Este trabajo aportará con el conjunto de imágenes de pruebas, las cuales son imágenes de rayos X de equipajes con armas pequeñas con diferente nivel de oclusión. Estas imágenes

quedaran a disposición de futuras investigaciones por parte de la comunidad científica de la Pontificia Universidad Católica de Chile.

1.6. Solución Propuesta

Esta investigación profundizó y mejoró el método propuesto por Mery (2011a) que consiste en la detección de una falla u objeto en una vista, su seguimiento en otras vistas y la posterior eliminación de falsos positivos. La propuesta del método plantea que una sola vista no proporciona suficiente información para la detección del objetivo buscado, por lo que mediante geometría epipolar se incorpora información geométrica aumentando el desempeño de la detección.

El clasificador de múltiples-vistas opera con información combinada de geometría epipolar (Richard Hartley, 2003) e información de intensidad, lo cual disminuye falsos positivos proporcionando niveles mayores de certeza en la clasificación (Carrasco y Mery, 2007). Si en cada vista se logra tener una clasificación parcial de los objetos que están presentes, mediante el análisis de múltiples vistas se logrará una identificación global de pertenencia del objeto en la escena o secuencia.

Se propone una metodología que capaz de detectar un posible objeto peligroso en múltiples imágenes de rayos X. Se clasifica en una sola vista para realizar una detección, se realiza un seguimiento de las detecciones en todas las vistas, por último, se aplican criterios de decisión considerando información geométrica y descriptores SIFT para detectar la presencia de un objeto peligroso en la escena. El diseño de un detector de pistolas requiere que sea lo suficientemente general como para detectar diferentes modelos, pero además lo suficientemente preciso para descartar falsos positivos. Hay que considerar que existen diferencias en el tamaño, el cañón, los resortes, los martillos y los mangos de las pistolas casi por cada modelo. Por lo tanto, no sería sensato confeccionar un detector que de cuenta de todas estas diferencias. En esta investigación se definió buscar el gatillo por ser este un elemento presente y similar que existe en todos los modelos de pistolas. Es decir, se acotó la búsqueda de pistolas a una búsqueda de gatillos, por lo que en esta tesis se puso a prueba la metodología para la detección

de armas pequeñas mediante una búsqueda exhaustiva de gatillos en el interior de diversos tipos de equipajes utilizando múltiples vistas de imágenes de rayos X.

El método propuesto en la presente tesis cuenta con varias etapas las que se presentan en la Figura 1.1. *Obtención de Imágenes* Se obtuvieron secuencias de imágenes en una ambien-

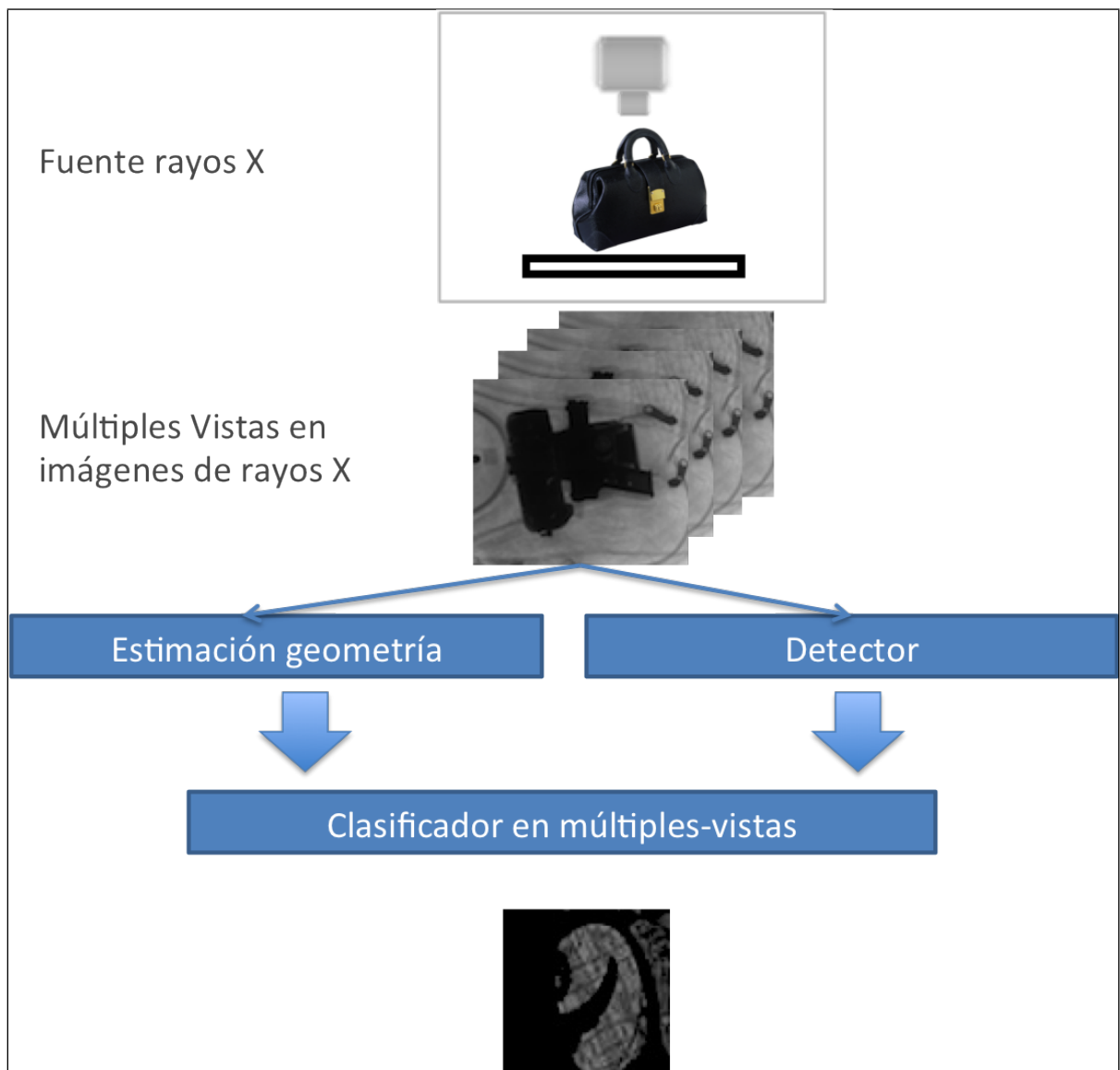


FIGURA 1.1. Descripción general del sistema

te poco controlado. Las secuencias se construyeron a partir de imágenes adquiridas donde se

utilizaron 6 diferentes tipos de armas pequeñas, pistolas automáticas, colocadas aleatoriamente dentro de un equipaje, luego el equipaje se fue moviendo aleatoriamente para adquirir el resto de las imágenes. *Pre-procesamiento de Imágenes* Se pre-procesaron las imágenes para aumentar el contraste y disminuir el ruido. *Estimación Geométrica* A partir de cuatro imágenes o más pertenecientes a una secuencia, se extrajeron puntos característicos y correspondientes para calcular las matrices de proyección de cada vista, con esto se logra establecer el modelo geométrico en que está el objeto representado por las vistas. *Entrenamiento Detector* Se diseñó un clasificador mono ocular específico ad-hoc para la detección de gatillos. Fue entrenado con un conjunto de imágenes de pistolas adquiridas desde internet logrando establecer el método de clasificación como las características representativas del gatillos. *Clasificador mono ocular* Una vez que se tienen las imágenes, se busca en ellas los posibles gatillos con el clasificador mono ocular. Para esto se obtiene una ventana que va recorriendo de manera exhaustiva toda la imagen y va evaluando el clasificador. *Proyecciones y validación* Se realizaron proyecciones en todas las vistas de los posibles gatillos detectados. De esta forma, combinando información geométrica y descriptores SIFT, se van proyectando y re-clasificando para descartar falsos positivos.

1.7. Organización de la Tesis

La tesis consta de los siguientes capítulos:

1. Introducción: en este capítulo se verán aspectos relacionados con la descripción del problema, la motivación, la hipótesis, los objetivos, los antecedentes previos y cómo en la actualidad la comunidad científica está abordando la problemática.
2. Conceptos Básicos: La finalidad de este capítulo es brindarle al lector un conjunto de antecedentes para que éste pueda conocer tanto la metodología propuesta como los fundamentos teóricos que le dieron sustento.
3. Método Propuesto: en este capítulo se revisaran los aspectos relacionados con la propuesta específica que se desarrolla en la tesis para el diseño de una metodología de un clasificador en múltiples vistas.

4. Experimentos y Resultados: en este capítulo se expondrán los experimentos realizados en el transcurso de la investigación con los resultados obtenidos.
5. Conclusiones: en este capítulo se expondrán los aspectos relevantes que se concluyen además de los posibles trabajos futuros.
6. Anexo: La finalidad de este capítulo es proporcionar al lector mayores antecedentes gráficos mediante el despliegue de Tablas con rendimientos e imágenes de rayos X de las secuencias.

2. CONCEPTOS BÁSICOS

En este capítulo se expone el marco conceptual que le da soporte teórico a las técnicas utilizadas. Los temas abordados son diversos, en general han sido abordados en diferentes áreas de visión por computador, reconocimiento de patrones y procesamiento de imágenes. La finalidad de este capítulo es brindarle al lector un conjunto de antecedentes para que éste pueda conocer tanto la metodología propuesta como los fundamentos teóricos que le dieron sustento.

El capítulo se divide en dos grandes secciones. El primero tiene relación con el reconocimiento de patrones, esto es una explicación de las técnicas utilizadas para el diseño de un detector mono-ocular, especificación del clasificador utilizado, extracción y selección de características del objeto, disminución de las dimensiones de las características, por último, evaluación de clasificadores. La segunda sección tiene relación con el modelo geométrico que representan las imágenes, calibración, geometría epipolar y re-proyección de puntos en 3D.

Ambos ámbitos teóricos son necesarios para comprender cómo se diseñó la propuesta de la metodología para el clasificador múltiples-vistas.

2.1. Reconocimiento de Patrones

El reconocimiento de patrones es el proceso de clasificación (etiquetado) de uno o más objetos, que previamente han sido conocidos. El objetivo es realizar una interpretación de los datos extraídos de los objetos para lograr una representación precisa de ellos. De esta manera el reconocimiento de patrones es esencial para una interpretación automática de objetos o fenómenos. Existen dos formas de diseñar un clasificador para el reconocimiento de patrones:

- Supervisado: donde las características han sido previamente definidas para cada clase y la clasificación se realiza después de un entrenamiento.
- No supervisado: donde se desconocen las clases a la que pertenecen las características.

El reconocimiento de patrones consta de las siguientes etapas: 1) la extracción de información, o características, 2) su representación, 3) y la toma de decisión. Para ello, se busca

establecer las propiedades o relaciones, conservando solo la menor información discriminatoria. Un patrón, es una entidad, una característica, que pudiera ser identificada, es uno o más atributos que logran identificar el objeto.

Existen diferentes modelos de reconocimiento de patrones (Duin y Mao, 2000; Bishop, 2006): el de correspondencia, el estadístico, el sintáctico y de redes neuronales. El reconocimiento de patrones estadísticos ha sido utilizado con éxito en aplicaciones comerciales como en una diversidad de investigaciones (Viola y Jones, 2001).

En este estudio se utilizó un clasificador estadístico. Estos requieren de la representación de cada patrón como una medición característica que se considera como un punto en el espacio multi-dimensional. Los patrones característicos de un mismo objeto tendrían que ser los mismos, sin embargo, tienen una variabilidad debido a ruido, deformaciones, diferencias en la captura, etc. por lo que el patrón puede ser descrito mediante un modelo probabilístico como un vector x donde $x = [x_1, x_2, \dots, x_n]$ de n -dimensiones perteneciente a m posibles clases ω_i donde $i \in \{1, \dots, m\}$. Cada clase ω_i tiene una probabilidad *a priori* igual a $P(\omega_i)$.

$$\sum_{i=1}^m P(\omega_i) = 1 \quad (2.1)$$

Lo que interesa es calcular es la probabilidad una vez que ha sido observado el patrón; *a posteriori* y está definida por la fórmula de Bayes que relaciona probabilidades condicionales:

$$P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{P(x)} \quad (2.2)$$

La clasificación es la decisión que toma el sistema, esto no es más que la asignación de una categoría a los patrones de una clase desconocida *a priori*. Es buscar el conjunto de características o mediciones ω_i que se encuentran más compactas y que logren separar la región que representa a cada objeto. El clasificador estará definido con el mejor criterio que logre separar las regiones. Para la búsqueda del criterio se analizan discriminantes (por ejemplo lineales o cuadráticos) que logren establecer las fronteras entre las regiones de características que representan a las clases.

Los clasificadores estadísticos requieren de supervisión, esto significa que en una etapa inicial se entrena el clasificador con un conjunto de entrenamiento del cual se conocen las clases y sus características. De esta manera se analizan las características y se determinan los discriminantes que logren separar mejor la muestra.

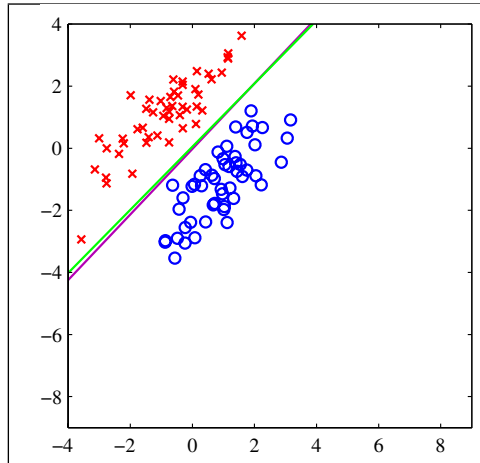


FIGURA 2.1. Clasificación de dos clases mediante Lineas Discriminative Analysis, LDA.

En la Figura 3.1.6 se muestra la gráfica de las características de dos clases, con cruces rojas, con círculos azules la otra. También se muestra el discriminador lineal que logra segregar de mejor forma las clases.

2.1.1. Discriminante Linear de Fisher *Fisher's linear discriminant*

Si se considera que tenemos dos clases, y tenemos un vector x de n -dimensiones, la proyección en una sola dimensión sería:

$$y = w^T x \quad (2.3)$$

Si las características no estuviesen ocluidas entre las dimensiones se podría establecer un umbral para la separación de las clases $y > \omega_0$, sin embargo, las características regularmente se encuentran solapadas entre las dimensiones.

Considerar que tenemos w_1 puntos de la clase C_1 y w_2 puntos de la clase C_2 . Los vectores de las medias serán:

$$m_1 = \frac{1}{w_1} \sum_{n \in C_1} x_n, m_2 = \frac{1}{w_2} \sum_{n \in C_2} x_n \quad (2.4)$$

La separación de las clases se podría realizar a partir de la proyección de las medias en w y buscar una función de maximización de w

$$m_2 - m_1 = w^T (m_2 - m_1) \quad (2.5)$$

Utilizando el multiplicador de Lagrange se puede maximizar tal que $w^T (m_2 - m_1)$. Se define el total de la varianza para los datos como $(s_1^2 + s_2^2)$.

$$J(w) = \frac{(m_2 - m_1)^2}{(s_1^2 + s_2^2)} \quad (2.6)$$

El criterio de Fisher (Welling, 2009) se determina a partir de la matrices de covarianza y está definido como la menor distancia intra-clase y la mayor distancia inter-clases (Webb, 2005).

Entonces la definición de Fisher queda:

$$J(w) = \frac{w^T S_B w}{w^T S_W w} \quad (2.7)$$

donde

$$S_B = (m_2 - m_1)(m_2 - m_1)^T \quad (2.8)$$

$$S_W = \sum_{n \in C_1} (\mathbf{x}_n - m_1) - (\mathbf{x}_n - m_1)^T - \sum_{n \in C_2} (\mathbf{x}_n - m_2) - (\mathbf{x}_n - m_2)^T \quad (2.9)$$

En realidad el criterio de Fisher no alcanza a ser un discriminante pero si se elige un umbral y_0 podemos establecerlo como tal ya que se podría clasificar un punto como clase C_1 si $y(\mathbf{x}) > y_0$ o C_2 en otro caso.

2.1.2. Características

Las características son uno o más atributos que logran identificar al objeto. La importancia de encontrar las mejores características que representen al objeto está en directa relación con el desempeño del clasificador. Existen dos grandes grupos de características extraídas en reconocimiento de patrones de imágenes, las cuales son características geométricas y cromáticas (Mery, 2006b). Las características geométricas tienen relación con la forma, posición, tamaño, orientación de la imagen, etc. Las características cromáticas tienen relación con la intensidad, con los colores, las texturas, etc.

Las características utilizadas en el diseño del detector mono-ocular fueron características geométricas por lo que se verán algunas definiciones de ellas.

Las Características Geométricas

Se expondrán algunas definiciones de características geométricas utilizadas en la presente investigación. El comando de `regionprops` de matlab proporciona varias características geométricas estándar del toolbox (MathWorks, 2010).

Redondez: Esta característica contiene información acerca de cuan redonda es la región, está definida por:

$$R = \frac{4A\pi}{L^2} \quad (2.10)$$

Donde A es el área y L es el perímetro. La Redondez también llamada compacidad R es para un círculo perfecto $R = 1$ y para una línea $R = 0$

Radios: R_{max} radio máximo y R_{min} radio mínimo, son las distancias máximas y mínimas que tiene el contorno a su centro de masa.

Altura y Anchura: Es la distancia de la altura y de la anchura de la región:

$$h = i_{max} - i_{min} + 1 \quad \text{y} \quad w = j_{max} - j_{min} + 1 \quad (2.11)$$

Donde i_{max} e i_{min} son el valor máximo y mínimo en la coordenada i , y j_{max} y j_{min} para la coordenada j .

Momentos Estadísticos: Los momentos estadísticos son :

$$m_{rs} = \sum_{ij \in \mathfrak{R}} i^r j^s \quad \text{para } r, s \in \mathcal{N} \quad (2.12)$$

Donde $\mathfrak{R}$ es el conjunto de pixeles de la región.

Con las coordenadas del centro de gravedad se definen los momentos invariantes:

$$\nu_{rs} = \sum_{ij \in \mathfrak{R}} (i - \bar{i})^r (j - \bar{j})^s \quad \text{para } r, s \in \mathcal{N} \quad (2.13)$$

Los momentos de Hu (Hu, 1962) son las características derivadas de los momentos centrales:

$$\begin{aligned}
\phi_1 &= \mu_{20} + \mu_{02} \\
\phi_2 &= (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2 \\
\phi_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\
\phi_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\
\phi_5 &= (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} - \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + \\
&\quad (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\
\phi_6 &= (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + \\
&\quad 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\
\phi_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - \\
&\quad (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[(3\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]
\end{aligned} \tag{2.14}$$

con

$$\mu_{rs} = \frac{\nu_{rs}}{\nu_{00}^t} \quad t = \frac{r+s}{2} + 1$$

Los Descriptores de Fourier: Los descriptores de Fourier (Persoon y Fu, 1977) caracterizan la forma. Los descriptores de Fourier pueden representar a una curva cerrada, si tenemos un contorno $x(n)$ constituido por N puntos en el plano XY cada punto puede ser considerado como un punto complejo:

$$x(n) = x_n + jy_n \quad \text{con } j = \sqrt{-1} \tag{2.15}$$

Si seguimos el contorno y tomamos una secuencia de puntos en la misma dirección la transformada discreta de Fourier queda:

$$X(n) = \sum_{k=0}^{L-1} (x(n)) e^{(-j \frac{2\pi kn}{L})} \quad \text{para } n = 0, \dots, L-1 \quad (2.16)$$

La fase de los coeficientes de Fourier contienen información acerca de la orientación y de la simetría de la región. La ventaja es que con pocos términos se logra una buena descripción para representaciones de contornos cerrados. Otra ventaja es que los descriptores de Fourier son invariantes a la rotación, el primer descriptor contiene información de la ubicación de la región en la imagen y los otros descriptores son invariantes a la ubicación.

Elipse de Ajuste (best-fit ellipse):

Las características elípticas (Fitzgibbon, Pilu, y Fisher, 1999) representan la región envuelta en una elipse, mediante el ajuste del borde de la region a una elipse se obtienen características de la elipse como el eje mayor, orientación, posición, excentricidad, etc.

Descriptor SIFT *Scale-Invariant Feature Transform*

En la presente investigación se utilizó una sola característica cromática conocida como el descriptor SIFT (Lowe, 2004) . La ventaja de esta característica es que no solo considera información del punto evaluado, si no también considera la información de los puntos de su entorno proporcionando un vector con información del punto y la vecindad.

El nombre SIFT proviene de la abreviación *Scale-Invariant Feature Transform* (Lowe, 2004) y es un algoritmo para describir características locales de una imagen mediante un vector. SIFT está basado en histogramas con orientación de gradiente. Es invariante a la rotación lo que permite comparar puntos en varias orientaciones. Se considera el punto con su orientación como la diferencia de gradientes a su vecindad. El gradiente es un vector compuesto por las primeras derivadas de las direcciones verticales y horizontales. Es un algoritmo robusto frente a cambios en la escala, ruido, distorsión geométrica, orientación y parcialmente invariante a la iluminación. Las principales etapas de cálculo son:

1. Se calcula la magnitud y la orientación del gradiente en la vecindad del punto utilizando la imagen suavizada.

2. Se crea un histograma, donde su máximo determinará la orientación del punto.
3. Si en el histograma existen picos superiores al 80 % del máximo se genera un nuevo punto para cada pico con su orientación correspondiente.
4. Se segmenta la vecindad en regiones de 4×4 píxeles.
5. Se genera un histograma de orientación de gradiente para cada región usando una ponderación Guassiana de ancho 4 píxeles.
6. Un píxel contribuye a todos sus vecinos, multiplicando la contribución por un peso $1-d$, donde d es la distancia al centro del segmento.
7. Por último se obtiene un histograma tridimensional de $4 \times 4 \times 8 = 128$ características considerando 8 direcciones principales y formando un vector con el valor de todas las casillas.
8. Este vector se genera para cada punto de interés.

2.1.3. Selección de Características, *Sequential Forward Selection*

Se busca encontrar el conjunto más pequeño de características que logre realizar la mejor clasificación, este procedimiento se llama selección de características. La idea es que un subconjunto de m características menor que el total de características extraídas realice la clasificación disminuyendo el error.

Sequential Forward Selection SFS (Duin y Mao, 2000) es la sigla de búsqueda secuencial hacia adelante. Es un algoritmo que se encarga de seleccionar las mejores m características de un conjunto de M características con $m < M$. Su funcionamiento básico consiste en que una función $F(\cdot)$ va evaluando un conjunto de características, comienza con el conjunto vacío y va agregando características, su funcionamiento es iterativo y en cada iteración va incorporando una nueva característica y evaluando $F(\cdot)$ para ver si se conserva o se descarta.

El criterio de selección utilizado con SFS fue el de Fisher (Sección 2.1.1).

2.1.4. Evaluación del Clasificador

En el diseño de clasificadores será necesario establecer métricas que proporcionen información acerca del nivel de certidumbre que estos logran en una clasificación. Para esto existe

un método de prueba llamado validación cruzada el cual busca aislar los sesgos de la muestra como estabilizar el resultado iterando en varias ocasiones sobre la clasificación. Por otra parte se describirá la tabla de confusión la cual sirve para clarificar cómo es el comportamiento de lo clasificado, validando cuántos son falsos positivos, negativos y cuántos fueron clasificados bien.

Validación Cruzada

Un problema habitual en el diseño de clasificadores es el sobre entrenamiento, es decir, el clasificador es capaz de realizar una buena clasificación solo en la muestra que fue utilizada para su entrenamiento, pero al momento de realizar cambios en el conjunto de imágenes, por ejemplo, al integrar nuevas muestras, estas no son clasificadas de forma correcta.

Para evitar el sobre entrenamiento es que la muestra se divide en un sub-conjunto utilizado para realizar el entrenamiento y otro para realizar el test o comprobación. Realizar este procedimiento varias veces tomando diferentes sub-conjuntos para entrenar y para comprobar se conoce como la técnica de validación cruzada.

Este método de validación consiste en dividir la muestra en n partes conservando la distribución global en cada subconjunto (Duda, Hart, y Stork, 2001). Si la muestra cuenta con N datos cada subconjunto quedará de N/n datos. Después se realiza el entrenamiento con las $N - N/n$ como conjunto de entrenamiento y se evalúa la clasificación con el N/n que quedó fuera del entrenamiento. Luego se separa otro subconjunto y se realiza el mismo procedimiento, esto se repite n veces con cada subconjunto, hasta que se tiene una validación cruzada con el total de la muestra.

Tabla de Confusión

Con ella se logra organizar el resultado para determinar tanto la cantidad de falsos positivos como de falsos negativos. En esta investigación tendremos dos posibles clases, a) pertenencia, b) no pertenencia, por lo que la tabla es la matriz $T_{2 \times 2}$ donde se ordenan los resultados de la clasificación.

$$T = \begin{bmatrix} TP & FP \\ TN & FN \end{bmatrix}$$

Donde:

- TP Verdadero Positivo: indica la cantidad de detecciones correctas.
- FP Falso Positivo: la cantidad de detecciones incorrectas.
- FN la cantidad de objetivos no detectados.
- TN indica la cantidad de elementos correctamente no detectados.
- TP + FP es la cantidad total de detecciones realizadas por el sistema.
- TP + FN son la cantidad real de objetos que el sistema debió detectar.

2.2. Geometría de las Múltiples Vistas

En esta sección se expondrá la teoría geométrica de múltiples vistas, las líneas epipolares, la definición de la matriz fundamental, y la restricción epipolar utilizada para determinar correspondencia de puntos entre imágenes. Luego se expondrá un método de auto-calibración sin asistencia, el cual provee las matrices de proyección contando al menos con cuatro vistas y puntos correspondientes entre ellas. Se verá cómo realizar proyecciones de un punto en 3D, esta técnica se puede utilizar para hacer reconstrucción o seguimiento de objetos en el espacio. Por último se explicará un método utilizado para descartar puntos en la correspondencia, seleccionando solo los puntos denominados *inliers*.

2.2.1. Geometría de dos vistas

Un punto en el espacio visto desde una cámara genera proyecciones en un plano 2D perdiendo la información de profundidad, por lo que la representación espacial 3D de un punto en una imagen se puede lograr si conocemos el modelo matemático que contiene el espacio donde se encuentra el punto. Esto se realiza con dos o más cámaras, donde se buscan puntos correspondientes entre ellas y se calcula matemáticamente su representación.

Si tenemos dos observadores, como se muestra en la Figura 2.2, que miran el mismo objeto; la proyección de cada punto de este objeto tendrá una correspondencia a lo largo de una línea del otro observador, esta línea se denomina línea Epipolar. Los lugares donde la línea epipolar

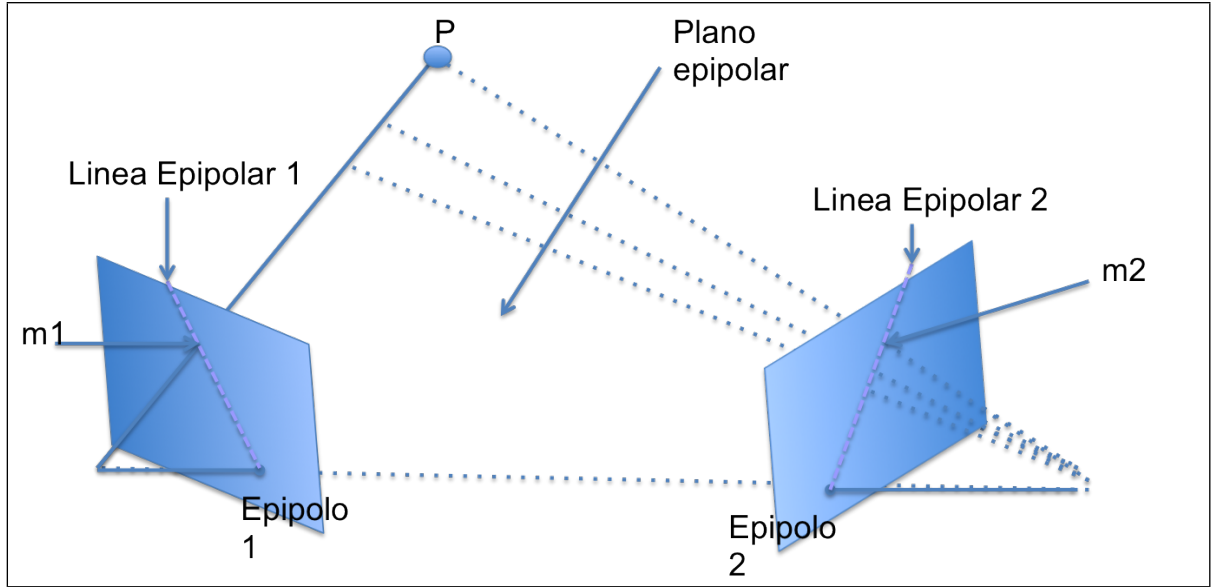


FIGURA 2.2. Líneas epipolares

une a los centros de proyección de los observadores y corta a los planos de proyección se llaman epipolos.

Si tenemos un punto definido como $M = [X \ Y \ Z \ 1]^T$, y su proyección en cada plano es m_1 y m_2 con $m_i = [x_i \ y_i \ z_i]^T$ con $i \in \text{planos}$. Si hacemos la transformación proyectiva de ambos puntos, se puede describir como:

$$\lambda_1 m_1 = AM \quad (2.17)$$

$$\lambda_2 m_2 = BM$$

Donde A y B son las matrices de proyección de 3×4 por cada imagen, m_1 , y m_2 están en coordenadas homogéneas.

La matriz fundamental fue introducida por Faugeras (Luong y Faugeras, 1995). Se utiliza en visión estéreo y es una matriz de 3×3 , la cual representa la correspondencia entre un punto de una imagen en una línea de la otra imagen. Esta matriz se utiliza en la geometría Epipolar. Un punto en el espacio 3D representado de forma homogénea podrá ser manipulado si se opera matemáticamente con él a través de transformaciones proyectivas y mediante la Matriz Fundamental se realizan escalamientos, traslaciones, rotaciones. Se conoce a F como la matriz

fundamental y es constante para una geometría bifocal dada, no depende de m_1 , m_2 , M . Las representaciones de las líneas epipolares :

$$l_2 = Fm_1 \quad (2.18)$$

$$l_1 = F^T m_2 \quad (2.19)$$

l_2 es la línea epipolar en la segunda imagen de m_1 A y B . l_2 es la proyección del rayo $\langle C_1, m_1 \rangle$.

Si m_2 pertenece a la recta entonces $m_2^T l_2 = 0$

$$m_2^T Fm_1 = 0 \quad (2.20)$$

Donde F es la matriz fundamental del modelo geométrico que soporta a los puntos m_1 m_2 , y la fórmula 2.20 se conoce como restricción epipolar, esto implica que m_2 tiene que estar sobre la línea epipolar de m_1 . Aunque se cumpla la restricción epipolar no se logrará poder saber el lugar exacto de M en el espacio 3D, ya que solo se cuenta con información para saber que pertenece a una línea.

Dos puntos correspondientes m_1 y m_2 satisfacen la condición epipolar con muy baja probabilidad, es por ello que es necesario introducir un criterio de correspondencia para determinar la cercanía del punto proyectado con la línea epipolar llamada restricción bifocal práctica:

$$d = \frac{|m_2^T Fm_1|}{\sqrt{l_1^2 + l_2^2}} < d_0 \quad (2.21)$$

Donde d_0 es la distancia de lejanía que se tolera, l_1 , l_2 son las líneas epipolares respectivas de los puntos m_1 , m_2 .

2.2.2. Geometría de tres vistas

Si se busca una correspondencia de un punto en dos vistas, a partir de la matriz fundamental contamos con la información para buscarlo sobre la línea epipolar. Si lo que se busca es una representación del punto en 3D y generar sus proyecciones en más vistas se requiere contar con puntos correspondientes en al menos dos vistas y las matrices de proyección de cada vista.

Una representación geométrica del espacio visto desde 3 puntos se puede realizar con los tensores trifocales (Hartley, 1997), estos pueden ser obtenidos a partir de las matrices de proyección de tres cámaras. La expresión algebraica de los tensores trifocales es:

$$\mathbf{T}_i^{jk} = (-1)^{i+1} \begin{vmatrix} \sim \tilde{\mathbf{a}}_i \\ \tilde{\mathbf{b}}_j \\ \tilde{\mathbf{c}}_k \end{vmatrix} \quad \text{para } i, j, k = 1, 2, 3 \quad (2.22)$$

Donde T es el tensor trifocal. El término $\sim \tilde{\mathbf{a}}_i$ representa la matriz A sin la fila i . Los términos $\tilde{\mathbf{b}}_j \tilde{\mathbf{c}}_k$ corresponden a la fila j y la fila k de las matrices B y C respectivamente. Las matrices A B C son las matrices de proyección para cada una de las vistas.

Los que deben cumplir las trilinearidades de Shashua (Shashua y Werman, 1995)

$$\begin{aligned} D_1 &= m_1^T (x_3 \mathbf{T}^{13} - x_3 x_2 \mathbf{T}^{33} + x_2 \mathbf{T}^{31} - \mathbf{T}^{11}) = 0 \\ D_2 &= m_1^T (y_3 \mathbf{T}^{13} - y_3 x_2 \mathbf{T}^{33} + x_2 \mathbf{T}^{32} - \mathbf{T}^{12}) = 0 \\ D_3 &= m_1^T (x_3 \mathbf{T}^{23} - x_3 x_2 \mathbf{T}^{33} + y_2 \mathbf{T}^{31} - \mathbf{T}^{21}) = 0 \\ D_4 &= m_1^T (y_3 \mathbf{T}^{23} - y_3 y_2 \mathbf{T}^{33} + y_2 \mathbf{T}^{32} - \mathbf{T}^{22}) = 0 \end{aligned} \quad (2.23)$$

Re-proyección en la tercera

Una vez obtenidas las matrices de proyección, se pueden proyectar puntos en cada imagen y ellos tendrán su representación 3D en 2D .

Si m_1, m_2, m_3 en imágenes diferentes y son puntos correspondientes la proyección de $\hat{m}_3$

$$\hat{m}_3 = \begin{bmatrix} \hat{x}_3 \\ \hat{y}_3 \\ 1 \end{bmatrix} = \frac{1}{m_1^T (T^{13} - x_2 T^{23})} \begin{bmatrix} m^T (T^{11} - x_2 T^{31}) \\ m^T (T^{12} - x_2 T^{32}) \\ m^T (T^{13} - x_2 T^{33}) \end{bmatrix} \quad (2.24)$$

Para tres vistas la correspondencia de tres puntos se evalúa con $\|m_3 - \hat{m}_3\| < \epsilon$.

2.2.3. Auto-calibración con *Bundle Adjustment*

El proceso de analizar y extraer la información métrica de imágenes 2D es conocido como el proceso de calibración de la cámara. Lo que se busca es el modelo de proyección que relaciona los modelos de coordenadas entre la imagen y el mundo. La calibración es una expresión matemática con una matriz de proyección de cada cámara. En esta Sección se revisarán los elementos utilizados para realizar una estimación geométrica, mediante un proceso de auto-calibración no asistido.

Una imagen 2D tiene múltiples distorsiones, algunas propias del sensor que captura la imagen, otras distorsiones debidos a la ubicación, rotación, etc, del objeto. Los parámetros que influyen en la transformación de 3D a 2D de una imagen son los parámetros intrínsecos, los parámetros extrínsecos de la cámara, la distorsión del lente, como los parámetros del manipulador.

En visión por computador se han desarrollados dos métodos para la calibración: el de auto-calibración y el de calibración fotogramétrica. La calibración fotogramétrica se utiliza para buscar una reconstrucción precisa del objeto. En esta calibración es necesario utilizar un objeto del cual se conozca su geometría previamente. Se escogen N puntos del objeto de referencia y se adquiere una imagen donde los puntos serán vistos en 2D con coordenadas $w_i = [u_i \ v_i \ 1]^T$, para $i = 1, \dots, N$. Si conocemos el modelo de proyección podríamos calcular una estimación teórica de cada punto en 3D, $\hat{w}_i = f(M'_i)$, donde M'_i son los puntos de interés medidos en 3D y f es la función de proyección que involucra los parámetros de la cámara, del lente y del manipulador. De esta manera el problema de la calibración se transforma en un problema de optimización, el cual busca el error mínimo entre w_i y $\hat{w}_i$.

La auto-calibración (*self-calibration*) requiere de varias imágenes de una misma escena, en todas ellas se buscan puntos que sean correspondientes. La rigidez de la escena impone restricciones sobre los parámetros de la cámara, por lo que se puede encontrar el mejor modelo que otorgue esta correspondencia (Dornaika y Chung, 2001). Richard Hartley (2003) describe un método de auto-calibración llamado *Bundle Adjustment* que ha sido ampliamente adoptado en visión por computador debido a sus buenos resultados.

El método calibración fotogramétrica logra mayor precisión en la reconstrucción 3D, sin embargo, frente a cualquier cambio, por ejemplo, una pequeña variación en la posición de la cámara, se descalibraría el sistema, lo que se presenta como una gran desventaja. El método de auto-calibración, si bien, presenta un problema en que la reconstrucción 3D queda afectada por un factor de escala, para lo que buscamos (seguimiento entre vistas), el factor será mismo aplicado a todas las imágenes por lo que resultaría irrelevante.

Bundle Adjustment consiste en estimar las matrices de proyección para cada cámara. Se trata de un método de auto-calibración y se utiliza en pos de realizar una reconstrucción proyectiva (Richard Hartley, 2003).

Este método consiste en minimizar el error de la proyección con el punto real. Lo que busca, es ajustar los rayos de los centros ópticos de cada cámara con la correspondencia 3D entre los puntos. De esa manera sin información a priori es posible determinar las matrices de proyección a partir de la correspondencia de puntos 3D.

Bundle Adjustment trabaja con al menos cuatro imágenes y sus puntos correspondientes $M_j^i, j = 1, \dots, n$ puntos $i = 1, \dots, m$ imágenes, se determinan las matrices de proyección de cada cámara $\{P^i\}$ y los puntos 3D $\{M_j\}$ con una rotación t^i

$$\sum_{ij} \|x_j^i - (P^i M_j + t^i)\|^2$$

Se busca minimizar P^i, t^i, M_j con P^i es una matriz de 2×3 , M_j es un vector, $x_j^i = (x_j^i, y_j^i)^T$

1. Se trasladan los puntos calculando su centro de gravedad.

$$t^i = \langle x^i \rangle = \frac{1}{n} \sum_j x_j^i$$

2. Cada punto de cada imagen se expresa en coordenadas con respecto al centro de masa.

$$x_j^i \longleftarrow x_j^i - \langle x^i \rangle$$

3. Se construye $2m \times n$ la matriz de puntos centrados y se descompone en SVD.

4. La matrices M^i se obtiene de las tres primeras filas de U multiplicado por sus valores singulares.

2.2.4. Método RANSAC para Estimar Puntos Inliers

RANSAC, RANdom SAMpling Consensus (Fischler y Bolles, 1981), es un algoritmo que sirve para descartar puntos que se encuentran fuera de un rango. Lo que supone es que la muestra obtenida posee errores de medición por lo que los puntos extremos o fuera de su distribución se desechan. El método consiste en que se eligen n puntos aleatorios y se evalúa una función objetivo, y luego se calcula la proyección de todos los puntos sobre la estimación encontrada. A partir de un rango de tolerancia se determina cuales puntos quedan dentro de lo estimado por las muestras y cuales fuera, los que son considerados como outliers. Para la presente investigación es de interés utilizar este método porque con él se podrán estimar los puntos que no son correspondientes. Para esto se construye el modelo geométrico y se evalúan las proyecciones, los puntos que quedan como outliers se descartan y se vuelve a construir el modelo.

3. MÉTODO PROPUESTO

Esta investigación propone un método que cuenta con tres grandes partes. La primera tiene relación con el diseño de un **detector** de gatillos. El segundo, tiene relación con la **estimación geométrica** de las matrices de proyección de cada vista. La técnica utilizada en este trabajo tiene la particularidad de que opera sin asistencia humana para su estimación. El tercero, tiene relación con la propuesta que se realizó para el diseño de un **clasificador en múltiples-vistas** que opera con información geométrica y cromática de forma simultánea.

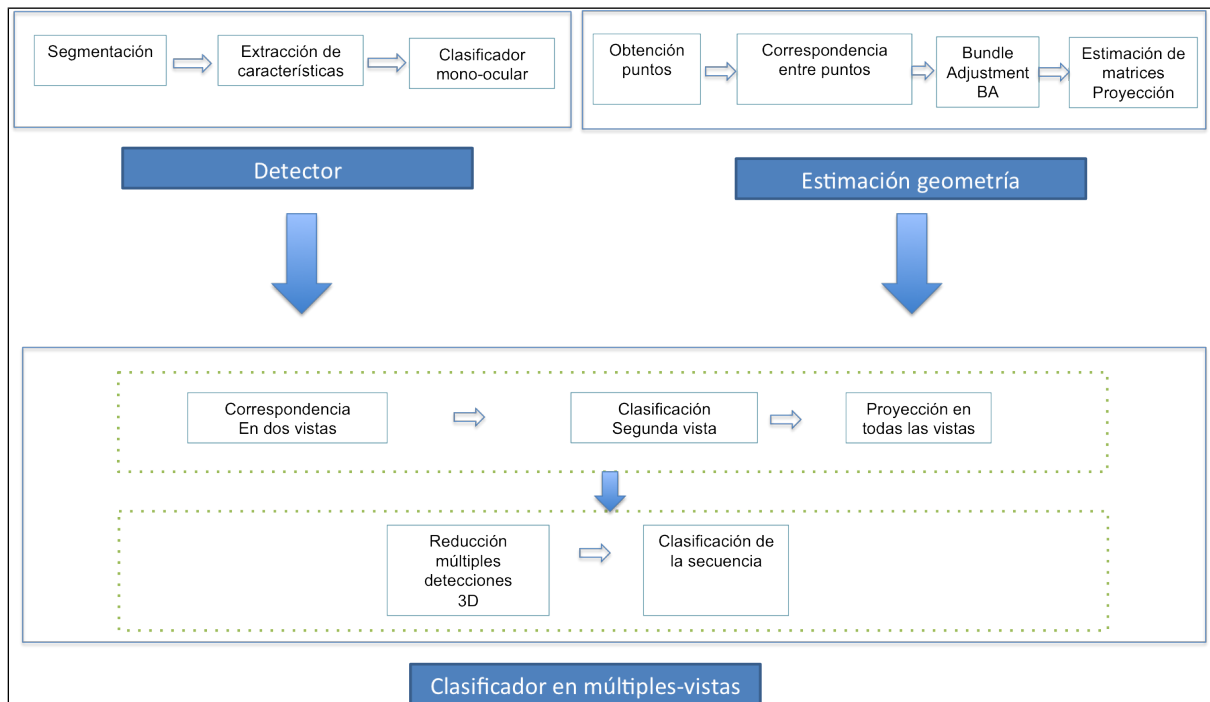


FIGURA 3.1. Diagrama de bloques del clasificador múltiples-vistas

Como se muestra en la Figura 3.1 el clasificador múltiples-vistas cuenta con una etapa de estimación del modelo geométrico, en paralelo se analiza cada vista para la detección de posibles gatillos de manera mono ocular. Luego, utilizando la información de manera complementaria se realiza correspondencia en dos vistas. Posteriormente, se proyectan los puntos detectados en todas las vistas, para finalizar con un cálculo de puntos representativos y realizar una clasificación final de la secuencia.

3.1. Detector

Un detector es un programa que logra entregar información acerca de la presencia de un patrón conocido dentro de una imagen. Para diseñar un detector y operarlo, lo mejor sería disponer de imágenes con poco ruido y alto contraste, sin embargo, las imágenes de rayos X son altamente ruidosas y con poco contraste, lo cual impone mayores desafíos. En esta sección se revisarán algunos conceptos relacionados con pre-procesar la imagen, segmentación, búsquedas de patrones en las imágenes, características y clasificadores.

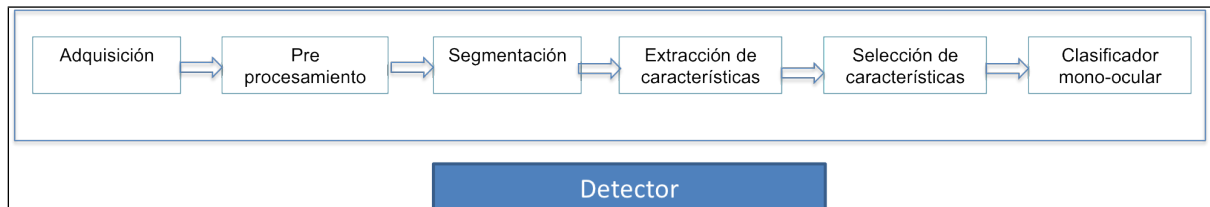


FIGURA 3.2. Diseño del detector

3.1.1. Adquisición de Imágenes

Se adquirieron imágenes con poco control del entorno, intentando simular el transporte de una pistola dentro de un bolso de forma natural. Se utilizó una mochila y un bolso pequeño de cintura como contenedores de los objetos a analizar.

Entre los objetos utilizados se contó con armas pequeñas, un laptop, una cámara fotográfica, una grabadora de cassettes, cables, monedas, pendrives, cuadernos, ropa y toallas.

Se llenaron los bolsos hasta su capacidad máxima para que no sufrieran deformaciones en los movimientos. Se buscó que el bolso tuviese un comportamiento rígido, sin que los objetos se pudieran mover libremente por su interior, ya que de esta manera se puede aplicar técnicas descritas en la Sección 2.2. Todos los objetos fueron introducidos de forma aleatoria.

Como el Laboratorio cuenta con solo una máquina de rayos X, al momento de adquirir las secuencias se realizan movimientos aleatorios de rotación y traslación en los planos X , Y , Z del bolso para cada imagen de la secuencia. Los movimientos de los bolsos se realizaron de forma manual sin mayor control, pretendiendo simular un caso real. Se tuvo la precaución de

que las rotaciones no fueran mayores de 20° para encontrar correspondencia entre las imágenes, SIFT no logra encontrar correspondencia si la diferencia entre las vistas es mayor (Lowe, 2004). Para lograr un mayor control de los movimientos se tendría que utilizar un manipulador robótico, lo que fue descartado ya que le resta espontaneidad a la toma de secuencias.

3.1.2. Procesamiento de las Imágenes

Para lograr una buena segmentación se realizaron dos tipos de procesamiento en las imágenes. Uno se realizó a nivel global y otro a nivel local o sobre una ventana (Mery, Filbert, y Jaeger, 2005; R. R. da Silva, 2010).

El proceso que se realiza a nivel global es aplicado sobre toda la imagen y su objetivo es normalizar la imagen para eliminar distorsiones producidas por la máquina y aumentar el contraste. También se redimensionaron las imágenes para aumentar la velocidad en su procesamiento.

El proceso realizado a nivel local se hace para aumentar el contraste y solo produce cambios en forma temporal sobre una ventana de la imagen a la vez.

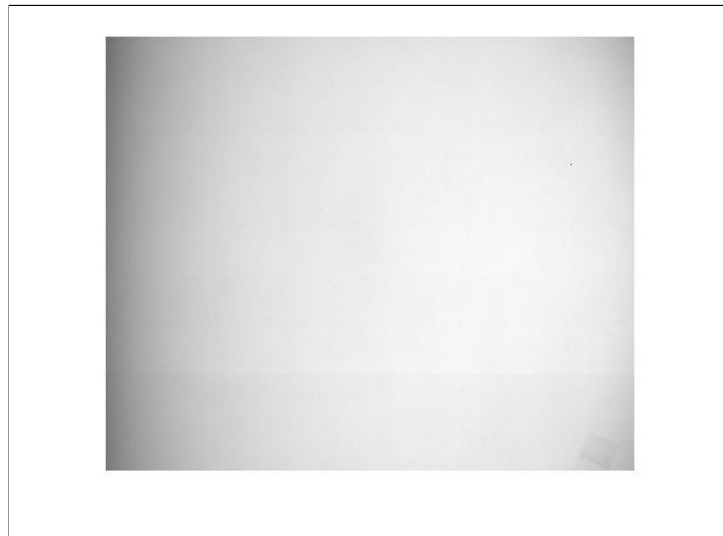


FIGURA 3.3. Imagen de referencia para normalización

En la Figura 3.3 se muestra una imagen adquirida con la máquina vacía, esta imagen es utilizada como imagen de referencia para normalizar todas las imágenes.

Para la normalización de las imágenes se calculó la diferencia entre la imagen de referencia con el resto de las imágenes. La utilización de la diferencia entre las imágenes para la normalización es método utilizado en el procesamiento de imágenes para aplicaciones industriales (R. R. da Silva, 2010). Al utilizar esta técnica no se aumenta el ruido en las imágenes y no es necesario ningún cálculo extra para realizar la corrección como lo sería si se aplicara corrección mediante la multiplicación de un gradiente.

$$I'_n = I_n - I_o$$

Donde I'_n es la imagen normalizada, I_n es la imagen original y I_o es la imagen de referencia.

La imagen original se adquiere con un tamaño 2208×2688 px lo que representa un costo al momento de su procesamiento. Para buscar una mayor velocidad de procesamiento la reducción del tamaño es una alternativa a considerar. Cuando se realizan escalamientos en las imágenes se pierde información dado que un pixel no se puede dividir. Esto podría llegar a ser un problema si lo que se está buscando es algo pequeño y se requiere un alto nivel de detalle. Como el objetivo buscado son armas, una disminución en el tamaño de la imagen no afectará la detección. Las imágenes se escalaron a un 20% (441×556 pixeles) del tamaño original, de esa forma el costo en tiempo de procesamiento se redujo considerablemente.

La imagen original cuenta con una resolución de $2^{12} = 4096$ tonos de gris por cada pixel. Cuando los rayos X interactúan con la materia una parte de los rayos son absorbidos y otra transmitidos. La absorción depende de la intensidad como de la distancia que atraviesa

$$I_x = I_o e^{(-\mu/\rho)\rho x}$$

Donde μ es el coeficiente de absorción, ρ la densidad del material, x es la distancia que atraviesa el rayo.

Se pueden diferenciar variaciones en los objetos debido al grosor que tenga o variaciones entre objetos debido al cambio del coeficiente lineal de absorción.

Para estudios relacionados con áreas biológicas podría ser de interés tener un nivel mayor de resolución, por lo que la máquina tendría que proporcionar un rango mayor de variaciones en intensidad como en la cantidad de tonos de gris por cada pixel. Sin embargo, para una alta gama de estudios, la resolución y el rango de intensidades con que cuenta la máquina es suficiente¹. Para el caso de las armas hay que considerar que estas son metálicas y que los metales tienen un alto coeficiente de absorción como consecuencia de su elevado número atómico. Esto hace que los metales se diferencien claramente de otros objetos orgánicos como inorgánicos. Considerando lo anterior, es que para cada imagen de rayos X se realizó una transformación lineal de colores mediante función LUT *Look Up Table* reduciendo de 4096 a 256 la escala de grises.

La función LUT *Look Up Table* realiza una transformación lineal de tonos de gris modificando el histograma.

La imagen original está representada por $f(x, y)$ y calculamos la imagen transformada como una transformación lineal.

En este caso la función de transferencia se calculó según el valor mínimo $\min(f(x, y))$ y el valor máximo $j = \max(f(x, y))$ de la imagen original. La función de transferencia T va desde el valor mínimo i hasta el valor máximo j por un factor de escalamiento $df = \frac{255}{j-i}$. Por último se aplica la función de transferencia a la imagen quedando una imagen transformada en $g(x, y)$ dentro del rango $[0 \dots 255]$.

Una vez que las imágenes han sido normalizadas se procesan para realizar detecciones en ella. Como no se sabe si el objeto se encuentra presente o en qué lugar de la imagen podría estar ubicado, es que se tiene que realizar una búsqueda exhaustiva en toda la imagen. Un enfoque muy utilizado por la comunidad, para la detección de objetos al interior de las imágenes es el de ventana deslizante (*Sliding Windows*).

¹En un reciente estudio se logró detectar diferencias en firmeza de kiwis mediante rayos X (Mondragon, Leiva, Aguilera, y Mery, 2011), a pesar de que estos estudios por lo general, se realizan con máquinas que soportan rangos inferiores de intensidad.

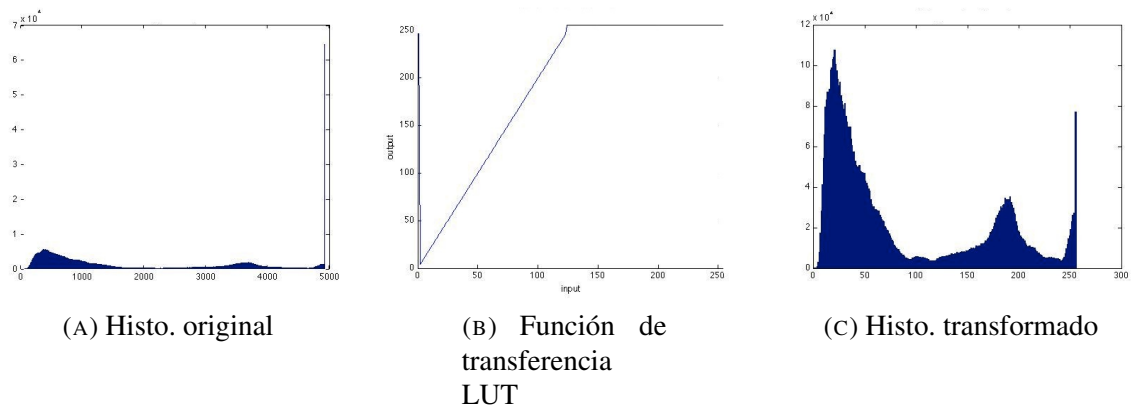


FIGURA 3.4. Histograma original, transformado y función de transformación LUT

3.1.3. Búsqueda en la Imagen *Sliding Windows*

Existen aplicaciones exitosas que utilizan este método como (Viola y Jones, 2001) para la detección de caras y de personas (Dalal y Triggs, 2005). En su funcionamiento se van adquiriendo submuestras de la imagen que quedan en una ventana o parche. Este se va deslizando en el eje x y en el eje y . La ventana recorre toda la imagen varias veces, el tamaño y el salto de avance son parámetros que normalmente varían en cada pasada. Esta variación se realiza para detectar el objeto en diferentes tamaños o perspectivas. Cada ventana se van obteniendo características las cuales son clasificadas con algún clasificador ad-hoc.

Como la captura de imágenes se realizó dentro de un dispositivo preparado para tal efecto, la distancia para la captura de imágenes en la máquina se mantuvo constante entre las imágenes, de esta manera se previno un efecto de escalamiento por distancia. De manera heurística se determinó el tamaño de la ventana deslizante *Sliding Windows*: 80×80 px.

Cada punto de la imagen tiene que pasar 4×4 veces por la ventana, es por ello que se calculó que el salto para el desplazamiento de la ventana es de 20 px en ambos ejes.

Ahora el problema se redujo a diseñar un detector que opere sobre la ventana deslizante o *Sliding Windows* y que entregue un resultado binario: se encuentra presente o no se encuentra presente en la ventana.

3.1.4. Segmentación

El diseño de un detector busca resaltar el objeto del resto de la imagen y hacer este proceso se denomina segmentación. La segmentación consiste en dejar en una imagen binaria los objetos de interés. En este caso, el problema de la segmentación tiene que considerar que las imágenes son de rayos X y cuentan con un alto nivel de ruido y un bajo contraste. Además, las imágenes pueden presentar distintos niveles de oclusión y superposición de objetos.

Como se describió en la Sección 1.3 existen métodos que realizan segmentación de imágenes de rayos X considerando que la absorción del rayo varía de según el coeficiente de absorción del material, por lo que, a partir de la utilización de energía dual se puede calcular una aproximación del número atómico del componente, este método se utiliza para reconocer el tipo de material y es útil para identificar explosivos (Chan y cols., 2010). Pensar en utilizar energía dual en esta investigación no es de mayor aporte y encarecer los costos en maquinaria y en tiempos de procesamiento. Otro enfoque es buscar bordes como el descrito en (Oertel y Bock, 2006), la detección de bordes presenta la debilidad de que si el objeto está superpuesto con otro objeto de similar coeficiente, o se encuentran presentes coeficientes de absorción similar, el método presentará problemas.

Para determinar la segmentación en la presente investigación se utilizó una corrección de la imagen y umbralización. Debido a la dificultad que presentan las imágenes es que se diseñaron cuatro métodos de segmentación y después de realizar pruebas se determinó cual método es el más eficiente: Umbral, Promedio, Otsu, LUT.

La segmentación por umbral establece un valor límite el cual se comporta como umbral para dejar la imagen binaria.

$$g(x, y) = \begin{cases} 1 & \text{si } f(x, y) < a \\ 0 & \text{si } f(x, y) \geq a \end{cases} \quad (3.1)$$

Como se muestra en la ecuación 3.1 se establece un criterio el cual opera frente a cada pixel verificando que sea menor que a , si no, es quedará en 0 el valor del pixel.

Considerar un umbral estático para toda la imagen podría ser útil para ciertas regiones de la imagen pero defectuoso para otras. Un umbral variable podría adecuarse según el nivel de contraste local que posea la imagen. Se puede encontrar un umbral variable en relación al promedio de tonos de gris de la ventana, una ventaja es que se puede realizar de manera automática y no supervisada.

$$\frac{1}{n * m} \sum_{i=1}^n \sum_{j=1}^m f(i, j) \quad (3.2)$$

Donde $f(x,y)$ es una matriz de $n \times m$ que contiene la imagen.

Otro método utilizado para calcular el umbral es el que propuso Otsu (1979). Este método consiste en buscar en el histograma el punto donde se encuentre un máximo según el criterio de Fisher. Como se en la Figura 3.5 se establece el Umbral en línea roja. Como se muestra en

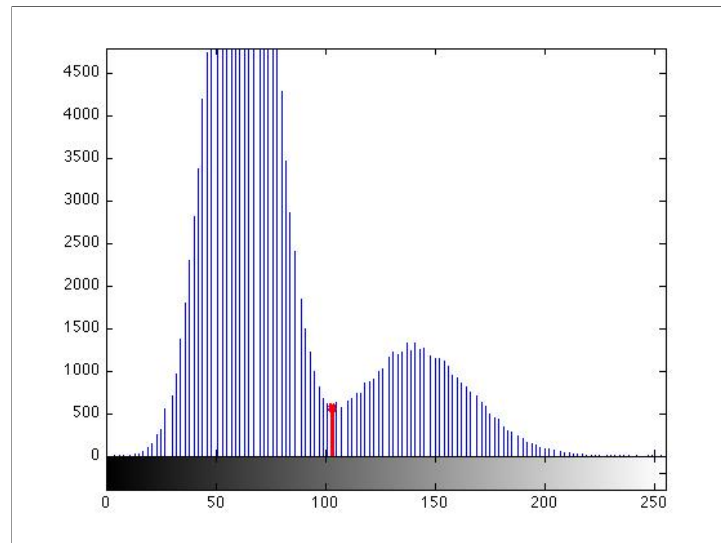


FIGURA 3.5. Determinación de umbral mediante otsu

la Figura 3.5 el umbral se define según el criterio de Fisher.

Por último se realizó una corrección del histograma sobre la ventana deslizante calculado una función de transferencia LUT que modificó el histograma a nivel local, para dejarlo con valores entre $[0 \dots 255]$. De esta manera se resalta el contraste a nivel local, después se aplica un umbral sobre el histograma transformado.

Cada método propuesto para realizar segmentación es aplicado sobre la ventana deslizante en cada paso.

A todos los métodos de segmentación propuestos se les aplicó una constante, de esa forma se realizaron variaciones sobre la vecindad del umbral definido.

3.1.5. Características

Extracción de Características

Una vez que se ha definido el método de segmentación por cada ventana deslizante se obtendrán características. Se consideró que toda arma de fuego pequeña posee un gatillo. Los gatillos presentan similitudes fundamentalmente en su forma, son redondeados, tienen una cantidad de huecos constante, son de un tamaño similar, son metálicos, es por ello que se considero la extracción de características geométricas.

La características geométricas obtenidas se pueden dividir en cuatro grupos:

1. Características geométricas estándares, se obtienen con el comando `regionprops` (MathWorks, 2010) donde se obtiene el area, orientación, número de Euler, y otros.
2. Características invariantes de momentos de Hu (Hu, 1962; Sonka, Hlavac, y Boyle, 1998).
3. Descriptores de Fourier (Persoon y Fu, 1977).
4. Características elípticas (Fitzgibbon y cols., 1999).

Selección de características

Se extrajeron 44 características geométricas por cada ventana obtenida mediante la metodología descrita en la Sección 3.1.3 de las cuales se fueron seleccionando las mejores mediante SFS-FISHER como se describe en la Sección 2.1.3. Se realizó una búsqueda secuencial para encontrar el mejor conjunto de características entre 1 a 17 características.

El paso siguiente es diseñar un clasificador que logre detectar gatillos de manera mono ocular.

3.1.6. Clasificación

Como se mencionó en la Sección 2.1 el reconocimiento de un patrón se podría realizar de manera supervisada o no-supervisada. En este trabajo se utiliza la metodología supervisada.



FIGURA 3.6. Etapas para el diseño del clasificador

Como se muestra en la Figura 3.6 en una etapa se obtienen las imágenes para el entrenamiento, luego se realiza el entrenamiento el cual va siendo re-alimentado a partir de las clasificaciones, por último se realiza una validación cruzada para evaluar el clasificador. En el desarrollo del presente estudio se confeccionaron cuatro clasificadores de gatillos mono oculares y se realizaron comparaciones entre ellos.

Imágenes para Entrenamiento

Para el entrenamiento del detector se realizó una búsqueda de imágenes de armas pequeñas en Internet (Fig. 4). Se seleccionaron 64 modelos diferentes los que fueron utilizados en el entrenamiento. Los modelos seleccionados incluyeron desde revólveres hasta pistolas automáticas, considerando armas de fuego y de fogeo.

Las imágenes obtenidas de internet fueron pre-procesadas dejándolas del mismo tamaño y se corrigió el contraste.

Se fue recorriendo las imágenes con una ventana deslizante para construir un conjunto de sub-imágenes en las cuales se encontraba el gatillo, las que se dejó como clase 1 (Fig. 3.7b) y las que no, como clase 2 (Fig. 3.7c). Para el conjunto de gatillos se seleccionaron los parches de

forma manual. Para el conjunto de entrenamiento de la clase no-gatillo se eligieron 64 parches al azar. Durante el proceso de entrenamiento la cantidad de muestras de la clase no-gatillo fue aumentando y re-alimentando el entrenamiento.

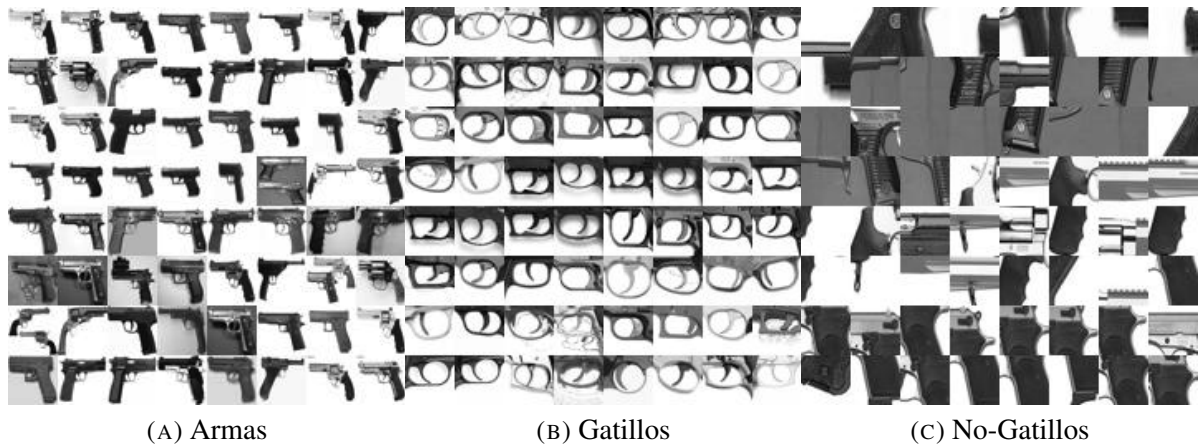


FIGURA 3.7. Imágenes de pistolas, gatillos , no-gatillos

La Figura 3.7a muestra pistolas seleccionadas desde internet, definidas como las imágenes de entrenamiento. La Figura 3.7b tiene una muestra de las imágenes de entrenamiento para la Clase 1: Gatillo. La Figura 3.7c tiene una muestra de las imágenes de entrenamiento para la Clase 2: No-Gatillo.

Entrenamiento

La Figura 3.8 muestra al lado izquierdo detecciones realizadas por un clasificador *LDA*. Al lado derecho, detecciones realizadas por un clasificador *Mahalanobis*.

El proceso de entrenamiento fue iterativo y re-alimentado en cada iteración (Fig. 3.6). El entrenamiento comenzó con dos clases, 64 imágenes de gatillos y 64 imágenes de no-gatillos. Se evaluó el clasificador con las imágenes de entrenamiento y se fueron guardando las clasificadas como gatillos que no lo eran en el conjunto de entrenamiento de no-gatillo. Al final la clase de no-gatillos quedó con 635 imágenes y el conjunto de gatillos se mantuvo en 64 imágenes.

Se fue aumentando el número de la muestra en la clase de no gatillos con el interés de disminuir los falsos positivos.

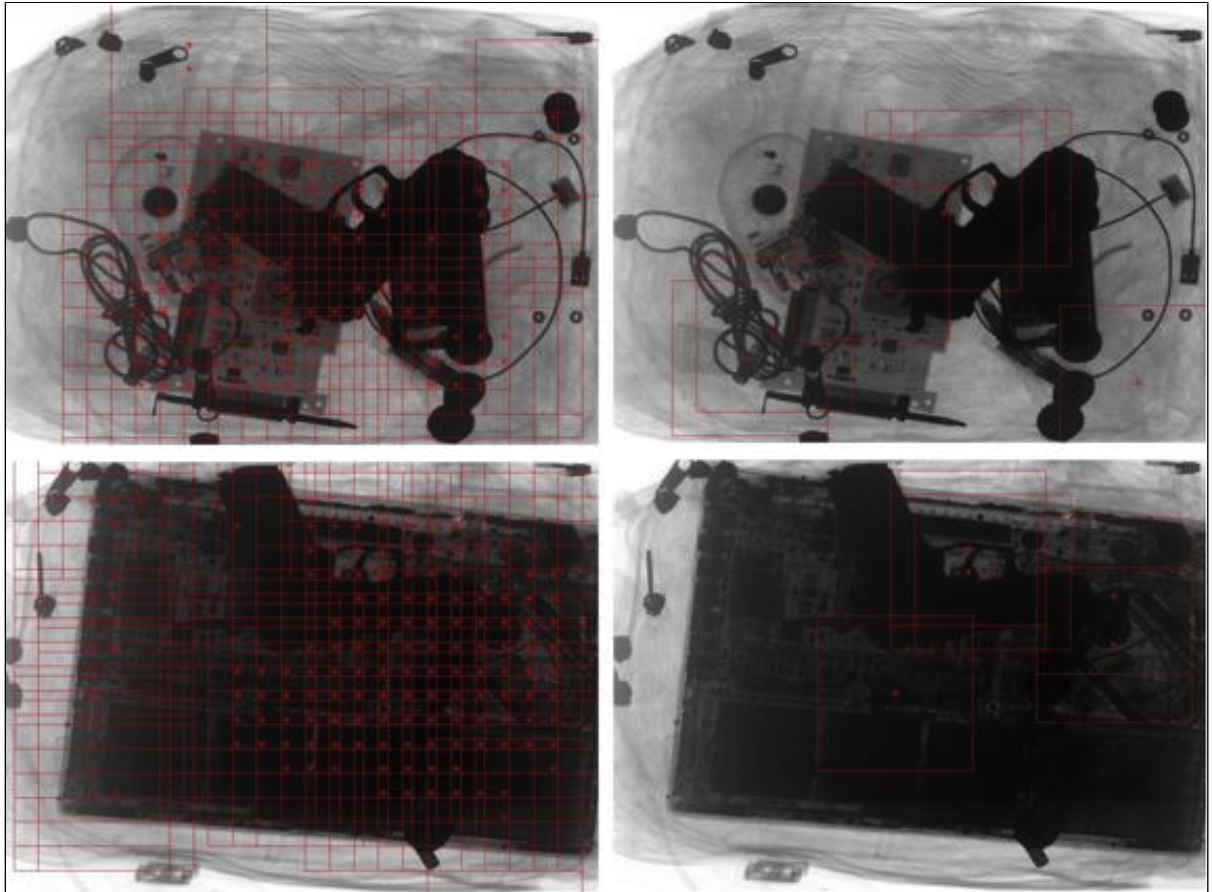


FIGURA 3.8. Comparación de clasificadores en fase de entrenamiento.

Este procedimiento de re-alimentación en el entrenamiento se realizó porque el diseño de clasificadores que utilizan *Sliding Windows* requieren de una gran cantidad de muestras para el entrenamiento (Viola y Jones, 2001).

Para prever que el clasificador no se sobre entrenara, se consideró que las imágenes utilizadas en el diseño del clasificador (fotografías adquiridas desde Internet) fueran diferentes a las imágenes de rayos X utilizadas.

Validación

Por último, para evaluar el desempeño del clasificador se utilizó la técnica de validación cruzada como se describe en la Sección 2.1.4. Se realizó una estratificación en 10 grupos con un intervalo de confianza de un 95 % de probabilidad (Duda y cols., 2001). La estratificación

permitió evaluar el desempeño de cada clase con la misma probabilidad de ocurrencia independiente del número de muestras. También se utilizó una tabla de confusión como la explicada en Sección 2.1.4, para verificar los falsos positivos, falsos negativos y de esta forma evaluar el clasificador.

3.2. Estimación Geométrica

La implementación de un detector en múltiples vistas tiene que considerar que el ambiente productivo es poco controlado, con altos niveles de movimiento de objetos como de personas. Las posibilidades de controlar los factores que intervienen en la inspección dentro de los aeropuertos o las industrias son bajas. Es por ello que se ha elegido el método de auto-calibración propuesto por Mery (2011a) como el factible para la estimación geométrica.

La auto-calibración requiere de puntos correspondientes entre la imágenes. La labor de encontrar puntos correspondientes se puede realizar de forma manual (asistida) o automática (no asistida). Realizar la correspondencia asistida presenta una gran desventaja, la cual es el tiempo y precisión que tiene que tener una persona para señalar cuales son los puntos correspondientes en la imágenes. Sería absurdo pensar que los inspectores realicen la labor de asistir el proceso de calibración para cada secuencia en un aeropuerto. El problema de obtener puntos característicos y garantizar su correspondencia entre las vistas de forma automática no es trivial.

Descriptor visuales describen características elementales tales como la forma, el color, la textura o el movimiento. En este caso, lo que se requiere es un descriptor visual que represente solo a un punto. Uno de los descriptores ampliamente utilizados en visión por computador es el descriptor SIFT, con él es posible encontrar puntos correspondientes entre las vistas de manera no asistida (Lowe, 2004).

Como se muestra en la Figura 3.9 para el cálculo de la estimación geométrica lo primero es la obtención de puntos de interés, luego se busca la correspondencia entre ellos para que con *Bundle Adjustment* se puedan calcular las matrices de proyección.

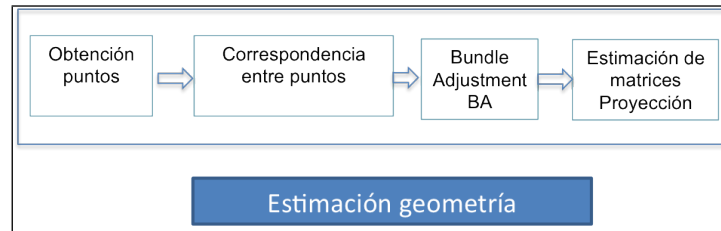


FIGURA 3.9. Estimación geométrica

Para descartar posibles errores en la correspondencia, se acotó la búsqueda a aquellos puntos que poseen mayor información, los que llamamos **puntos de interés**. Estos puntos podrían ser máximos locales de operadores diferenciales como bordes o esquinas. Un detector de esquinas popular es el detector de Harris y Stephens (1988), es rápido y con su utilización reducimos la búsqueda de correspondencia a puntos que contienen alta información en la imagen. A cada punto detectado como esquina, se le extraen características SIFT.

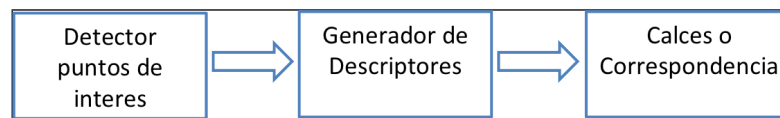


FIGURA 3.10. Obtención puntos correspondientes

Por último, para determinar si los **puntos de interés** de diferentes imágenes corresponden al mismo punto hay que comparar la distancia Euclideana de los descriptores SIFT (Lowe, 2004).

Una vez que se logra establecer cuáles puntos son correspondientes entre las vistas se procede al cálculo del modelo proyectivo que posee el sensor que realiza las capturas de las imágenes.

Para evitar cualquier distorsión en la correspondencia debido a alguna deformación del objeto entre las imágenes, lo ideal sería contar con diferentes imágenes adquiridas de forma simultánea o sincronizada. Sin embargo tomar cuatro imágenes de rayos X de manera sincronizada requiere contar con cuatro máquinas que operen al mismo tiempo, lo cual sería un hardware extremadamente costoso. Para estas situaciones es que la adquisición de secuencias

de imágenes es una solución óptima, solo hay que prever que los objetos no sufran deformaciones entre las vistas.

Encontrar el modelo geométrico a partir del movimiento en la escena es conocido como SfM, *Structure from Motion*, estructura a partir del movimiento (Jebara, Azarbayejani, y Pentland, 1999). Mediante su implementación se logra obtener una representación de la geometría epipolar a partir de al menos cuatro imágenes que posean ocho puntos correspondientes. Como se explicó en la Sección 2.2.3, mediante *Bundle Adjustment* se logra calcular las matrices de proyección para cada vista.

Con las matrices de proyección de cada vista, se puede calcular la geometría epipolar y la representación del espacio 3D de las imágenes en cada secuencia, por lo que una vez concluido el cálculo de las matrices de proyección de cada vista, se cuenta con el modelo geométrico de ellas.

3.3. Clasificador en Múltiples Vistas

El diseño de un clasificador en múltiples vistas utiliza de forma simultánea la información adquirida en la Sección 3.1 como en la Sección 3.2.

Cada secuencia cuenta con al menos cuatro imágenes, cada imagen tiene identificado posibles gatillos. A su vez poseemos para cada vista la matriz de proyección que ésta posee. Con ambas informaciones iremos completando el clasificador en múltiples-vistas.

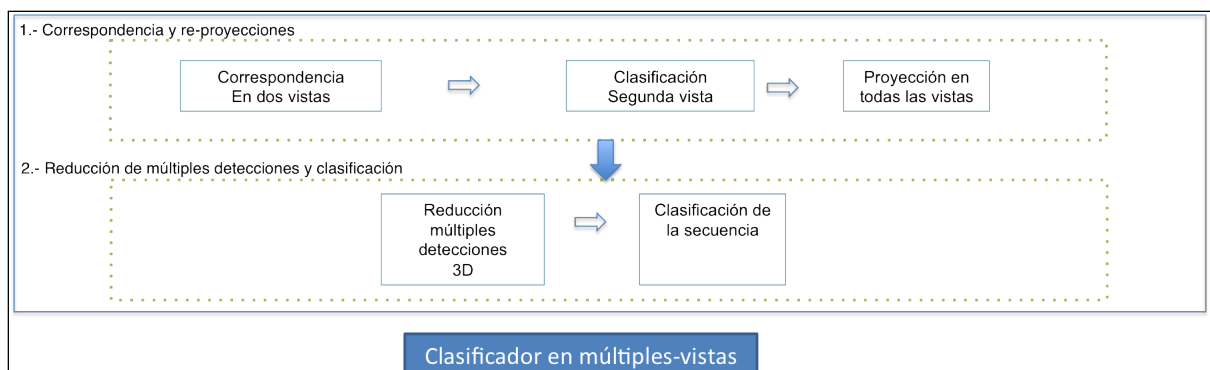


FIGURA 3.11. Clasificador en múltiples vistas

Como se muestra en la Figura 3.11 para el diseño del clasificador en múltiples-vistas se dividió el problema en dos grandes partes. La primera **correspondencia y re-proyecciones**, tiene relación con la correspondencia en dos vistas, proyectar los posibles gatillos en todas las vistas y su posterior validación. La segunda **reducción de múltiples detecciones y clasificación**, está relacionado con la representación por un solo punto de múltiples detecciones, para finalizar con una clasificación de la secuencia que completa su validación.

3.3.1. Correspondencia y Re proyecciones

Correspondencia

Se buscó correspondencia de los gatillos pre-seleccionados en el resto de las imágenes que componen la secuencia.

Para determinar esta correspondencia se realizaron cinco pasos (Fig. 3.12).

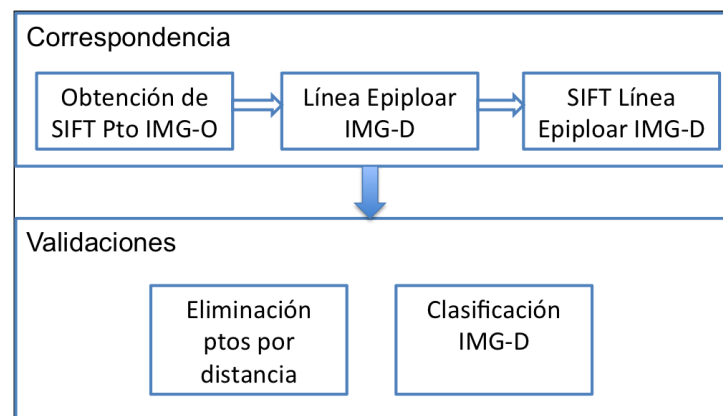


FIGURA 3.12. Diagrama de bloques para establecer correspondencia entre dos vistas.

La Figura 3.12 muestra que para cada detección de gatillo se buscó el descriptor SIFT del punto central, esta ventana fue llamada $IMG - O$ como se muestra en la Figura 3.13a.

Con la información de las matrices de proyección se calculó la línea epipolar en la imagen siguiente de la secuencia, que fue llamada $IMG - D$. Se extrajeron descriptores SIFT en una región de 10 píxeles de ancho centrada en la línea epipolar de $IMG - D$ como se muestra en la Figura 3.13b. Se realizó una comparación de todos los descriptores de $IMG - D$ con

el descriptor de $IMG - O$ quedando pre-seleccionado el punto que presenta menor distancia euclidiana.

Validaciones

Se procedió a realizar validaciones, ya que existe la posibilidad de que el punto pre-seleccionado no sea correspondiente:

1. Si la distancia entre los descriptores SIFT del punto proyectado en $IMG - D$ con el punto origen $IMG - D$ es menor a λ será seleccionado, si es mayor, será descartado.
2. Si la ventana del punto seleccionado en $IMG - D$ es clasificada como gatillo será seleccionado, si no, se descarta.

Aplicando estas validaciones se descartaron todos los puntos que no contaron con un grado de similitud cromática y textural, o que no se encuentren en el espacio correspondiente a la geometría epipolar (Fig. 3.13c).

El mismo procedimiento se aplicó a todas las imágenes de la secuencia de forma iterativa para las vistas [1 2][2 3][3 4][4 1].

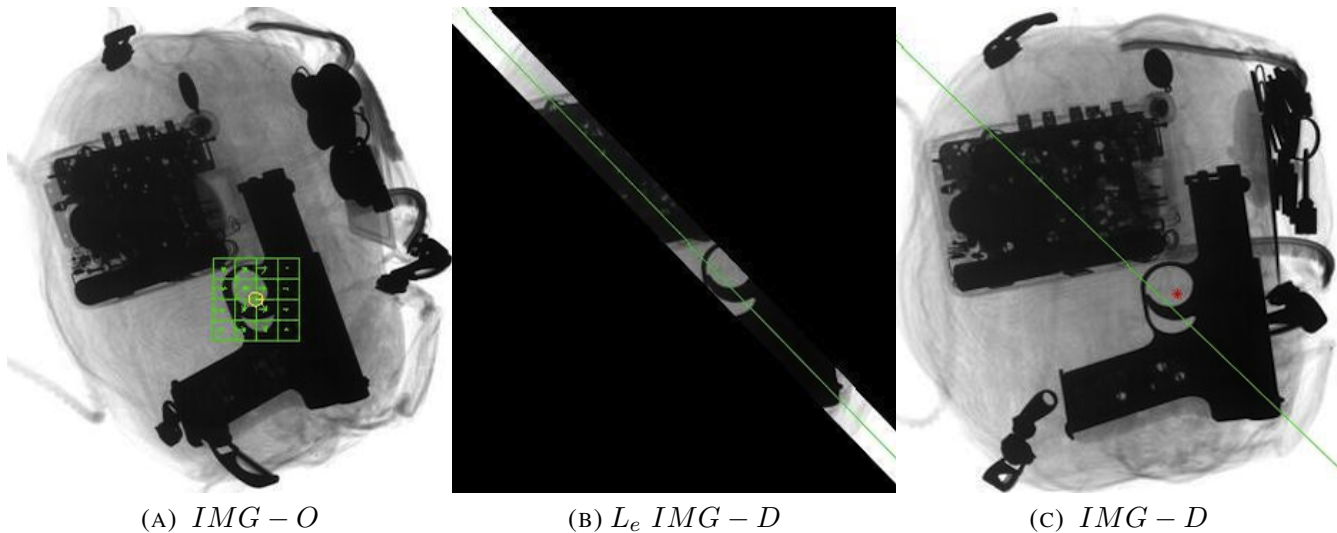


FIGURA 3.13. Correspondencia entre dos imágenes

Como se muestra en la Figura 3.13 el punto en la imagen origen $IMG - O$ es correspondido por un punto que está en la vecindad de la línea epipolar l_e de la imagen destino $IMG - D$.

Todos los puntos que no encontraron un punto correspondiente fueron descartados como puntos pre-seleccionados. De esta forma todos los puntos seleccionados tienen una re-proyección en dos vistas.

Re-Proyección de Puntos en Todas las Vistas

Como se plantea en la Sección 2.2.2 mediante los tensores trifocales es posible re-proyectar un punto en una tercera vista a partir de dos vistas. Como ya se definió la correspondencia de puntos en dos vistas se utilizarán los tensores trifocales para proyectar los puntos en todas las vistas.

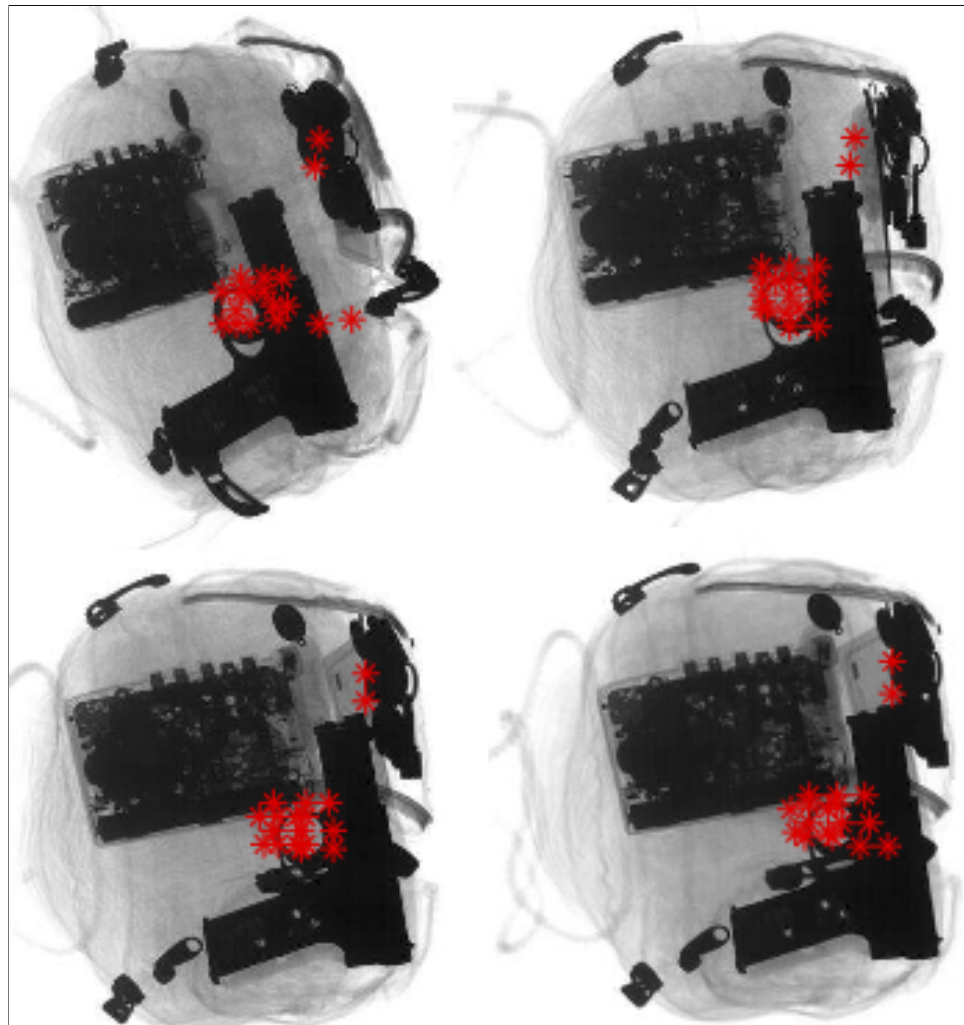


FIGURA 3.14. Puntos proyectados en todas las vistas

En la Figura 3.14 se muestran múltiples detecciones, esto es debido, a que en cada vista cada punto detectado como posible gatillo fue re-proyectado en todas las vistas.

3.3.2. Reducción de Múltiples Detecciones y Clasificación

Con las re-proyecciones de los puntos sobre todas las vistas se producen dos efectos no deseados. Como se puede apreciar en la Figura 3.14 se encuentran múltiples detecciones, además un falso positivo.

Para eliminar las múltiples detecciones, cada punto fue expresado en su coordenada 3D. Se buscó un punto capaz de representar a todos los puntos dentro de un rango o *voxelbox*. Esto se realiza creando clusters o conglomerados entre puntos que cumplen cierta similitud, en este caso se buscó similitud por distancia, es decir, que los puntos se encontraran dentro de una distancia en el espacio 3D, el criterio para determinar la pertinencia de un punto M_i a un cierto conglomerado o cluster fue el propuesto por Cheng (1995) con el algoritmo Meanshift. Meanshift entrega los puntos representativos de las múltiples detecciones M_r expresando las coordenadas del punto como $M_r = [x, y, z]$. En la Figura 3.15 los puntos desplegados con azul

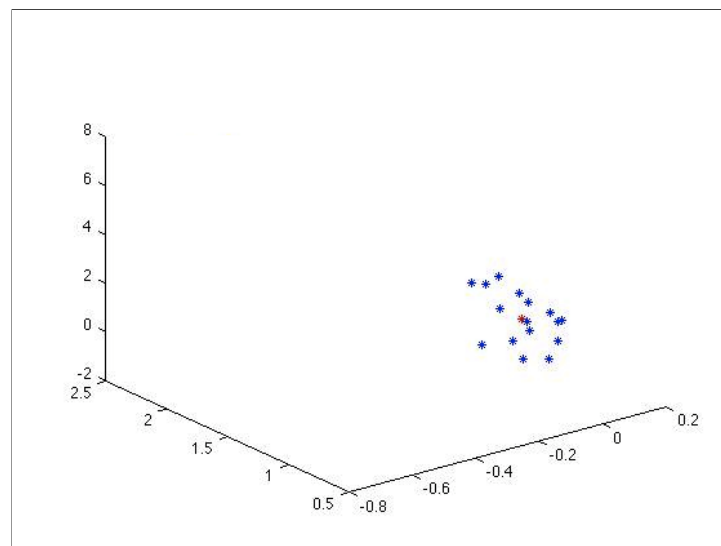


FIGURA 3.15. MeanShift para múltiples detecciones

son las múltiples detecciones, y el punto rojo es el punto representativo obtenido con meanshift.

Por último para eliminar falsos positivos se proyecta el punto M_r en cada vista y cada proyección es clasificada como se describe en la Sección 3.1.6. Para que la secuencia sea seleccionada tendrá que ser clasificada en por lo menos 2 proyecciones del punto M_r como gatillo, de lo contrario el punto M_r será descartado.

Para evitar que M_r sea un falso positivo se realizaron pruebas con un clasificador diferente al utilizado en el detector de gatillo mono ocular.

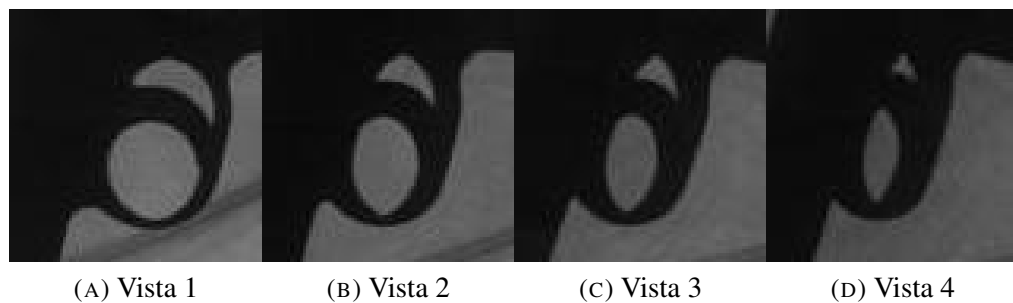


FIGURA 3.16. Proyecciones del punto M_r

En la Figura 3.16 se despliegan las proyecciones en todas las vistas del punto representativo que semuestra en rojo en la Figura 3.15.

4. EXPERIMENTOS Y RESULTADOS

Materiales y Toolboxes Para la investigación se confeccionaron programas computacionales los cuales tienen la función de procesar las imágenes, analizarlas y entregar resultados en la detección y clasificación de gatillos. La programación se realizó en Matlab (MathWorks, 2010) y se utilizaron los siguientes toolboxes disponibles para el desarrollo de investigaciones

- VIFeat (Vedaldi y Fulkerson, 2008)
- Balu (Mery, 2011b)
- Piotr's Image & Video Matlab Toolbox (Dollár, s.f.)

Pistolas para pruebas Las armas utilizadas en el presente trabajo fueron 5 pistolas de diferentes modelos. Se trabajó con armas de fuego, de fogueo e imitaciones de armas de fuego normales. Las pistolas utilizadas fueron:

- Beretta 92 en calibre 9 mm F92
- Glock calibre 9 mm
- Retay Falcon 9mm
- Blow Class 9 mm
- Mauser 25 mm



(A) Beretta



(B) Glock



(C) Retay Falcon



(D) Blow Class

FIGURA 4.1. Modelos de pistolas

Las pistolas que son parte de la muestra son armas utilizadas por la industria cinematográfica chilena¹.

Máquina de Rayos X La imágenes fueron capturadas en el Laboratorio de Robótica de la Pontificia Universidad Católica con una máquina de rayos X Cannon CXDI-50G. La máquina

¹Las armas fueron proporcionadas por la productora GRILLOFILMS S.A.

utiliza el método de Flat panel detector: scintillator & Amorphous Silicon (a-Si), y tiene un sensor LANMIT 4 (Large Area New-MIS sensor and TFT). Para la adquisición de Rayos X se utilizó una técnica de 3.6 mA 60 kV. Cada imagen tiene píxeles con 4096 valores en escala de grises (12 bit), con una resolución de 5,9 Mp por imagen. *Máquina Fotográfica*

Se utilizó una máquina fotográfica para la adquisición de secuencias fotográficas, esta fue una Cannon Reflex, EO5 de resolución 6.3 Mp por imagen.

4.1. Secuencias

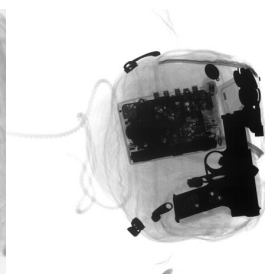
Como se explicó en la Sección 4 se utilizaron diferentes modelos de pistolas en la adquisición de secuencias, estas fueron cambiando en la medida que se adquiría una nueva secuencia.

Para la construcción de las secuencias se organizó el material de tal forma que simulara una situación lo más real posible. De esta manera las imágenes adquiridas presentaron diferentes niveles de dificultad. Se construyeron secuencias de imágenes de rayos X con 4 imágenes cada una y los niveles de dificultad se dividieron en:

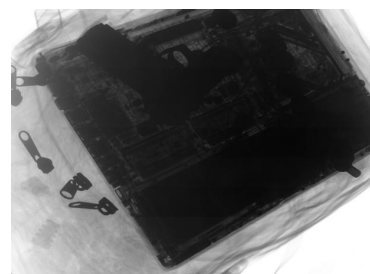
- Bajo (Fig. 4.2a).
- Media (Fig. 4.2b).
- Alta (Fig. 4.2c).



(A) Clase de dificultad baja



(B) Clase de dificultad media



(C) Clase de dificultad alta

FIGURA 4.2. Categorías de dificultades en las secuencias

En la Figura 4.2 se muestran los tres niveles de dificultad.

Se incorporaron elementos que produjesen oclusiones como superposiciones con otros objetos buscando que la imagen resultante fuera una mezcla de diferentes coeficientes de atenuación, de esta forma se intentó dotar de un sentido de realidad a las muestras.



FIGURA 4.3. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia con dificultad baja

Las secuencias con dificultad baja (Fig. 4.3) tienen un buen nivel de información ya que cada imagen cuenta con varios objetos en ella y estos se encuentran ocluidos en no más de un 10 %. En estas secuencias se procuró dejar el arma separada del resto de los objetos y estos no tienen forma similar a la del gatillo.

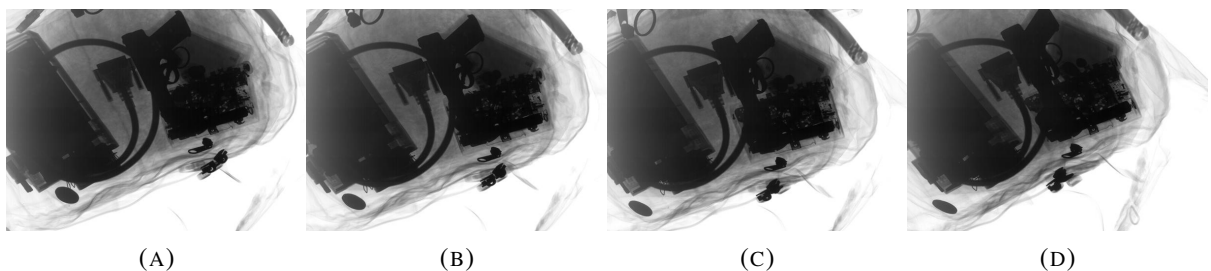


FIGURA 4.4. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia con dificultad media, imágenes con cables y chips

Las secuencias con dificultad media (Fig. 4.4) tienen oclusiones sobre el gatillo. El tipo de oclusión que se produce en estas secuencias fueron a partir de cables metálicos. Además estas secuencias contaron con imágenes de bolsos pequeños por lo que poseen pocos elementos al interior lo que dificulta encontrar puntos correspondientes y con ello las matrices de proyección de las vistas.

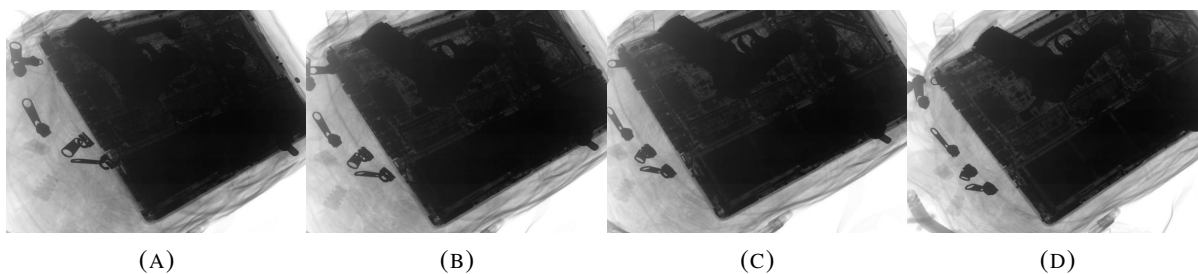


FIGURA 4.5. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia con dificultad alta, imágenes superpuestas sobre un laptop

Las secuencias con dificultad alta (Fig. 4.5) se encuentran ocluidas, además cuentan con altos niveles de ruido². Esta secuencias contaron con oclusiones completas del gatillo producidas por un laptop y por una cámara fotográfica.

Secuencias de control para validación del método con Fotografías:

Se consideró tomar tres secuencias de fotografías de las armas solas para validar el método propuesto.



FIGURA 4.6. Diferentes imágenes desplazadas y rotadas pertenecientes a una secuencia de arma sola

Las secuencias de las armas solas (Fig. 4.6) no contaron con ningún nivel de dificultad. Se adquirieron las imágenes con movimientos de la cámara de forma manual.

El total de las secuencias adquiridas con radiografías fue de 27. Las secuencias contaron con diferentes niveles de dificultad. Además se tomaron tres secuencias de fotografías, lo que dio un total de 30 secuencias para la muestra.

²Los niveles de ruido y dificultad en imágenes donde se encuentra superposición con laptops ha determinado que en la generalidad de los aeropuertos en el mundo obliguen a las personas a sacarlos de sus bolsos del momento de realizar las inspecciones.

TABLA 4.1. Distribución de la muestra de secuencias por nivel de dificultad

Nivel de dificultad	Número de muestras
Control (fotografías)	3
Simple	6
Mediano 1	1
Mediano 2	4
Mediano 3	6
Difícil 1	2
Difícil 2	8

Como se muestra en la Tabla 4.1, la distribución de la muestra quedó homogénea. Las secuencias de fotografías fueron consideradas en la clase de secuencias con dificultad baja.

Se sub-dividió la muestra en siete grupos, quedando los niveles de dificultad específicos para identificar precisamente lo que se logra clasificar³.

4.2. Segmentación

Se realizaron pruebas con las diferentes técnicas de segmentación expuestas en la Sección 3.1.4. En todas las técnicas se fue modificando un parámetro para evaluar el comportamiento de la segmentación. El parámetro se fue modificado en el intervalo $\lambda \in \{0, \dots, 0,9\}$.

Se construyeron imágenes con el resultado de la segmentación. Recordemos que la segmentación opera a nivel local en la ventana, por lo que, en base a la concatenación de todos las ventanas (*Sliding Windows*) se construyó una imagen resultante.

Como se muestra en las Figuras 4.7a, 4.7b la imagen final es una concatenación de cada ventana a nivel local. Se puede distinguir como cada ventana fue traslapada según su desplazamiento. Al final se obtiene una imagen que resulta ser una imagen distorsionada de la original, pero donde se distingue el gatillo cuando éste logra ser segmentado.

En las zonas centrales de las Figuras 4.7a, 4.7b se puede distinguir los gatillos de las diferentes imágenes. En la Figura 4.7a el gatillo se encuentra rotado en 170°. En la Figura 4.7b el gatillo se encuentra en el centro de la imagen.

³Para tener un mayor detalle visual revisar el Anexo 5.1.

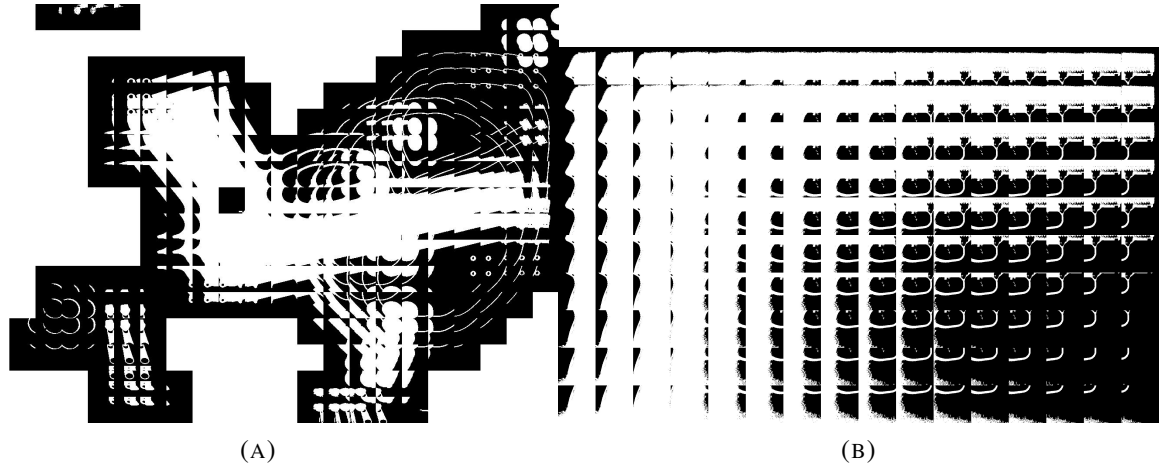


FIGURA 4.7. Imagen generada para comprobación de la segmentación

Para seleccionar la técnica de segmentación se realizaron pruebas cambiando el segmentador y el parámetro λ , después de la evaluación final se tabularon los resultados. El criterio para determinar si una técnica logró segmentar el gatillo se realizó de manera visual en las imágenes resultantes. Se etiquetó con 1 la secuencia si en todas las imágenes se logró identificar el gatillo, si no, la secuencia fue etiquetada con un 0.

$$\text{Evaluación de la técnica de segmentación} = \sum I_i \quad (4.1)$$

$$I_i = \begin{cases} 1 & \text{si se identificó} \\ 0 & \text{en otro caso} \end{cases}$$

Como se expresa en la Formula 4.1 la evaluación de la técnica de segmentación fue el acumulado de las detecciones.

Cada secuencia fue evaluada con los 4 segmentadores y a su vez con cada valor del parámetro λ , lo que da un total de $9 * 27 = 234$ evaluaciones.

En la Tabla 4.2 la primera columna muestra el parámetro λ . Las columnas siguientes muestran la cantidad de veces que fue identificado el gatillo con el segmentador (Umbral, Umbral promedio, Otsu, LUT) en el total de 27 secuencias.

TABLA 4.2. Cantidad de identificaciones de gatillos: Métodos de segmentación y λ

λ	Umbral	Umbral promedio 2	Otsu	LUT + Umbral
0,1	14	1	18	18
0,2	13	1	18	21
0,3	11	1	5	21
0,4	6	1	0	21
0,5	3	0	0	21
0,6	0	0	0	20
0,7	0	0	0	20
0,8	0	0	0	20
0,9	0	0	0	21

El total de secuencias evaluadas fue de 27 ya que se eliminaron las secuencias de fotografías dada su simpleza.

La Tabla 4.2 muestra que la mejor técnica de segmentación es la que utiliza LUT. Esta técnica presenta la condición de hacer un ajuste del histograma a nivel local lo que facilita la segmentación de objetos cuando se encuentran ocluidos. Al aumentar el contraste a nivel local es posible separar objetos superpuestos. El parámetro modificado no fue significativo frente a esta técnica de segmentación y λ se dejó fijo en 0,9. La Tabla 4.2 muestra que el segmentador en base a LUT con el parámetro de 0,9 logró detectar 21 secuencias de un total de 27, lo que da un 78 % sobre las secuencias adquiridas.

4.3. Selección de Características

Del total de 44 características geométricas que se obtuvieron se seleccionaron las mejores mediante *Sequential Forward Selection* SFS descrito en la Sección 2.1.3. Se buscó la menor cantidad de características iterando y agregando entre 1 y 17.

Como se muestra en la Figura 4.8 se tabuló la cantidad de características con su rendimiento.

La Figura 4.8 muestra cómo la cantidad de características fue alterando el rendimiento del clasificador. Fueron 11 características las seleccionadas ya que fue la menor cantidad que logró el mayor desempeño (98 %) ⁴.

⁴Notar que con solo cuatro características se obtiene sobre un 95 %.

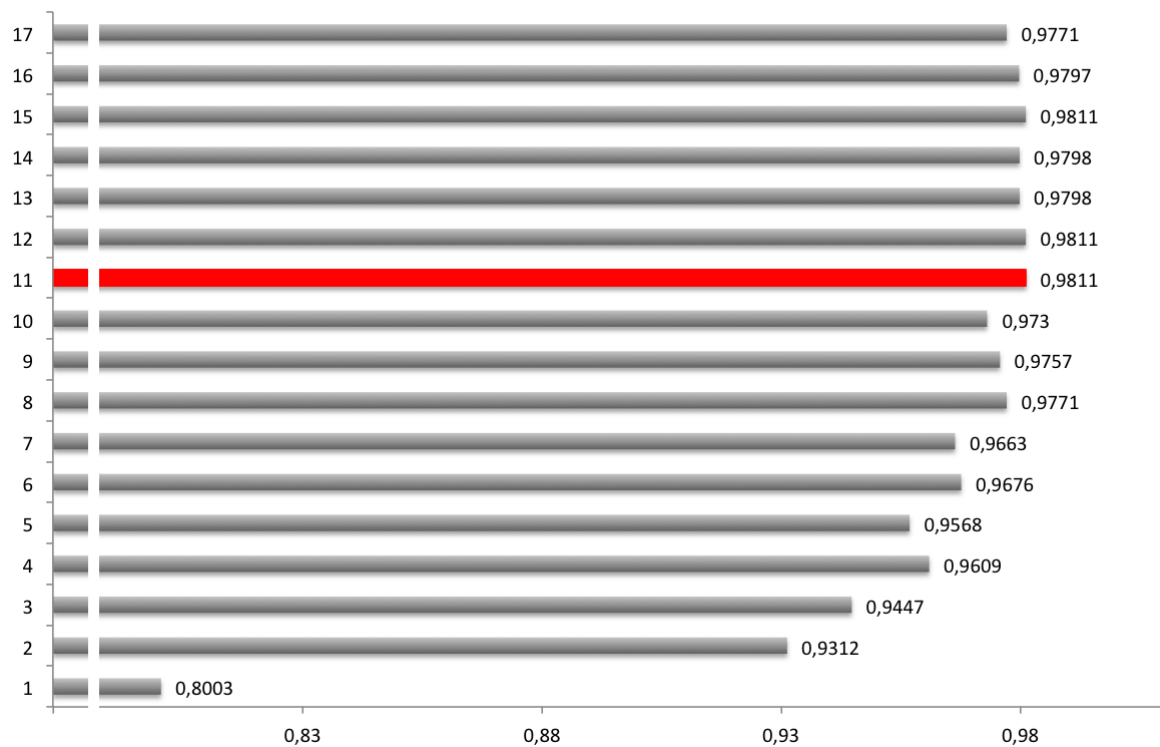


FIGURA 4.8. Desempeño de clasificador v/s cantidad de características, con rojo la cantidad de características seleccionada

Las características seleccionadas son:

TABLA 4.3. Características seleccionadas

Número	Característica
1	Roundness
2	Filled Area [px]
3	Area [px]
4	Danielsson factor
5	Euler Number
6	Fourier-des 4
7	Equivalent Diameter [px]
8	Fourier-des 8
9	Minor Axis Length [px]
10	Hu-moment 1
11	Major Axis Length [px]

1. La mejor característica seleccionada fue la redondez. La redondez es una medida de la forma y está definida en el rango $[0-1]$. Pensando en un gatillo esta característica hace sentido en la medida que el gatillo como su contexto es redondeado 4.9c.
2. La segunda y tercera característica están relacionadas con el área. Hay que recordar que la clasificación se realiza en la ventana deslizante, por lo que es consistente que las relaciones de las áreas estén correlacionadas con el gatillo.
3. La cuarta característica es la distancia de Danielsson para describir la forma de un objeto (P. Danielsson, 1980)
4. La quinta característica es el número de Euler, y representa el número de objetos menos el número de huecos en una región. Como se ve en la Figura 4.9 los gatillos tienen siempre una cantidad de huecos definido: 1.

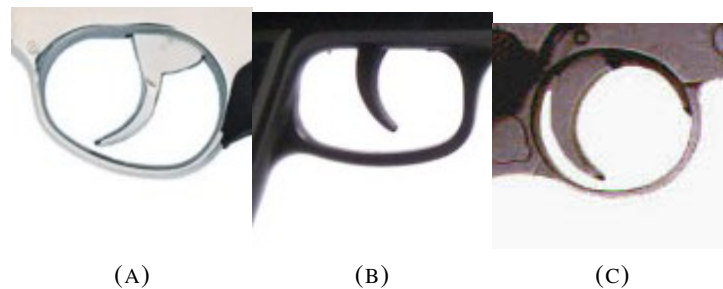


FIGURA 4.9. Muestra de gatillos

4.4. Clasificador

Se evaluaron diferentes clasificadores para determinar cuál tiene mejor rendimiento. Los clasificadores evaluados fueron LDA-Fisher, distancia Mahalanobis, QDA, SVM (Bishop, 2006).

Las pruebas se validaron con el método descrito en la Sección 2.1.4 y el análisis de los resultados se realizó sobre tablas de confusión para cada clasificador.

Los resultados fueron:

La Tabla 4.4 muestra en la primera columna el método de clasificación, en la segunda columna los TP (verdaderos positivos), en la tercera columna TN (verdaderos no detectados),

TABLA 4.4. Matrices de confusión y performance

Método	TP	FN	FP	TN	p	$ci95\%$
LDA	632	3	9	97	0.9797 %	[0.9703 0.9892]
MAHA	631	4	9	97	0.9838 %	[0.9621 0.9839]
QDA	632	3	18	88	0.9797 %	[0.9702 0.9892]
SVM	623	12	19	87	0.9582 %	[0.9421 0.9639]

en la cuarta columna FP (falsos positivos), en la quinta columna FN (verdaderos negativos), en la sexta y séptima columna se muestra el performance como el intervalo de confianza.

Como muestra la Tabla 4.4 de matrices de confusión el clasificador Mahalanobis presenta mayor desempeño con 0.98 % con 9 falsos negativos y 4 falsos positivos. LDA presenta un desempeño de 0.97 % con 9 falsos negativos con 3 falsos positivos. Se busca encontrar la mayor cantidad de gatillos con la menor cantidad de falsos positivos, por lo que se determinó que el clasificador LDA es el indicado a utilizar en la fase de detección.

4.5. Estimación de Matrices de Proyección

Debido a que las matrices de proyección son obtenidas a partir de *Bundle Adjustment* y este método utiliza puntos correspondientes entre las vistas de forma aleatoria no entrega el mismo modelo cada vez que se aplica. Se realizó una prueba para verificar el cambio en el cálculo de las matrices de proyección. La prueba consistió en calcular las matrices de proyección de una secuencia diez veces. La primera iteración se eligió manualmente un punto en la imagen 1 y su correspondiente en la imagen 2 los que se dejaron constantes para todas las iteraciones. A esos puntos se calculó sus re-proyecciones en las tercera y cuarta vista para cada iteración mediante el modelo geométrico adquirido con *Bundle Adjustment*. Se tabuló la desviación estándar de los puntos en el eje X como en el eje Y medidos en pixeles.

TABLA 4.5. Desviación de puntos re-proyectados en pixeles

vista	std(x)	std(y)
3	20,52	39,17
4	6,96	16,54

La Tabla 4.5 muestra las desviaciones para las vistas 3 y 4 que son significativas. Las desviaciones estándar de los puntos para las vistas 1 y 2 son 0 ya que no se re-proyectan.

Como el modelo geométrico calculado en cada ocasión entrega valores diferentes se realizaron todos los experimentos de detección 4 veces por lo que la muestra aumenta de 30 secuencias a 120 secuencias.

4.6. Reducción de puntos

Como se menciona en la Sección 3.3.2 para eliminar las múltiples detecciones se realizan clusters con meanshift buscando puntos representativos. Se realizaron pruebas buscando puntos representativos por cada imagen (2D), como buscando los puntos representativos en el modelo geométrico (3D). Al hacer la búsqueda de puntos representativos por cada imagen en 2D dio que la cantidad de puntos por cada imagen no eran correspondientes para todas las imágenes, por lo que había que volver a buscar las correspondencia entre ellos. Además el rendimiento final de la detección bajó entre un 10 % y 20 %.

4.7. Configuración de Clasificador en Múltiples Vistas.

Para el diseño de un clasificador basado en múltiples-vistas se proponen varias etapas que podrían ser reemplazadas dependiendo del objetivo buscado. El diseño del detector mono ocular considera una etapa segmentación y otra de clasificación. Como se explicó en la Sección 4.7b se puede construir un segmentador ad-hoc específico, lo mismo sucede con las características y con el clasificador utilizado. Gran parte del diseño de un clasificador en múltiples-vista consiste en diseñar el segmentador como los clasificadores que proporcionan un óptimo rendimiento a nivel mono ocular.

Los clasificadores que se podrían diseñar de forma independiente dentro de la propuesta metodología son:

- El clasificador utilizado en el detector de gatillos.
- El clasificador utilizado en la clasificación final de la secuencia.

Dentro de los experimentos, se realizaron pruebas con distintos clasificadores organizados de forma diferente entre las etapas. Se realizaron comparaciones de dos configuraciones y además se compararon con una clasificación mono ocular.

La definición de las configuraciones fue:

- Configuración 1: Clasificador LDA en detector de gatillos y en verificación Mahalanobis (*LDA-Maha*).
- Configuración 2: Clasificador LDA (alto rendimiento) en detector de gatillos y en verificación el mismo (*LDA-LDA*).

Se realizó una comparación del desempeño en las secuencias con las diferentes configuraciones, además se comparó con una clasificación que se realizó de manera mono ocular.

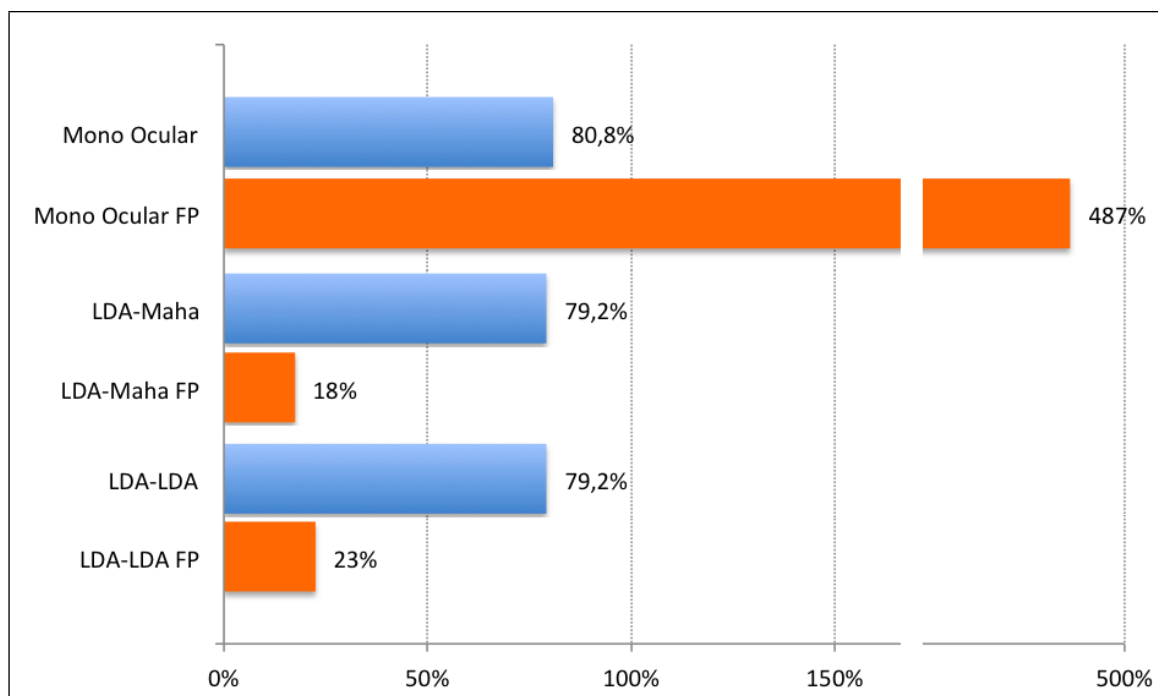


FIGURA 4.10. Comparaciones de cada configuración, con color azul el porcentaje de desempeño, con color naranja el porcentaje de falsos positivos

En el gráfico con color azul (Fig. 4.10) se muestra el mismo rendimiento (verdaderos positivos) de los clasificadores múltiples vistas y estos a su vez son levemente inferior al clasificador

mono ocular. En el gráfico con color naranja (Fig. 4.10) muestra el porcentaje de falsos positivos de los clasificadores. La Figura 4.10 muestra cómo el clasificador mono ocular presenta una elevada cantidad de falsos positivos, sin embargo los clasificadores en múltiples-vistas logran disminuirlos sustantivamente.

La configuración *LDA – Mahalanobis* fue la que presentó la menor cantidad de falsos positivos.

4.8. Comparación de detecciones con diferentes niveles de dificultad.

Se realizó una comparación de los clasificadores en múltiples-vistas Configuración 1 *LDA-Maha* y Configuración 2 *LDA-LDA* según los nivel de dificultad que presentaron las secuencias: En el cuadro 4.6 se muestra que la configuración *LDA-Maha* tiene un mayor desempeño que

TABLA 4.6. Clasificaciones en múltiples-vistas por nivel de dificultad, verdaderos positivos

Configuración	<i>LDA-Maha</i>	<i>LDA-LDA</i>
<i>Dificultad</i>	<i>TP %</i>	<i>TP %</i>
<i>Control</i>	100 %	100 %
<i>Baja</i>	100 %	100 %
<i>Media</i>	100 %	100 %
<i>Media</i>	100 %	93,75 %
<i>Media</i>	79,5 %	58,33 %
<i>Alta (cámara)</i>	37,5 %	87,5 %
<i>Alta (laptop)</i>	53,13 %	59,3 %

la configuración *LDA-LDA* exceptuando las secuencias con mayor nivel de dificultad.

TABLA 4.7. Falsos positivos por nivel de dificultad

Configuración	<i>LDA-Maha</i>	<i>LDA-LDA</i>
<i>Dificultad</i>	<i>FP %</i>	<i>FP %</i>
<i>Control</i>	0 %	0 %
<i>Baja</i>	0 %	4,16 %
<i>Media</i>	0 %	0 %
<i>Media</i>	0 %	25 %
<i>Media</i>	0 %	25 %
<i>Alta (cámara)</i>	12,5 %	50 %
<i>Alta (laptop)</i>	62,5 %	37,5 %

En la Tabla 4.7 se muestra que la configuración *LDA-Maha* presenta falsos positivos en las secuencias con mayor nivel de dificultad y no presenta falsos positivos en secuencias con dificultad baja o media. La configuración *LDA-LDA* presenta falsos positivos en la mayoría de las secuencias.

La configuración *LDA – Maha* fue la que presentó la menor cantidad de falsos positivos y mejor rendimiento, sin embargo, la configuración *LDA-LDA* tiene un rendimiento mejor para secuencias con niveles de complejidad y altos niveles de superposición de objetos.

Por último se muestran cuatro Figuras (Fig. 4.11, Fig. 4.12, Fig. 4.13 y Fig. 4.14) donde se muestran detecciones realizadas en secuencias con nivel de dificultad bajo, medio y alto.

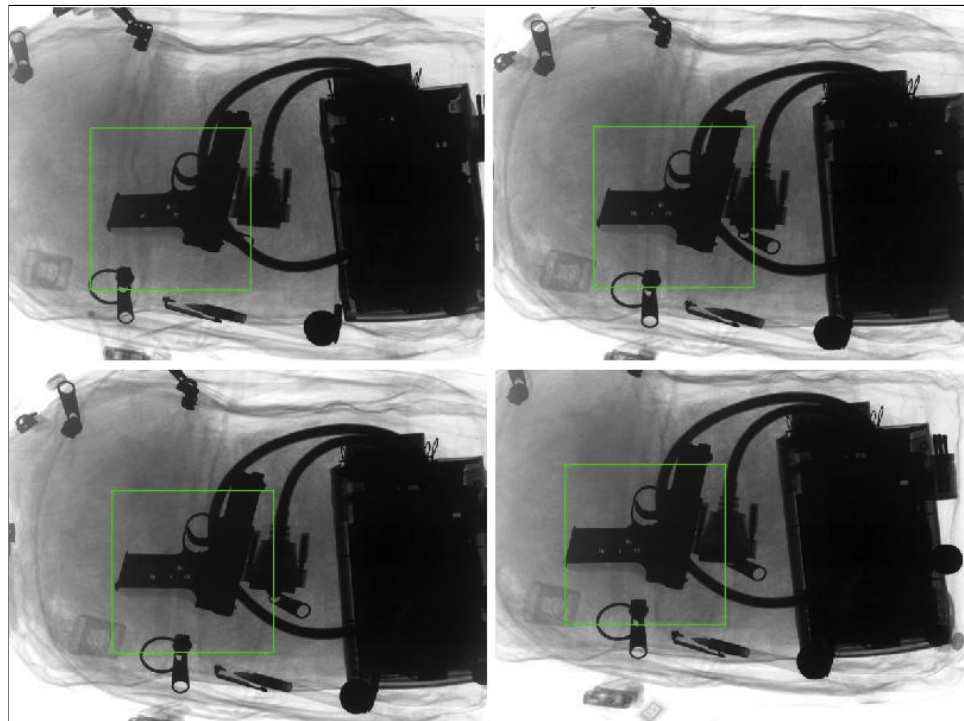


FIGURA 4.11. Detección en secuencia dificultada baja

Para tener mayor detalle de las clasificaciones de las secuencias podrá revisar la Sección 5.1.

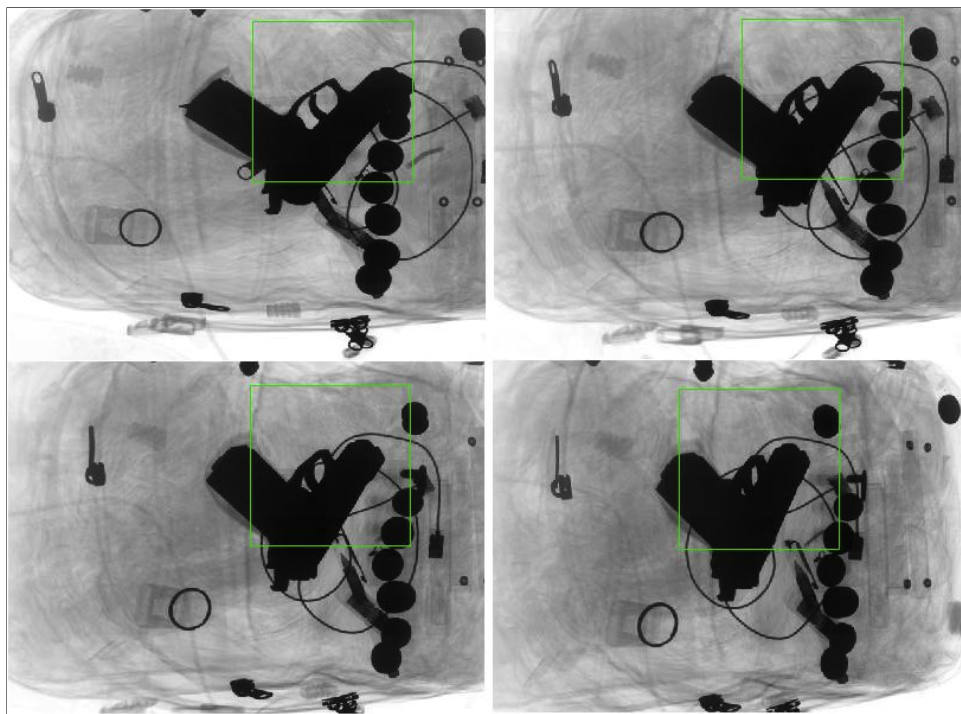


FIGURA 4.12. Detección en secuencia dificultada media

4.8.1. Comparación con Método Alternativo de Tracking entre Vistas

Se probó el método propuesto por Mery (2011a) con las secuencias utilizadas en la presente investigación para realizar una validación del método propuesto en la tesis. Se utilizó la misma técnicas de segmentación y los mismos clasificadores en esta prueba que los utilizados para la obtención de resultados de la investigación. Es esta prueba se logró un 43 % de detecciones:

TABLA 4.8. Detalle del desempeño método (Mery, 2011a)

<i>Dificultad</i>	Detecciones TP %	FP %
<i>Control</i>	0 %	0 %
<i>Baja</i>	33,3 %	16,6 %
<i>Media</i>	100 %	0 %
<i>Media</i>	60 %	60 %
<i>Media</i>	25 %	25 %
<i>Alta (cámara)</i>	100 %	100 %
<i>Alta (laptop)</i>	50 %	75 %

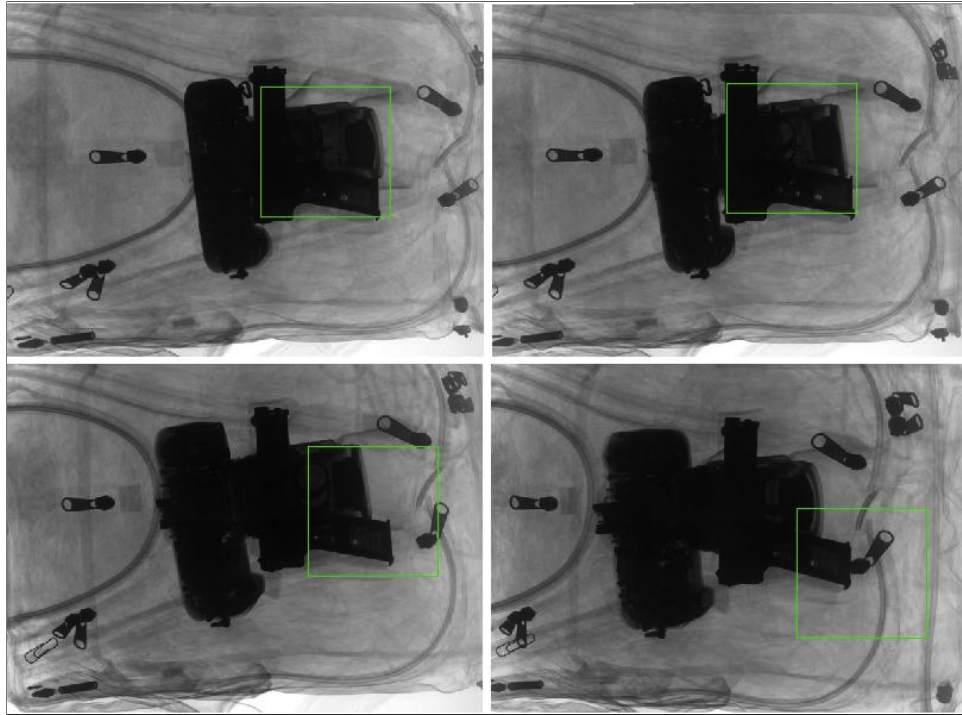


FIGURA 4.13. Gatillo ocluido por una cámara

La Tabla 4.8 muestra el desempeño del método propuesto en (Mery, 2011a). Un 56,6 % de las secuencias no logró obtener resultados debido a que este método no logra encontrar correspondencia entre los candidatos de gatillos en la diferentes vistas y justamente este es un punto que ambos métodos presentan una diferencia. El método de la presente investigación logra mejores resultados para un seguimiento o tracking específico.

4.8.2. Múltiples Detecciones

La cantidad de puntos detectados como gatillos fue variando según la etapa del clasificador múltiples-vistas. Los puntos iniciales fueron encontrados por el detector de gatillos mono ocular, a estos se les buscó correspondencia en dos vistas y se fueron descartando los que no fueron correspondidos. Por último se re-proyectaron los puntos en las 4 vistas.

En la Figura 4.15 se muestra el gráfico del promedio de puntos detectados en las 120 secuencias para cada imagen en cada etapa. Se puede apreciar cómo las cantidad de puntos

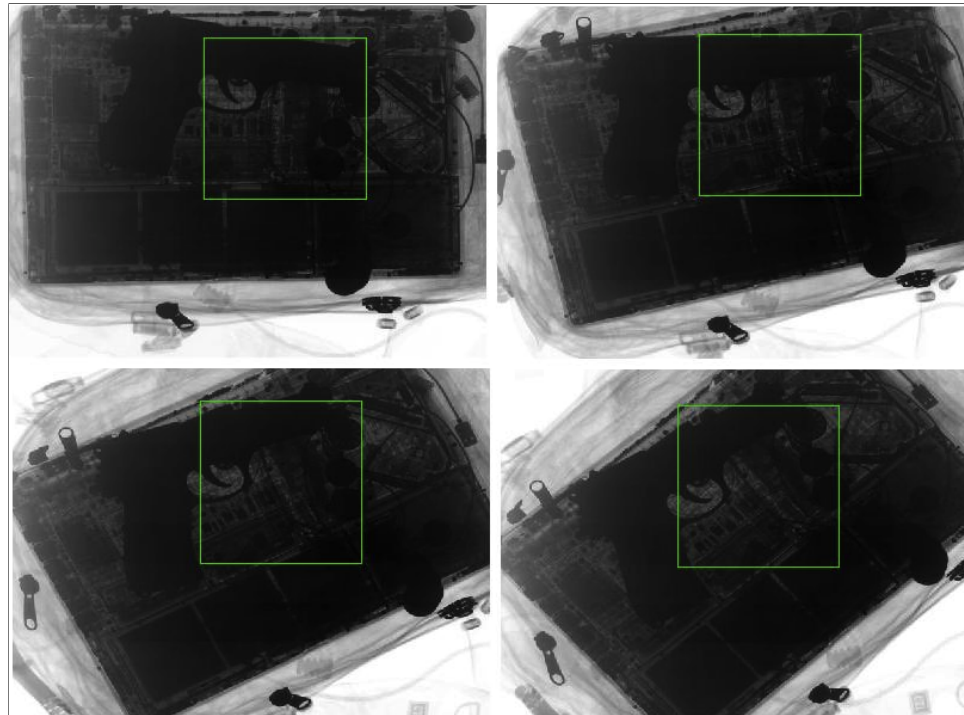


FIGURA 4.14. Gatillo ocluido con un laptop

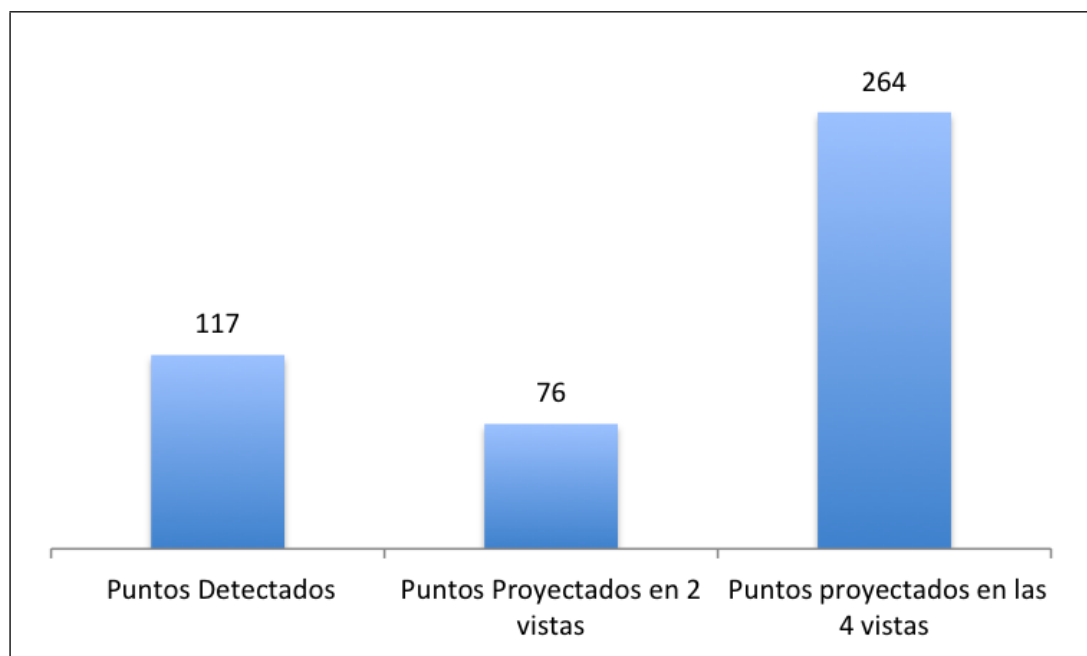


FIGURA 4.15. Cantidad de puntos seleccionados por etapa.

proyectados en las cuatro vistas no son cuatro veces los puntos iniciales. Esto es debido a que se fueron realizando validaciones y descartando falsos positivos.

4.8.3. Mediciones de Tiempo por Etapas

Se midió el tiempo de ejecución del clasificador múltiples-vistas en un Macbook 2.4 Ghz Core Duo, con Matlab R2010a (64-bit), para cada etapa de una secuencia, después se calculó el promedio de todas:

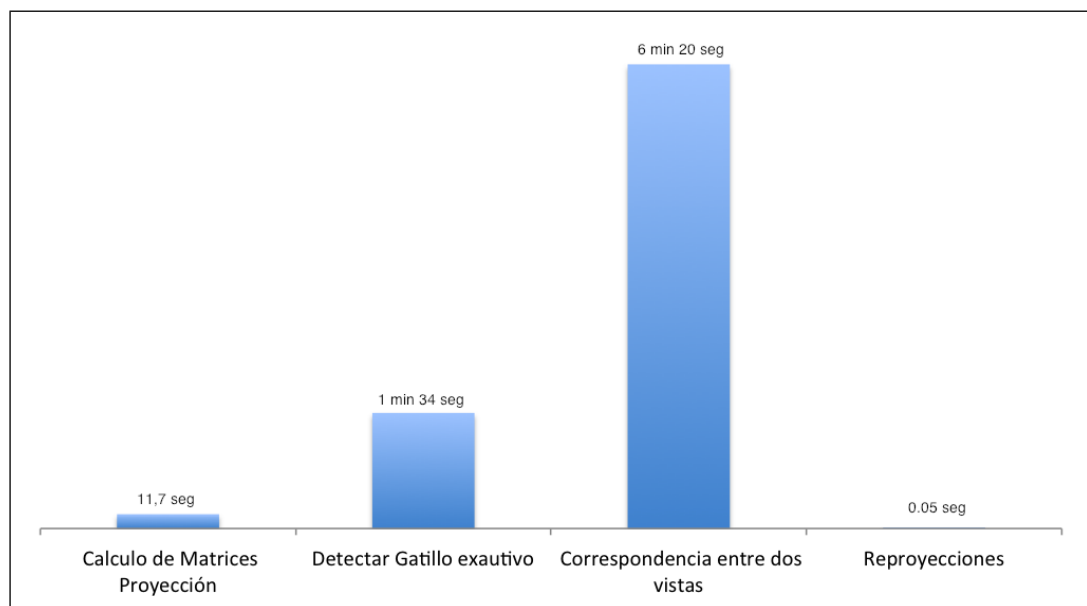


FIGURA 4.16. Tiempos de ejecución por etapas.

Como se puede apreciar en la Figura 4.16 la correspondencia entre dos vistas es lo que demoró mayor cantidad de tiempo. Esto es debido a que por cada punto se calculó la línea epipolar y luego SIFT para hacer comparaciones. En la medida que el detector de gatillos obtenga una mayor cantidad de puntos, el tiempo involucrado para el cálculo de la correspondencia será mayor.

5. CONCLUSIONES

El aporte de esta investigación es una respuesta a los problemas de automatización y de asistencia en las inspecciones de objetos mediante radiografías. Se desarrolló una metodología, la cual, al incorporar múltiples vistas obtiene en promedio una efectividad de detecciones de 81,4 % de gatillos al interior de bolsos con 3 diferentes niveles de dificultad: 100 % para secuencias simples, un 93 % para secuencias de mediana complejidad y un 73,4 % para secuencias alta dificultad. Las secuencias simples y medianas no presentaron falsos positivos, las secuencias con mayor dificultad (oclusiones por laptop) llegaron a tener un 65 % de falsos positivos. El detector mono-ocular logró un 80 % de rendimiento, sin embargo, presentó una tasa de falsos positivos en el orden de 487 %.

El desempeño de las configuraciones *LDA-Maha* y *LDA-LDA* muestra que la primera configuración presenta mejor rendimiento para secuencias simples y medianas, sin embargo, la segunda presenta mejor desempeño para las secuencias con mayor oclusión. En un ambiente productivo sería de interés contar con una aplicación que considere varias configuraciones y parámetros, cuestión de tener mayor información al momento de tomar una decisión o poder elegir diferentes modos de detección.

La metodología desarrollada puede desagregarse constituyendo cada parte del diseño como subproductos que pueden ser interpretados como aportes específicos en cada tema implicado. Resaltaremos a continuación tres dimensiones en las que es posible concluir aportes a la discusión científica en la materia: Clasificador mono-ocular, Estimación geométrica y Clasificador múltiples vistas.

En relación al clasificador mono-ocular se pueden concluir aspectos relacionados con dos temas centrales en la materia: segmentación y características.

- ▷ En cuanto al problema de la segmentación se observa que es necesario considerar al menos dos fases. La primera fase es de normalización de las imágenes: donde es necesario realizar una corrección global, aumentar el contraste, eliminar el ruido y las deformaciones producidas por la máquina que adquiere las imágenes de rayos X. La segunda es la fase de segmentación propiamente tal; al respecto se probaron y

compararon métodos de segmentación que utilizan parámetros y técnicas que operan a nivel global de la imagen, con métodos de segmentación que utilizan parámetros y técnicas que operan a nivel local. Los resultados fueron concluyentes respecto a la eficacia de las técnicas que utilizan parámetros a nivel local. Esto se explica debido que al existir oclusiones o superposición de objetos, la segmentación local logra encontrar umbrales que separan los objetos al aumentar el contraste localmente. Las técnicas que operan a nivel local logran mejores resultados que las técnicas globales, sin embargo, presentan una desventaja ya que requieren una mayor cantidad de tiempo al procesar cada ventana deslizante. Por lo general, las metodologías que trabajan con ventanas deslizantes realizan una transformación a la imagen integral lo que logra disminuir el tiempo de procesamiento.

La metodología propuesta en el presente trabajo no requiere una fase de segmentación robusta ya que al contar con múltiples vistas y seguimiento entre ellas se reduce la responsabilidad por cada vista. El problema de segmentación en las imágenes de rayos X logra encontrar un camino alternativo de resolución, pues, el método descansa fundamentalmente en el aporte que las múltiples vistas hacen a la detección en la imagen.

- ▷ Respecto a las características, se puede concluir que el conjunto de características geométricas seleccionadas en el diseño del clasificador son importantes y afectan directamente la eficacia de la aplicación. En este caso, las características fueron:
 1. La redondez: a simple vista se puede concluir que los gatillos se pueden representar de buena manera mediante su redondez.
 2. Relacionadas con el área: la segunda y tercera característica seleccionada en esta aplicación están relacionadas con el área, por lo que resulta evidente que un factor a controlar sería tamaño de la ventana, si este cambia, estas características tendrían que ser re-definidas.
 3. Factor de Danielsson y el número de Euler: que son la cuarta y la quinta características seleccionadas cuya función es representar una figura y representar el

número de huecos en la imagen. Ambas se entienden como pertinentes a la hora de representar un gatillo.

El diseño de un clasificador mono-ocular tiene que considerar los elementos planteados que hacen dificultoso el procesamiento de las imágenes de rayos X. Al dividir la responsabilidad realizando detecciones en varias vistas se limita la dependencia del detector mono-ocular como único agente de clasificación.

La implementación en un ambiente productivo tiene que considerar que existen características que no son invariantes al tamaño, lo que implica que el tamaño de la ventana, la distancia entre el tubo y el Flat Panel, como el tamaño de los gatillos tienen que ser factores que se encuentren controlados al momento del diseño del clasificador.

Respecto a la estimación geométrica, a partir de las pruebas realizadas se puede concluir que para la toma de secuencias la técnica *SfM Structure from Motion* es adecuada para trabajar con múltiples vistas. Además de efectiva es eficiente, pues mediante la implementación de esta técnica se logra utilizar una sola máquina, lo que reduce los costos en Hardware. Su limitación se encuentra en que los objetos no se pueden mover por lo que sería interesante pensar un diseño donde la máquina de rayos X sea quien gire en torno al objeto. Así mismo, la utilización de descriptores característicos SIFT permiten realizar correspondencias entre vistas de forma automática, sin embargo, si las imágenes presentan niveles de homogeneidad la correspondencia presentará problemas, lo que implicaría que el modelo geométrico calculado podría quedar defectuoso. A partir de los experimentos se puede concluir que el problema de calcular el modelo geométrico de forma defectuosa se presenta fundamentalmente en las secuencia con poca información (en particular las secuencias del bolso de cintura).

Hay que considerar que los resultado obtenidos podrían presentar mejoras si se usara un sistema calibrado. Si este fuera el caso, se tendría que prestar atención en el diseño de una metodología de calibración simple de implementar en un ambiente productivo.

Sobre el clasificador en múltiples vistas es posible concluir de manera general que la metodología propuesta de seguimiento entre vistas integrando información geométrica con información cromática, posibilita múltiples opciones para diseñar clasificadores de objetos internos.

Incluir múltiples vistas proporciona información suficiente para realizar detecciones de objetos internos y permite descartar un alto porcentaje de falsos positivos. La utilización de clasificadores diferentes para las distintas etapas, posibilita descartar una mayor cantidad de falsos positivos.

Como conclusión relacionada a las múltiples detecciones se logró encontrar puntos representativos mediante la inclusión de información espacial. Ello puede ser de alta relevancia para el diseño de futuras aplicaciones que busquen optimizar la detección de diversos objetos mediante la inclusión de tecnología basada en el análisis de imágenes de rayos X.

5.1. Trabajos Futuros

A partir de los experimentos y conclusiones obtenidas se han identificado algunos aspectos posibles de ser profundizados, para ello será necesario realizar algunas pruebas en relación a:

- Validar el método propuesto con una mayor cantidad de armas y diferentes niveles de oclusión. Dado que un civil en Chile tiene dificultades para el acceso a armas, sería importante desarrollar vías institucionales que favorezcan el acceso de forma legal.
- Validar el método propuesto realizando pruebas en detección de gatillos con armas que se encuentren desarmadas donde sus partes y piezas estén repartidas dentro de un contenedor.
- Validar el método propuesto en detección de armas en una variedad mayor de armas incorporando armas de asalto como rifles, fusiles, sub-ametralladoras, etc.
- Realizar pruebas y comparaciones del rendimiento reemplazando la auto-calibración geométrica por un método calibrado.
- Realizar pruebas con la imagen integral de las imágenes de rayos X para medir velocidad de procesamiento en la detección de gatillos en la fase mono-ocular.
- Disminuir los errores producidos en la estimación geométrica obtenida mediante *Bundle adjustment*.

BIBLIOGRAFIA

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)*.

Berman, E. G. (2011). *Small arms survey 2010 small arms survey 2010 small arms survey 2010 small arms survey 2010 the small arms survey 2011: States of security* (K. Krause, Ed.). Cambridge University Press.

Bishop, C. M. (2006). *Pattern recognition and Machine Learning*. Springer.

Carrasco, M., y Mery, D. (2007). Automatic multiple visual inspection on non-calibrated image sequence with intermediate classifier block. En *Proceedings of the 2nd pacific rim conference on advances in image and video technology* (pp. 371–384). Berlin, Heidelberg: Springer-Verlag. Disponible en <http://portal.acm.org/citation.cfm?id=1782074.1782113>

Chan, J., Evans, P., y Wang, X. (2010, oct.). 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).

Cheng, Y. (1995, agosto). Mean shift, mode seeking, and clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(8), 790–799. Disponible en <http://dx.doi.org/10.1109/34.400568>

Dalal, N., y Triggs, B. (2005). Histograms of oriented gradients for human detection. *Conference on Computer Vision and Pattern Recognition (CVPR2005)*, 1, 886–893.

Danielsson, P. (1980). Euclidean distance mapping. *Computer Graphics and Image Processing*, 14(3), 227–248. Disponible en <http://linkinghub.elsevier.com/retrieve/pii/0146664X80900544>

Danielsson, P. E. (1978, abril). New shape factor. , 7(2), 292–299.

Dollár, P. (s.f.). *Piotr's Image and Video Matlab Toolbox (PMT)*. <http://vision.ucsd.edu/~pdollar/toolbox/doc/index.html>.

Dornaika, F., y Chung, R. (2001). An algebraic approach to camera self-calibration. *Computer Vision and Image Understanding*, 83(3), 195 - 215. Disponible en <http://www.sciencedirect.com/science/article/pii/S1077314201909257>

Duan, X. (2009). X-ray cargo container inspection system with few-view projection imaging. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 598(2), 439-444. Disponible en http://xml.engineeringvillage2.org/controller/servlet/Controller?CID=expertSearchDetailedFormat&EISESSION=1_29c58e131d4707354M4494ses4&SYSTEM_USE_SESSION_PARAM=true&SEARCHID=acdd02131b4ebca327917prod4data1&DOCINDEX=1&PAGEINDEX=1&RESULTSCOUNT=1&database=1&format=expertSearchDetailedFormat (ID: EVII; ID: Engineering Village 2)

Duda, R., Hart, P., y Stork, D. (2001). *Pattern classification* (2.^a ed.). New York: John Wiley & Sons, Inc.

Duin, A. K. J. F. I. R. P., y Mao, J. (2000). Statistical pattern recognition: A review. *IEEE*.

Evans, J., Robinson, M., y Godber, S. (1998, jul). Pseudo-tomographic x-ray imaging for use in aviation security. *Aerospace and Electronic Systems Magazine, IEEE*, 13(7), 25 -30.

Fischler, M. A., y Bolles, R. C. (1981). Paradigm for model. *Communications of the ACM*, 24(6), 381–395. Disponible en <http://portal.acm.org/citation.cfm?id=358669.358692>

- Fitzgibbon, A., Pilu, M., y Fisher, R. (1999). Direct least square fitting ellipses. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(5), 476-480.
- Ginebra, S. de la Declaración de. (2011). *Global burden of armed violence 2011*. Cambridge University Press.
- GT, Z., y RZ, R. (1972). Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, C-21(3), 269 - 281.
- Harris, C., y Stephens, M. (1988). A combined corner and edge detector. En *Proc. 4th alvey vision conferences* (p. 147-152).
- Hartley, R. I. (1997). *Lines and point in three views and the trifocal tensor*.
- Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory*, IT(8), 179-187.
- Hussein E.M.A. Waller, E. (1998). Review of one-side approaches to radiographic imaging for detection of explosives and narcotics. *Radiation Measurements*, 29(6), 581-591. Disponible en <http://www.scopus.com/inward/record.url?eid=2-s2.0-0032289373&partnerID=40&md5=455fe1aabedb7be44fb332a0287904fc> (cited By (since 1996) 21)
- Internacional, A. (2008). Datos que matan.
- Jebara, T., Azarbayejani, A., y Pentland, A. (1999, <http://es.scribd.com/doc/36732609/SfM-Pentland>). 3d structure from 2d motion. *IEEE signal processing Magazine*.
- Leiva, G., Mondragon, G., Mery, D., y Aguilera, J. (2011). The automatic sorting using image processing improves postharvest blueberries storage quality. En *Proceedings of international congress on engineering and food (icef-2011)*. Disponible en <http://dl.dropbox.com/u/1791703/Prints/Conferences/International/2011-ICEFa.pdf>

- Liu, D., y Wang, Z. (2008, nov.). A united classification system of X-ray image based on fuzzy rule and neural networks. En *Intelligent system and knowledge engineering, 2008. iske 2008. 3rd international conference on* (Vol. 1, p. 717 -722).
- Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.
- Luong, Q.-T., y Faugeras, O. (1995). The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17, 43–75.
- Mahalanobis, P. C. (1936, 16 de abril). On the generalised distance in statistics. En *Proceedings national institute of science, india* (Vol. 2, pp. 49–55). Disponible en <http://ir.isical.ac.in/dspace/handle/1/1268>
- MathWorks. (2010). *Matlab image processing toolbox 7: User's guide*. The MathWorks Inc.
- Mery, D. (2002). New approaches for defect recognition with x-ray testing. *Insight*, 44(10), 614-615.
- Mery, D. (2004). *Vision por computador*. Santiago de Chile: Departamento de Ciencia de la Computacion, Pontificia Universidad Catolica de Chile. Disponible en <http://dl.dropbox.com/u/1791703/Prints/Books/2004-ApuntesVision.pdf>
- Mery, D. (2006a). Automated radiosopic testing of aluminum die castings. *Materials Evaluation*, 64(2), 135-143.
- Mery, D. (2006b). Extracción de características.
- Mery, D. (2011a). Automated detection in complex objects using a tracking algorithm in multiple x-ray views. *OTCBVS in Conjunction with Computer Vision and Pattern Recognition*.

Mery, D. (2011b). Balu.

Mery, D., y Filbert, D. (2002a). Automated flaw detection in aluminum castings based on the tracking of potential defects in a radiosopic image sequence. *IEEE Transactions on Robotics*, 18(6), 890-901. Disponible en <http://dl.dropbox.com/u/1791703/Prints/ISI-Journals/2002-IEEE.pdf>

Mery, D., y Filbert, D. (2002b, dec). Automated flaw detection in aluminum castings based on the tracking of potential defects in a radiosopic image sequence. *Robotics and Automation, IEEE Transactions on*, 18(6), 890 - 901.

Mery, D., Filbert, D., y Jaeger, T. (2005). Analytical characterization of aluminum and its alloys. En C. MacKenzie y G. Totten (Eds.), (p. 701-738). Florida: CRC Press, Taylor and Francis.

Mery, D., Jaeger, T., y Filbert, D. (2002). A review of methods for automated recognition of casting defects. *Insight*, 44(7), 428-436.

Mery, D., y Soto, A. (2008). Features: The more the better. En *The 7th wseas international conference on signal processing, computational geometry and artificial vision (iscgav-2008), rodos island, greece*. Disponible en <http://dl.dropbox.com/u/1791703/Prints/Conferences/International/2008-ISCGAV-TheMoreTheBetter.pdf>

Michel, S., Koller, S., Ruiter, J. de, Moerland, R., Hogervorst, M., y Schwaninger, A. (2007a, Oct.). Computer-based training increases efficiency in X-Ray image interpretation by aviation security screeners. En *Security technology, 2007 41st annual ieee international carahan conference on* (p. 201-206).

Michel, S., Koller, S., Ruiter, J. de, Moerland, R., Hogervorst, M., y Schwaninger, A. (2007b, oct.). Computer-based training increases efficiency in x-ray image interpretation by aviation security screeners. En *Security technology, 2007 41st annual ieee international carahan conference on* (p. 201 -206).

- Mondragon, G., Leiva, G., Aguilera, J., y Mery, D. (2011). Automated detection of softening and hard columella in kiwifruits during postharvest using x-ray testing. En *Proceedings of international congress on engineering and food (icef-2011)*. Disponible en <http://dl.dropbox.com/u/1791703/Prints/Conferences/International/2011-ICEFb.pdf>
- Murphy, E. (1989, nov). A rising war on terrorists. *Spectrum, IEEE*, 26(11), 33 -36.
- Murray, N., y Riordan, K. (1995, oct). Evaluation of automatic explosive detection systems. En *Security technology, 1995. proceedings. institute of electrical and electronics engineers 29th annual 1995 international carnahan conference on* (p. 175 -179).
- Nercessian, S., Panetta, K., y Agaian, S. (2008, july). A parametric method for edge detection based on recursive mean-separate image decomposition. En *Machine learning and cybernetics, 2008 international conference on* (Vol. 7, p. 3689 -3694).
- Oertel, C., y Bock, P. (2006, oct.). Identification of objects-of-interest in X-Ray images. En *Applied imagery and pattern recognition workshop, 2006. aipr 2006. 35th ieee* (p. 17).
- Otsu, N. (1979, enero). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1), 62-66.
- Persoon, E., y Fu, K. (1977). Shape discrimination using Fourier descriptors. *IEEE Trans. Systems, Man, and Cybernetics, SMC-7*(3), 170-179.
- Pieringer, C., y Mery, D. (2010). Flaw detection in aluminium die castings using simultaneous combination of multiple views. *Insight*.
- Richard Hartley, A. Z. (2003). *Multiple view geometry in computer vision*. Cambridge University Press.
- Riffo, V., y Mery, D. (2012). Active X-ray testing of complex objects. *Insight*. ((accepted Oct. 2011))

- Shashua, A., y Werman, M. (1995). Trilinearity of three perspective views and its associated tensor. En *In proceedings of the international conference on computer vision* (pp. 920–925).
- Silva, R., y D.Mery. (2007). State-of-the-art of weld seam inspection by radiographic testing: Part i - image processing. *Materials Evaluation*, 65(6), 643-647. Disponible en <http://dl.dropbox.com/u/1791703/Prints/ISI-Journals/2007-MatEval-Welding-1.pdf>
- Silva, R. R. da. (2010). *Radiografia industrial* (Vol. 1).
- Singh, M., y Singh, S. (2005, 31 2005-april 1). Optimizing image enhancement for screening luggage at airports. En *Computational intelligence for homeland security and personal safety, 2005. cihsp 2005. proceedings of the 2005 ieee international conference on* (p. 131 -136).
- Singh, S., y Singh, M. (2003). Explosives detection systems (eds) for aviation security. *Signal Processing*, 83(1), 31 - 55. Disponible en <http://www.sciencedirect.com/science/article/pii/S0165168402003912>
- Sivic, J., y Zisserman, A. (2009). Efficient visual search of videos cast as text retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), 591-605.
- Sonka, M., Hlavac, V., y Boyle, R. (1998). *Image processing, analysis, and machine vision* (2.^a ed.). Pacific Grove, CA: PWS Publishing.
- Strecker, H. (1998, jul.). Automatic detection of explosives in airline baggage using elastic X-ray scatter. En *Medicamundi* (Vol. 42, p. 30-33).
- Tuytelaars, H. B. T., y Gool., L. V. (2006). *Surf: Speeded up robust features*. In 9th European Conference on Computer Vision Graz Austria.
- Vedaldi, A., y Fulkerson, B. (2008). Vlfeat: An open and portable library of computer vision algorithm.

Viola, P., y Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. En *Proc. of ieee conf. on computer vision and pattern recognition (cvpr-01)* (p. 228-235).

Viola, P., y Jones, M. (2004). Robust real-time object detection. *International Journal of Computer Vision*, 57(2), 137-154.

Wales, A., Halbherr, T., y Schwaninger, A. (2009, oct.). Using speed measures to predict performance in X-ray luggage screening tasks. En *Security technology, 2009. 43rd annual 2009 international carnaham conference on* (p. 212 -215).

Wassenaar Arrangement, C. by the. (2012). Wassenaar arrangement on export controls for conventional arms and dual-use goods and technologies.

Webb, A. (2005). *Statistical pattern recognition*. England: Wiley.

Welling, M. (2009). Fisher linear discriminant analysis. *Science*, 1(2), 1-3. Disponible en <http://www.cs.huji.ac.il/~csip/Fisher-LDA.pdf>

Zentai, G. (2008). X-ray imaging for homeland security. *Imaging Systems and Techniques, 2008.IST 2008.IEEE International Workshop on*, 1-6. Disponible en <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4659929>

ANEXOS

Clasificación de cada secuencia utilizando la configuración *LDA-LDA*

Las Tablas presentadas en la presente sección muestran el rendimiento en la clasificación de cada secuencia. También se muestra una imagen de cada secuencia con la finalidad de proporcionar mayores antecedentes de los niveles de dificultad al lector. En la primera columna se especifica el nivel de dificultad evaluado entre $[0 \dots 7]$. La segunda columna especifica el número de la secuencia, la tercera y cuarta columna tiene los valores de las detecciones, la quinta y sexta columna los falsos positivos detectados. Por último la séptima columna contiene una imagen de muestra para apreciar el tipo de secuencia que se evaluó.

El modelo geométrico se calculó cuatro veces por lo que las detecciones se realizaron cuatro veces sobre una secuencia; de ahí que 4 detecciones son un 100 %, lo que significa que se logró detectar todas la veces que fue evaluada la secuencia.

TABLA 5.1. Rendimiento y falsos positivos de las secuencias de fotografías de control

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
0	1	4	100 %	0	0 %	

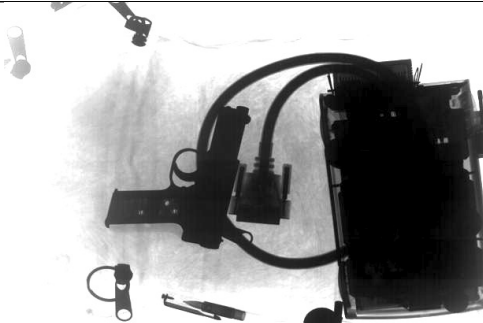
Continúa en la siguiente página. . .

Continuación de la Tabla 5.1

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
0	2	4	100 %	0	0 %	
0	3	4	100 %	0	0 %	
Fin Tabla 5.1						

Las secuencias expuestas en la Tabla 5.1 son imágenes fotográficas que se utilizaron como imágenes de control para la validación del método propuesto.

TABLA 5.2. Rendimiento y falsos positivos de las secuencias con dificultad baja

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
1	4	4	100 %	0	0 %	

Continúa en la siguiente página. . .

Continuación de la Tabla 5.2

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
1	5	4	100 %	1	25 %	
1	6	4	100 %	0	0 %	
1	7	4	100 %	0	0 %	

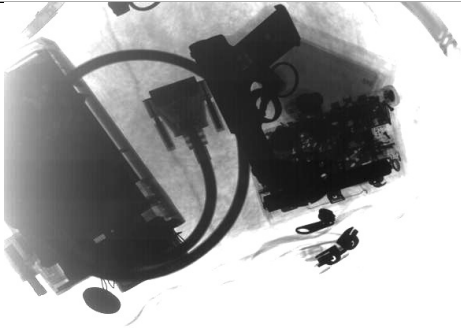
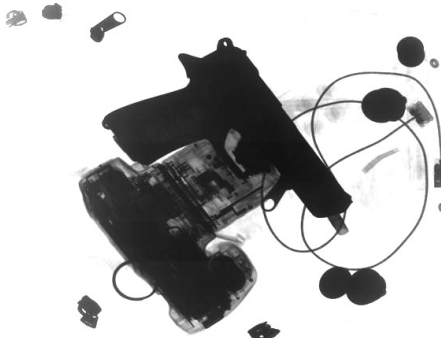
Continúa en la siguiente página. . .

Continuación de la Tabla 5.2

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
1	8	4	100 %	0	0 %	
1	9	4	100 %	0	0 %	
Fin Tabla 5.2						

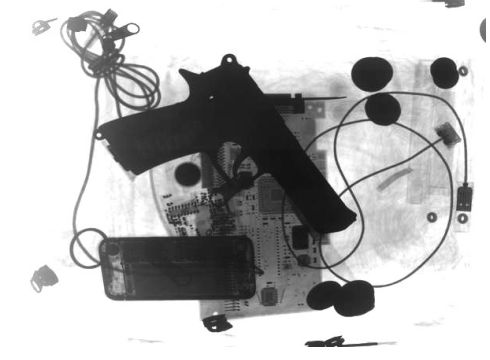
La secuencias de la Tabla 5.2 son secuencias que contienen un nivel de dificultad bajo.

TABLA 5.3. Rendimiento y falsos positivos de las secuencias con dificultad media (artefactos electrónicos)

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
2	10	4	100 %	0	0 %	
3	11	4	100 %	0	0 %	
3	12	1	25 %	0	0 %	

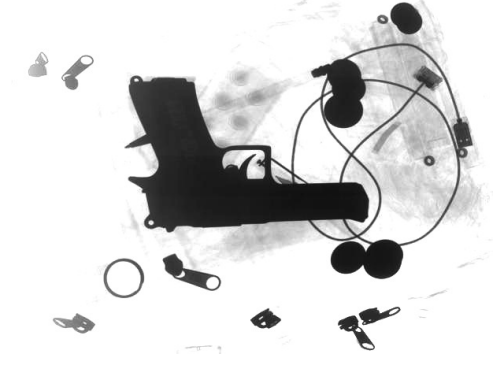
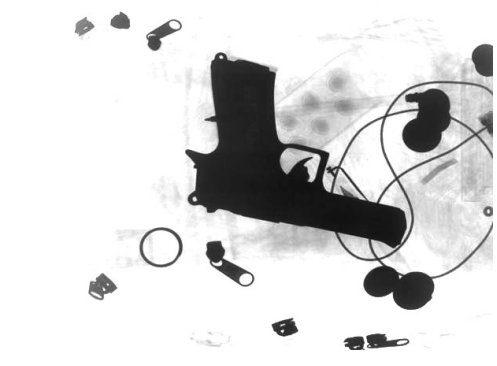
Continúa en la siguiente página. . .

Continuación de la Tabla 5.3

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
3	13	1	25 %	3	75 %	
3	14	4	100 %	1	25 %	
3	15	4	100 %	2	50 %	
Fin Tabla 5.3						

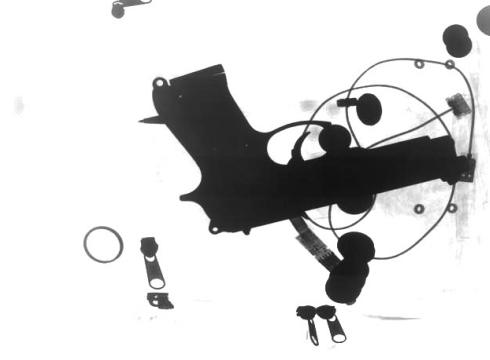
La secuencias de la Tabla 5.3 son secuencias que contienen un nivel de dificultad medio, en ellas se encuentran las pistolas con objetos electrónicos, además las mochilas tienen ropa y cables.

TABLA 5.4. Rendimiento y falsos positivos de las secuencias con dificultad media

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
4	16	4	100 %	0	0 %	
4	17	3	75 %	3	75 %	
4	18	4	100 %	1	25 %	

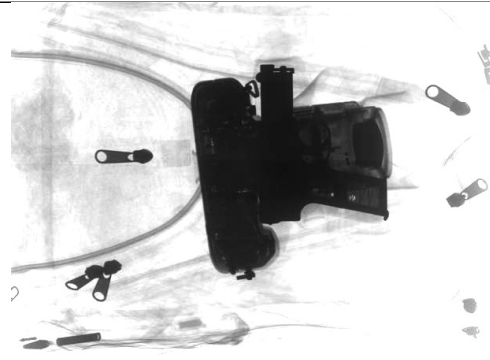
Continúa en la siguiente página. . .

Continuación de la Tabla 5.4

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
4	19	4	100 %	0	0 %	
Fin Tabla 5.4						

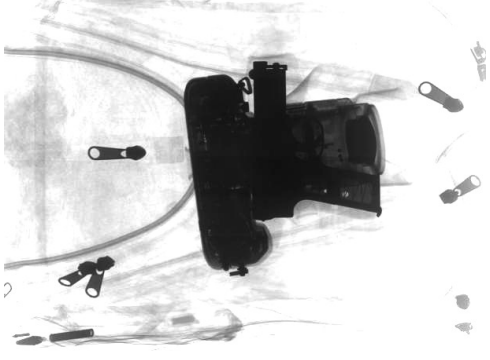
La secuencias de la Tabla 5.4 son secuencias que contienen un nivel de dificultad medio, en ellas se encuentran las pistolas las mochilas están llenas de ropa y cables.

TABLA 5.5. Rendimiento y falsos positivos de las secuencias dificultad alta (cámara)

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
5	20	3	75 %	0	0 %	

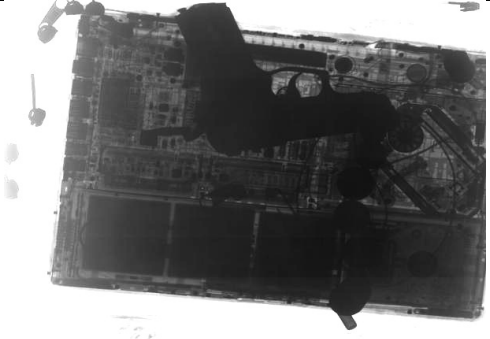
Continúa en la siguiente página. . .

Continuación de la Tabla 5.5

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
5	21	4	100 %	0	0 %	
Fin Tabla 5.5						

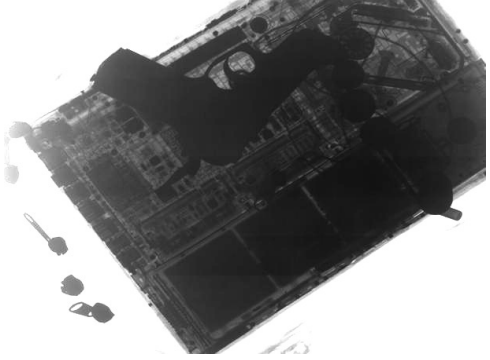
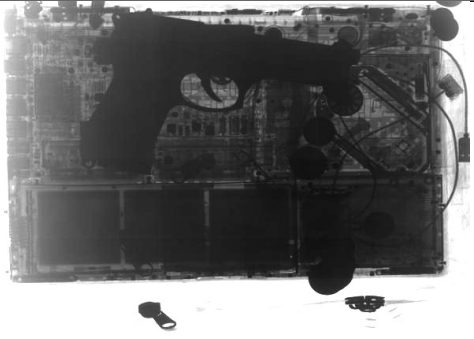
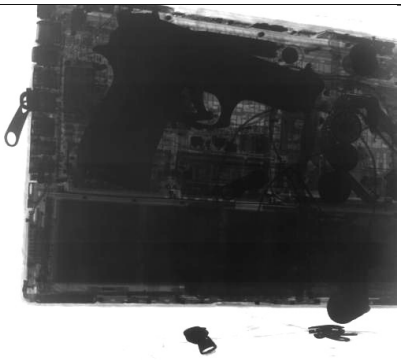
La secuencias de la Tabla 5.5 son secuencias que contienen un nivel de dificultad alto, en ellas se encuentran las pistolas superpuestas a una cámara fotográfica.

TABLA 5.6. Rendimiento y falsos positivos de las secuencias con dificultad alta (laptop)

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
6	22	0	0 %	4	100 %	

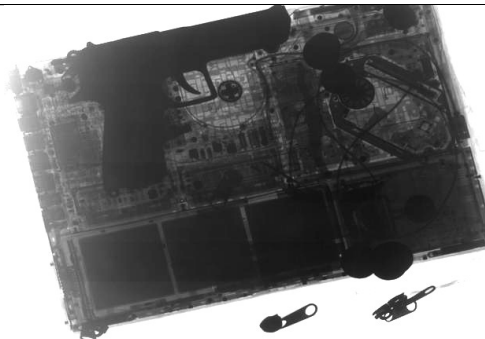
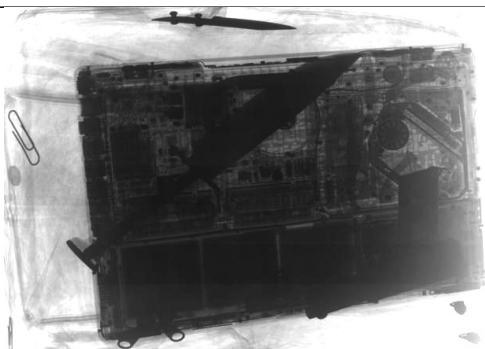
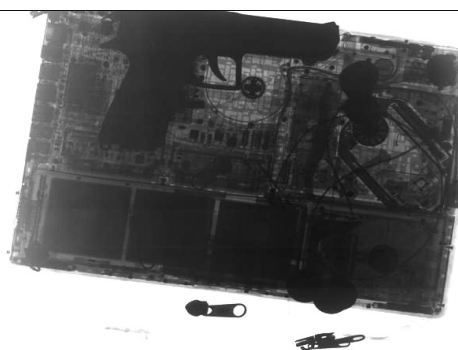
Continúa en la siguiente página. . .

Continuación de la Tabla 5.6

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
6	23	4	100 %	0	0 %	
6	24	4	100 %	0	0 %	
6	25	3	75 %	4	100 %	

Continúa en la siguiente página. . .

Continuación de la Tabla 5.6

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
6	26	0	0 %	4	100 %	
6	27	0	0 %	4	100 %	
6	28	4	100 %	0	0 %	

Continúa en la siguiente página. . .

Continuación de la Tabla 5.6

δ	ς	Detec.	% Detec.	FP	% FP	Imagen
6	29	4	100 %	0	0 %	
Fin Tabla 5.6						

La secuencias de la Tabla 5.5 son secuencias que contienen un nivel de dificultad alto, en ellas se encuentran las pistolas superpuestas a un laptop.

TABLA 5.7. Rendimiento y falsos positivos de las secuencia con problemas de estimación modelo geométrico

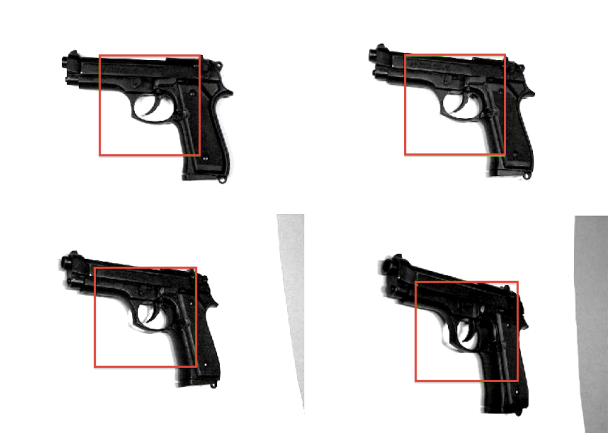
δ	ς	Detec.	% Detec.	FP	% FP	Imagen
7	30	0	0 %	0	0 %	
Fin Tabla 5.7						

La secuencia de la Tabla 5.7 es una secuencia de la cual no se logró estimar el modelo geométrico como detecciones sobre ella.

Detalle visual de las detecciones sobre las secuencias

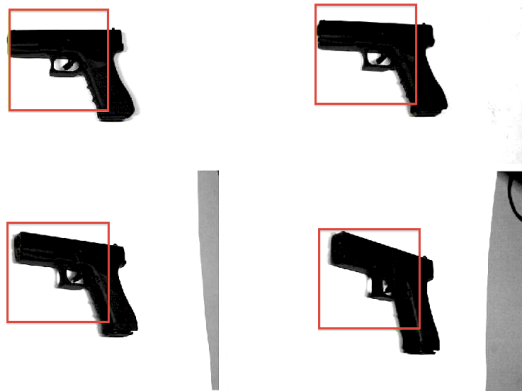
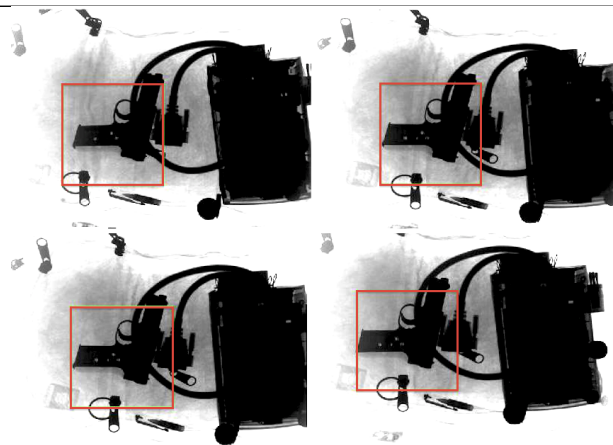
La siguiente Tabla muestra los resultados de las detecciones de gatillo realizadas sobre cada secuencia.

TABLA 5.8. Detecciones de las secuencias

δ	ς	Imagen
0	1	
0	2	

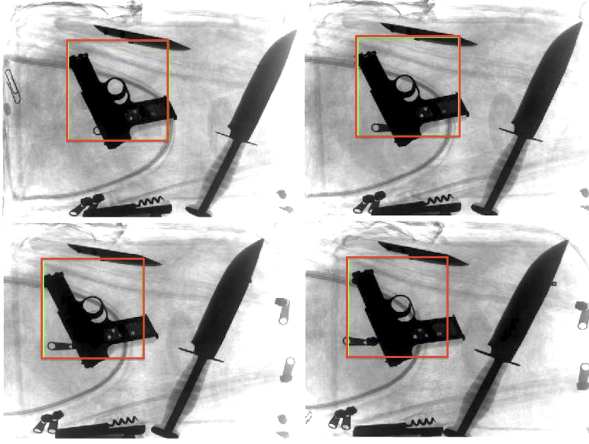
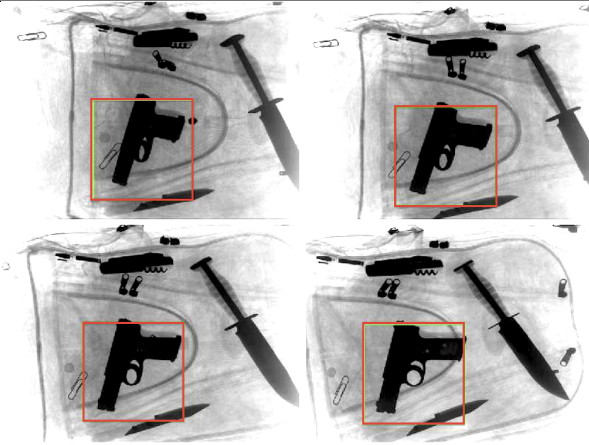
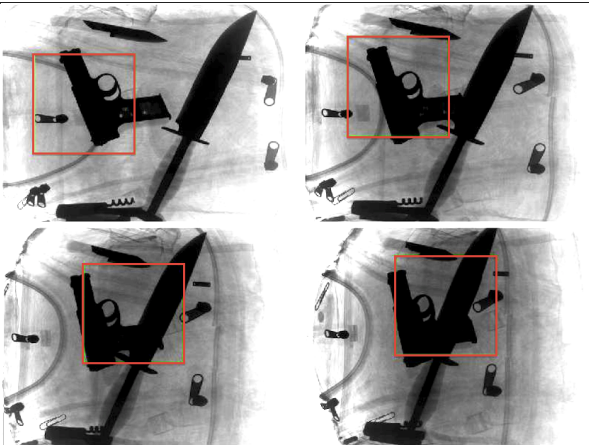
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
0	3	
1	4	
1	5	

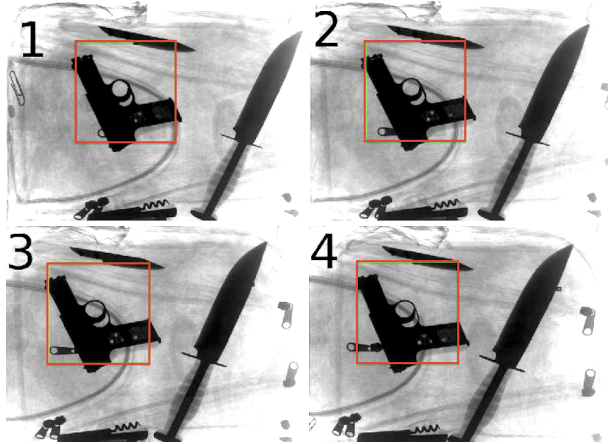
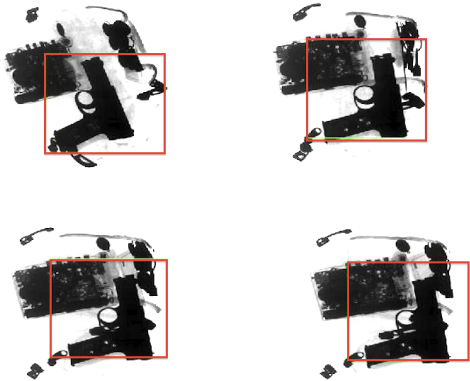
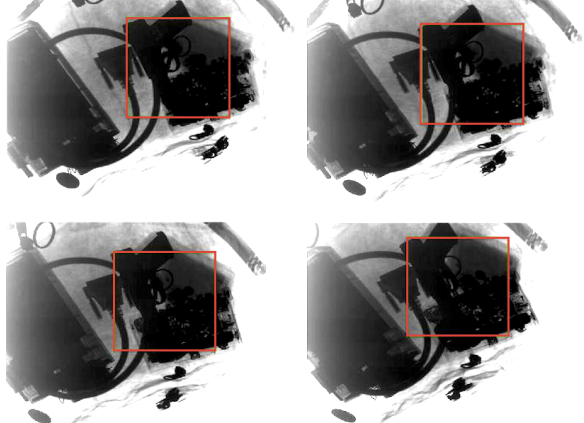
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
1	6	
1	7	
1	8	

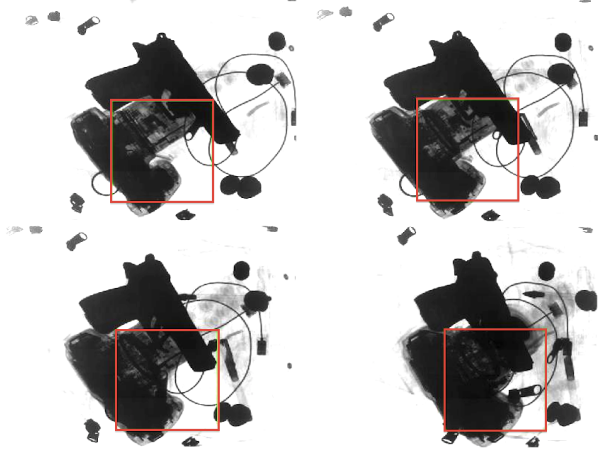
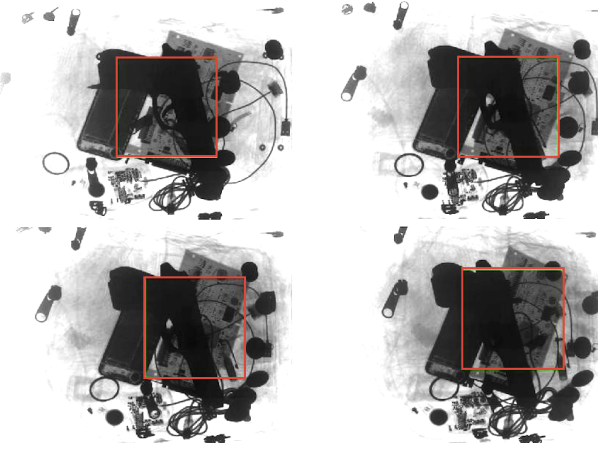
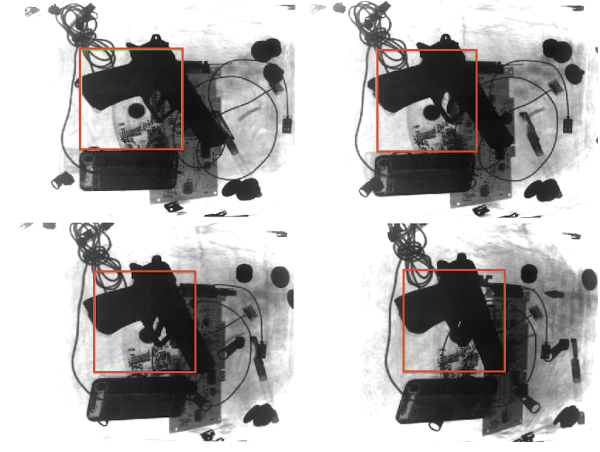
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
1	9	
2	10	
3	11	

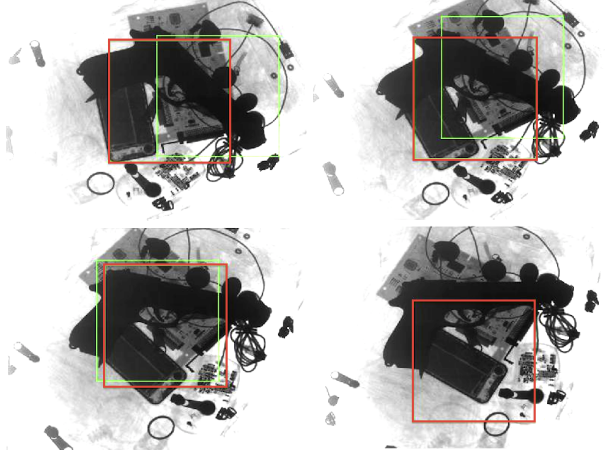
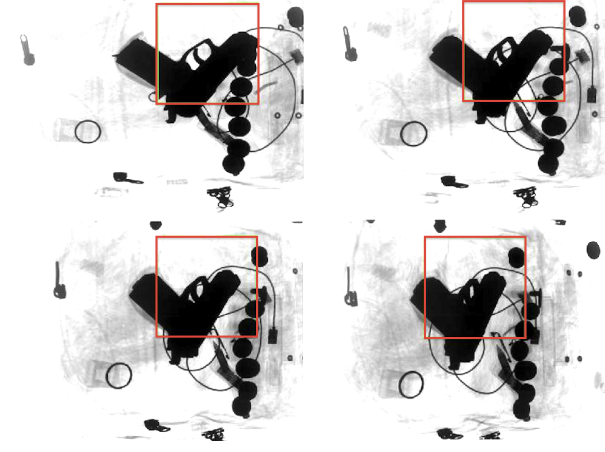
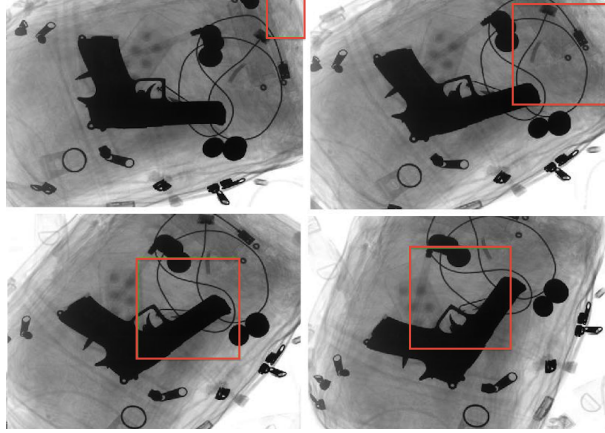
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
3	12	 This block contains four grayscale images of a mechanical assembly, arranged in a 2x2 grid. Each image features a red rectangular bounding box highlighting a specific component of the assembly. The assembly appears to be a complex mechanical part with various wires and connectors.
3	13	 This block contains four grayscale images of a mechanical assembly, arranged in a 2x2 grid. Each image features a red rectangular bounding box highlighting a specific component of the assembly. The assembly appears to be a complex mechanical part with various wires and connectors.
3	14	 This block contains four grayscale images of a mechanical assembly, arranged in a 2x2 grid. Each image features a red rectangular bounding box highlighting a specific component of the assembly. The assembly appears to be a complex mechanical part with various wires and connectors.

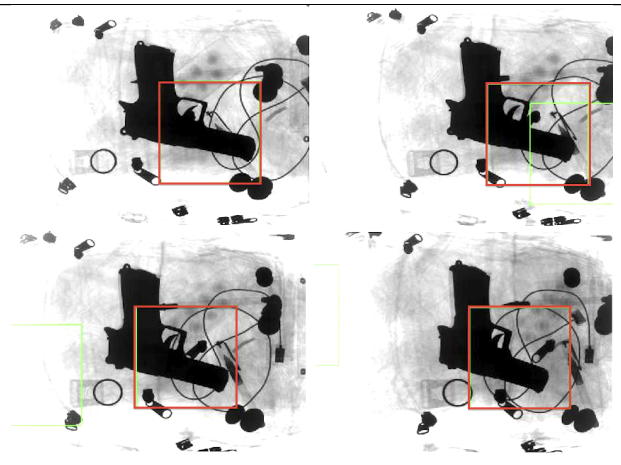
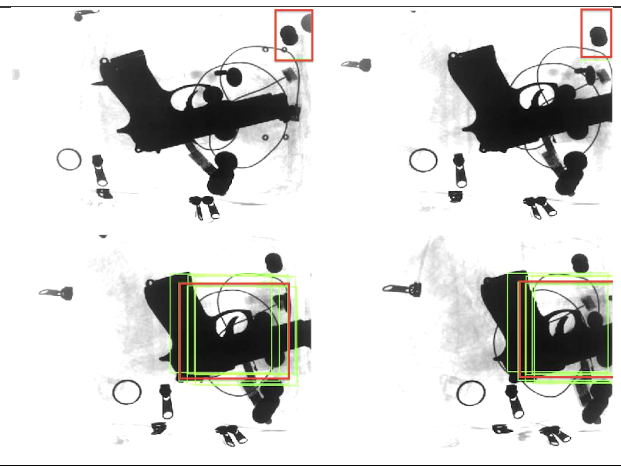
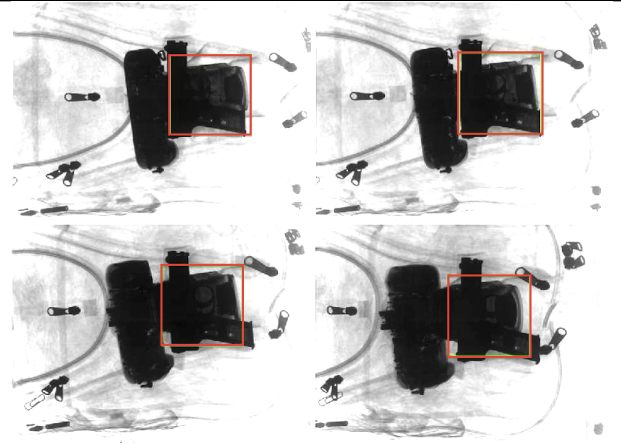
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
3	15	
4	16	
4	17	

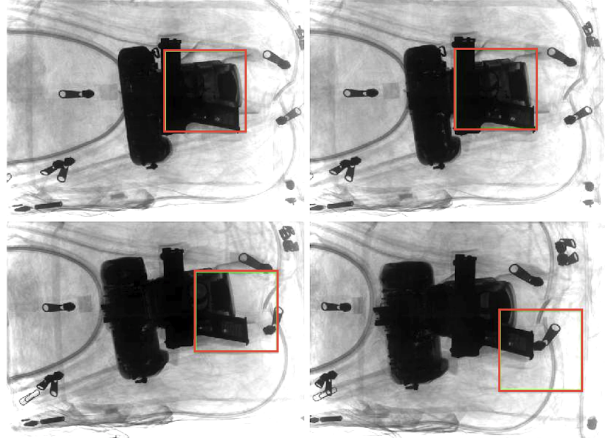
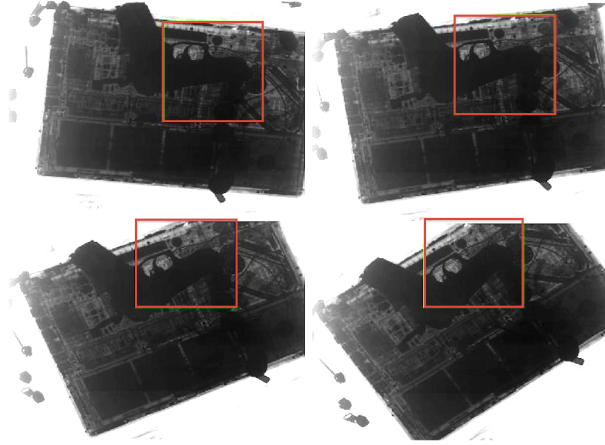
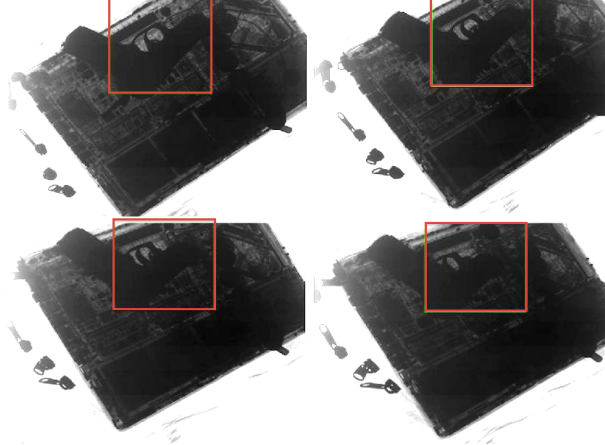
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
4	18	 Four grayscale images of a handgun, arranged in a 2x2 grid. Each image shows the handgun with a red bounding box around the main body and a green bounding box around the trigger area. The images are slightly different, showing different angles or lighting.
4	19	 Four grayscale images of a handgun, arranged in a 2x2 grid. Each image shows the handgun with a red bounding box around the main body and a green bounding box around the trigger area. The images are slightly different, showing different angles or lighting.
5	20	 Four grayscale images of a handgun, arranged in a 2x2 grid. Each image shows the handgun with a red bounding box around the main body. The images are slightly different, showing different angles or lighting.

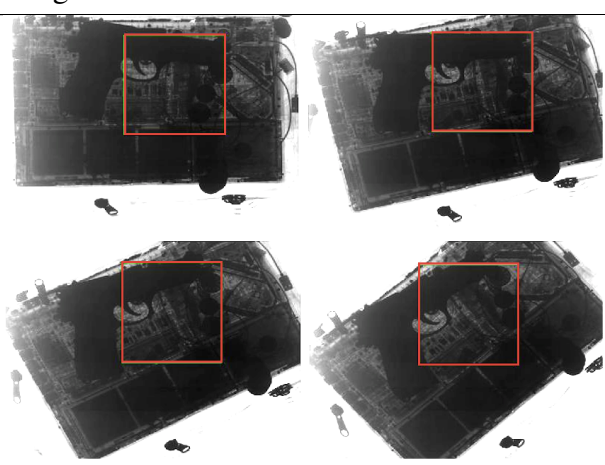
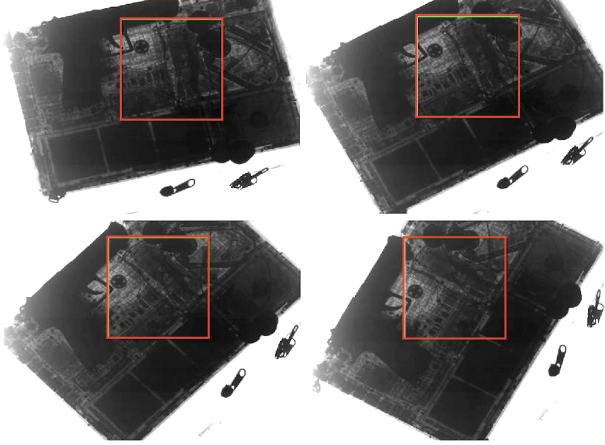
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
5	21	
6	22	
6	23	

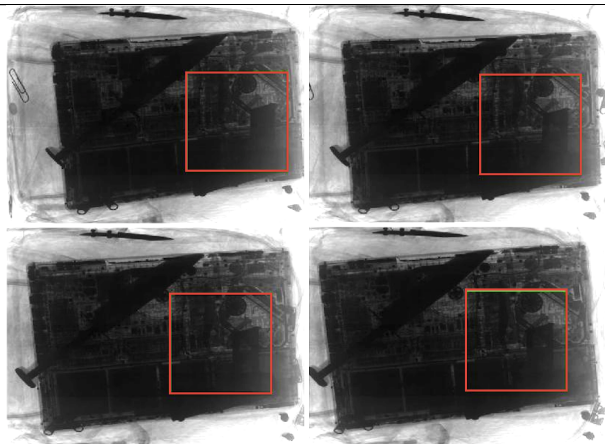
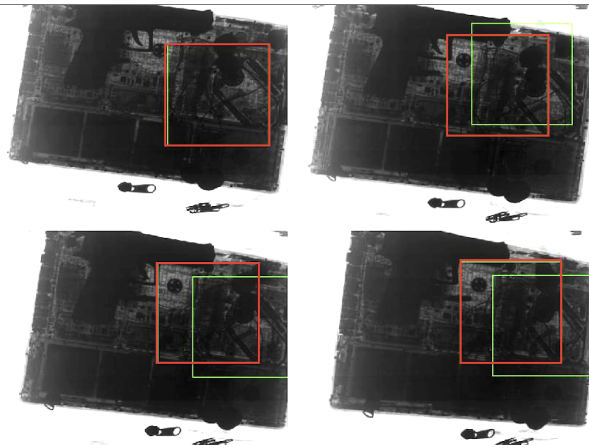
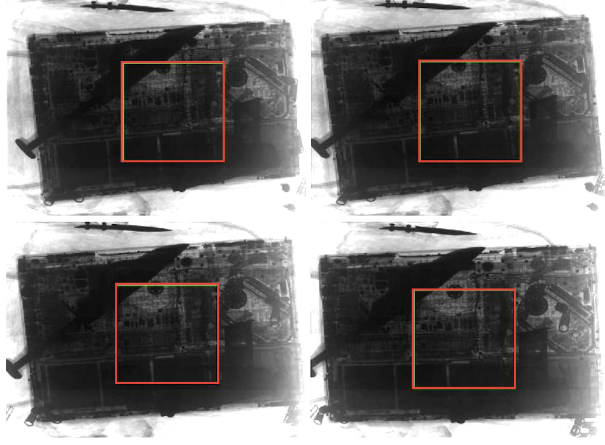
Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
6	24	
6	25	
6	26	

Continúa en la siguiente página. . .

Continuación de la Tabla 5.8

δ	ς	Imagen
6	27	
6	28	
6	29	
7	30	No existe imagen.