



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

RECONOCIMIENTO DE ROPA UTILIZANDO PERCEPCIÓN CONTINUA

TESIS PARA OPTAR AL GRADO DE DOCTOR EN INGENIERÍA
ELÉCTRICA

LUZ MARÍA MARTÍNEZ RAMÍREZ

PROFESOR GUÍA:

DR. JAVIER RUIZ DEL SOLAR SAN MARTÍN

MIEMBROS DE LA COMISIÓN:

DR. CLAUDIO ANDRÉS PÉREZ FLORES

DR. MARÍA JOSÉ ESCOBAR SILVA

DR. LUIS ENRIQUE SUCAR SUCCAR

Este trabajo ha sido parcialmente financiado por CONICYT- PCHA/Doctorado
Nacional/2014-21140280 y CONICYT-FONDECYT proyecto 1161500.

SANTIAGO DE CHILE

2019

**RESUMEN DE TESIS PARA OPTAR AL
GRADO DE: DOCTOR EN INGENIERÍA ELÉCTRICA
POR: LUZ MARÍA MARTÍNEZ RAMÍREZ
FECHA: 2019
PROFESOR GUÍA: DR. JAVIER RUIZ DEL SOLAR**

RECONOCIMIENTO DE ROPA UTILIZANDO PERCEPCIÓN CONTINUA

El reconocimiento de ropa a partir de una configuración inicial arbitraria y desconocida, es considerada una de las tareas más difíciles en el área de la visión computacional, debido a la gran cantidad de clases, modelos, diseños, y poses que puede llegar a tener cada ropa. Esta tesis propone una metodología que permite que el robot sea capaz de reconocer una ropa en una pose inicial aleatoria, al manipular la ropa y extraer la información necesario de una secuencia de vistas para clasificar la ropa en una de la 5 clases previamente definidas: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*.

Esta tesis presenta un método capaz de calcular los puntos de agarre para desplegar una prenda de ropa, que no requiere estos puntos almacenados previamente. Combinando el contorno de la región clave detectada y la información de rugosidad de la ropa obtenida. Para la evaluación del método, se intentó reconocer aquellos puntos de agarre que extienden la ropa y facilitan su doblado, logrando que para la clase *Pant*, *Shirt* y *T-Shirt* se podría desplegar satisfactoriamente un 43 %, 33 % y 29 % de las veces, respectivamente.

También se presenta un método de reconocimiento y clasificación de ropa utilizando percepción continua. En los experimentos se obtuvo una precisión de clasificación promedio del 67 %, con una precisión específica por clase de 58 %, 42 %, 84 %, 67 % y 84 % para las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*, respectivamente. Estos resultados son los mejores en comparación con otros métodos del estado-del-arte.

Finalmente, para el método de reconocimiento se presenta la primera base de datos de secuencias de ropa. Esta base de datos permite hacer evaluaciones de métodos que evalúan cuando la ropa se encuentra sobre la mesa, colgando por uno de los brazos del robot o utilizando percepción continua, que permite unificar las investigaciones de este problema.

Palabras Claves: Objetos deformables, reconocimiento de ropa, visión computacional, aplicaciones en robots de servicio.

**ABSTRACT OF THE THESIS TO OPT FOR THE
DEGREE OF: DOCTOR IN ELECTRICAL ENGINEERING
BY: LUZ MARÍA MARTÍNEZ RAMÍREZ
DATE: 2019
ADVISOR: PhD. JAVIER RUIZ DEL SOLAR**

CLOTHING RECOGNITION USING CONTINUOUS PERCEPTION

Clothing recognition from an arbitrary and unknown initial configuration is considered one of the most difficult tasks in the area of computational vision, due to a large number of classes, models, designs, and poses who can get every clothes. This thesis proposes a methodology that allows the robot to be able to recognize a piece of clothing in a random initial pose, by manipulating it and extracting the necessary information from a sequence of views to classify the clothes in one of the 5 previously defined classes: *T-Shirt*, *Shirt*, *Sweater*, *Pant* and *Hand-Towel*.

This thesis presents a system capable of calculating the grasp points to extend a garment, which does not require these points stored previously. Combining the contours of the key region detected and the roughness information of the clothing obtained. For the evaluation of the detection of the grip points, it was tried to recognize those grasp points that extend the clothes and facilitate their folding, achieving that for the class *Pant*, *Shirt* and *T-Shirt*, it could be satisfactorily deployed clothes a 43 %, 33 % and 29 % of the times, respectively.

Also, it presents a clothing recognition and classification system using continuous perception. In experiments where continuous perception is used, an average classification accuracy of 67 % was obtained, with a specific accuracy per class of 58 %, 42 %, 84 %, 67 % and 84 % for classes: *T-Shirt*, *Shirt*, *Sweater*, *Pant* and *Hand-Towel*, respectively. These results are the best compared to other state-of-the-art methods.

Finally, for the recognition system, the creation of the first database of clothing sequences is presented. This database allows the evaluation of methods that evaluate when the cloth is on the table, hanging by one of the arms of the robot or with the continuous perception, which allows unifying the investigations of this problem.

Keywords: Deformable objects, clothing recognition, computational vision, service robots applications.

*Gracias UChile Homebreakers,
estos años no habrían sido lo mismo sin ustedes.*

Luz María Martínez

Agradecimientos

Quiero comenzar esta sección agradeciendo a mi familia, a mis padres, hermanos, abuelos, tíos y primos por su constante apoyo y cariño. El camino que he escogido me ha llevado un poco lejos de ellos, pero siempre los tengo presente y lamento perderme algunos momentos importantes de sus vidas. Gracias a ellos siempre tengo un lugar al cual volver y del cual me cuesta mucho volverme a alejar.

Siendo un poco mas especifica, quiero agradecer a mi padre, por la dedicación a su trabajo para que nada nos faltara a mi o mis hermanos y por sobre todo por el inmenso amor que me tiene que lo demuestra siempre con pequeños y grandes detalles. A mi madre, por estar siempre ahí, por los lindos recuerdos de la casa reluciente y la rica comida cuando llegaba a comer. A mis hermanos, por siempre velar por mi, por cuidarme, por aguantarme, porque no importa la distancia a la que nos encontremos, siempre han estado y estoy muy segura que siempre estarán conmigo.

Por otro lado quiero agradecer al equipo UChile Homebreakers, aprendí mucho con ustedes y fue una genial experiencia. Estos años fueron increíbles, viajamos mucho, nos desvelamos trabajando, comimos mucha pizza, mucho esperanto, pasamos muchos problemas, pero fue genial. Recuerdo que cuando llegue al equipo se fueron muchos miembros y nos quedamos con el Mati y el Gonzalo solos, fue un poco demandante, pero logramos salir adelante y quedó un gran equipo. Gracias por apañar en todas: Mati, Gonzalo, Rodrigo, Diego, Gustavo Cristopher, Giovanni, Jose, David, Lukas, Leo y Nico.

Además, quiero agradecer a Matías Pavez, un gran compañero por muchos años y novio un poco más reciente. Gracias por ser una persona totalmente de confianza al momento de trabajar y por acompañarme en esta nueva aventura, con la que estoy feliz compartiendo mi vida contigo. Estoy segura que nos quedan muchos viajes y experiencias juntos.

Por otro... otro lado, no puedo dejar de lado a grandes amigos con los que compartí mucho estos años: Gustavi y Naxito, están lejitos ahora, pero espero que me sigan viniendo a visitar donde sea que esté, la paso bkn con ustedes y les tengo mucho cariño. A Christopher, gracias por el apoyo, por la ayuda y por las risas, has sido un gran amigo con el que siempre podría confiar y espero que tu también pienses lo mismo. A Joao y Phifo, principalmente gracias por las salidas, no saben cuanto extraño su terremoto en trova.

Finalmente pero no menos importante, quiero agradecer a mi profesor guía y profesor a cargo del proyecto Bender, el prof. Javier Ruiz-del-Solar. Gracias por aceptarme y dejarme realizar mi memoria de pregrado con usted, gracias a eso entre a trabajar en Bender y después me motive a realizar el doctorado. Gracias por su ayuda y consejos, fue un placer trabajar con usted.

Luz María Martínez Ramírez

Tabla de Contenido

Agradecimientos	vii
Índice de Tablas	xi
Índice de Ilustraciones	xiii
1. Introducción	1
1.1. Fundamentación General	1
1.2. Definición del Problema	3
1.3. Objetivos	4
1.3.1. Objetivo General	4
1.3.2. Objetivos Específicos	4
1.3.3. Hipótesis	5
1.4. Aportes de la Tesis	5
1.4.1. Publicaciones	6
1.5. Estructura de la Tesis	7
2. Revisión Bibliográfica	9
2.1. Robots Trabajando con Ropa	9
2.1.1. PR2	10
2.1.2. HRP-2	11
2.1.3. CloPeMa	11
2.1.4. Baxter	13
2.2. Reconocimiento de Ropa	13
2.2.1. Caso de Ropa Sobre una Superficie	13
2.2.2. Caso de Ropa Colgando	19
2.3. Puntos de Manipulación	24
2.4. Resumen y Análisis del Capítulo	27

3. Metodología para Reconocimiento y Clasificación de Ropa	29
3.1. Detección de Puntos de Agarre	29
3.1.1. Reconocimiento de Partes Claves	31
3.1.2. Análisis de Arrugas	35
3.1.3. Detección de Puntos de Agarre	37
3.2. Clasificación de Ropa Mediante Percepción Continua	39
3.2.1. Características Locales	43
3.2.2. Características Globales	46
3.2.3. Codificación de Características Locales	48
3.2.4. Clasificación	51
3.3. Resumen y Análisis del Capítulo	54
4. Experimentos y Resultados	57
4.1. Detección de Puntos de Agarre	57
4.1.1. <i>Clothing Part Dataset</i>	58
4.1.2. Resultados del Reconocimiento de Partes Claves	59
4.1.3. Resultados de la Detección de Puntos de Agarre	60
4.2. Clasificación con Percepción Continua	63
4.2.1. Bases de Datos	63
4.2.2. Experimentos de Percepción Continua	65
4.2.3. <i>Ablation Study</i>	67
4.2.4. Estrategia en Percepción Continua	70
4.2.5. Experimentos <i>single-shot</i>	71
4.3. Resumen y Análisis del Capítulo	73
5. Conclusiones y Trabajo Futuro	77
5.1. Conclusión	77
5.2. Trabajo Futuro	80
6. Bibliografía	81
Anexos	90
A. <i>Continuous Clothing Dataset</i>	91

Índice de Tablas

2.1. Resumen del estado-del-arte de reconocimiento de ropa sobre una superficie	18
2.2. Comparación de métodos utilizando el enfoque de la ropa colgando	23
2.3. Resumen del estado-del-arte de reconocimiento de ropa colgada	24
4.1. Matriz de confusión del reconocimiento de partes claves.	59
4.2. Resultados promedios de la detección de puntos de agarre utilizando la métrica IoU	60
4.3. <i>Recall</i> de las detecciones de puntos de agarre	61
4.4. Comparación de rendimiento entre el método propuesto con dos métodos del estado-del-arte.	66
4.5. <i>Ablation study</i> de las características del método propuesto	68
4.6. <i>Ablation study</i> de los algoritmos de codificación del método propuesto	69
4.7. Resultados de los experimentos al utilizar distintas estrategias en el algoritmo. . . .	71
4.8. Comparación de rendimiento entre el método propuesto con y sin características distintivas contra do	
A.1. Información y <i>rate</i> promedio de los tópicos de cada rosbag	92

Índice de Ilustraciones

2.1. Robots: PR2, HRP-2 y Baxter	11
2.2. Robot del proyecto CloPeMa	12
2.3. Ejemplos de detección de partes claves	14
2.4. Resultados de la detección de partes claves utilizando FINDDD	16
2.5. Comparación de descriptores 3D	17
2.6. Buffer de imágenes de profundidad	19
2.7. Reconocimiento de clase y de pose utilizando un clasificador jerárquico de dos etapas	20
2.8. Diagrama de reconocimiento y desdoblado de ropa utilizando el framework POMDP	21
2.9. Arquitectura de una red profunda para el reconocimiento de ropa	22
2.10. Ejemplo de imagen de rugosidad	26
3.1. Ejemplo de detección de puntos de agarre para desplegar la ropa	30
3.2. Diagrama de bloques del método de detección de puntos de agarre	30
3.3. Filtro de entropía	31
3.4. Cálculo de puntos más altos del filtro de entropía	32
3.5. Detección de contornos de partes claves	33
3.6. Partes claves en nube de puntos	34
3.7. Resultados del filtro de entropía y de <i>vesselness</i>	36
3.8. Cálculo de puntos más altos con el filtro de <i>vesselness</i>	37
3.9. Selección de puntos candidatos de agarre de la clase <i>Shirt</i> y <i>T-Shirt</i>	38
3.10. Selección final de puntos de agarre de la clase <i>Shirt</i> y <i>T-Shirt</i>	38
3.11. Selección de puntos candidatos de agarre de la clase <i>Waist</i>	39
3.12. Selección final de puntos de agarre de la clase <i>Waist</i>	39
3.13. Diagrama de clasificación de ropa utilizando percepción continua	41

3.14. Diagrama de reconocimiento utilizando percepción continua	42
3.15. Diagrama del descriptor FINDDD	44
3.16. Cálculo del descriptor SC	45
3.17. Categorías de <i>Shape Index</i>	46
3.18. Explicación del descriptor LBP	48
3.19. Composición del Vector de Características Compuestas	51
4.1. Imágenes de las clases evaluadas de la base de datos <i>Clothing Part Dataset</i>	59
4.2. Resultados de la detección de puntos de agarre.	62
4.3. Imágenes de las clases de la base de datos <i>Continuous Clothing Dataset</i>	64
4.4. Imágenes de las clases de la base de datos <i>Free-Configuration Clothing Dataset</i>	65
4.5. Matriz de confusión del método propuesto	67
4.6. Resultados del <i>ablation study</i> sobre las características	70
4.7. Resultados del <i>ablation study</i> sobre los algoritmos de codificación	70
4.8. Matriz de confusión del método propuesto en <i>single-shot</i>	72
A.1. Robot CloPeMa con los sensores indicados.	92
A.2. Modelos de la clase <i>Pant</i> de la base de datos	94
A.3. Modelos de la clase <i>Shirt</i> de la base de datos	95
A.4. Modelos de la clase <i>Sweater</i> de la base de datos	96
A.5. Modelos de la clase <i>Hand-Towel</i> de la base de datos	97
A.6. Modelos de la clase <i>T-Shirt</i> de la base de datos	98

Capítulo 1

Introducción

1.1. Fundamentación General

En la actualidad, las personas destinan bastante de su tiempo libre para limpiar y ordenar su casa. Según la Encuesta Nacional sobre el Uso del Tiempo [12] las personas destinan entre 1.83 y 3.84 horas diarias al trabajo doméstico en Chile. Es en este escenario que los robots de servicio llegan a tener tanta importancia. Su objetivo principal es ayudar en las distintas tareas del hogar, tales como: limpieza, cocina, entretenimiento, y vigilancia, entre otros. Algunas de las tareas más repetitivas que se realizan constantemente en los hogares y en los cuales los robots de servicio podrían llegar a ser de gran utilidad, son separar, lavar, secar, doblar y ordenar la ropa. Estas tareas pertenecen a las tareas del hogar de carácter irrenunciables, siendo realizadas cada 1 o 2 semanas dependiendo de los integrantes del hogar. Para que estas tareas puedan llegar a ser realizadas por un robot de servicio, el robot debe ser capaz de reconocer el tipo de ropa. Este conocimiento le daría al robot la información necesaria para interactuar con ella, detectar sus puntos de agarre y los parámetros necesarios para su manipulación.

Los robots de servicio llegarán a ser de gran utilidad para los seres humanos, proporcionándoles tiempo libre para sus intereses personales al librarlos de las tareas domésticas. Estos robots han sido definidos con el objetivo principal de realizar labores que sean de ayuda para el ser humano, normalmente mediante la realización de tareas de carácter diario o semanal en un ambiente parcialmente desconocido y constantemente cambiante. Si bien el concepto de robots de servicio

no tiene una definición fija aún, la IFR¹ propone como definición tentativa: “Un robot de servicio es un robot que opera de forma parcial o completamente autónoma para realizar servicios útiles para el bienestar de humanos y equipamientos, excluyendo labores de manufactura”. Al tener que trabajar en un entorno cambiante, la habilidad de detectar e identificar objetos que se encuentran a su alrededor es de suma importancia, dado que deben tener la capacidad de poder interactuar con ellos y así realizar las tareas para las que fueron diseñados. Son muchos los objetos que se pueden encontrar en el hogar, de distintas clases, tamaños y variedades. Siendo los objetos deformables uno de los más complicados de detectar y manipular. Los objetos deformables abarcan todas las clases de ropa, paños y bolsas, entre otros.

La interacción con objetos deformables es un tópico de investigación relativamente nuevo y cada vez más popular en el área de la visión computacional. Esto se puede deber a que la habilidad de reconocer y manipular objetos deformables es altamente necesaria para los robots que están planeados para trabajar en hogares, hospitales y algunas tiendas. Además, también se debe considerar que en los últimos años ha mejorado la tecnología que se enfoca en manipular materiales deformables, como por ejemplo, la aparición de nuevos *grippers* en el mercado que permiten agarrar materiales con muy poca profundidad y con una gran precisión.

La motivación principal en este trabajo fue utilizar la robótica para mejorar la calidad de vida de las personas, aumentando las horas libres de hombres y mujeres al reducir el tiempo que pasan realizando tareas del hogar, enfocando este trabajo de tesis en el reconocimiento y clasificación de ropa para robots de servicio.

¹IFR : International Federation of Robotics

1.2. Definición del Problema

El reconocimiento de ropa en un estado aleatorio es considerada una tarea difícil incluso para un ser humano, ya que existen una gran cantidad de clases, por ejemplo, pantalón corto, pantalón largo, camisa de hombre/mujer, polera sin mangas, toalla de mano y toalla de baño, entre otras. Además, por cada una de estas clases hay una gran variedad de modelos, tamaños y diseños. Tomando todo esto en consideración y sumando que los estados que puede llegar a tener cada prenda de ropa al ser un objeto deformable, lo convierte en una de las tareas más difíciles en el área de la visión computacional.

Este trabajo de tesis se enfocó específicamente en el reconocimiento de ropa para robots de servicio humanoides que operen en ambientes controlados, y que sean capaces de manipular objetos en forma precisa. Estos robots deben poseer al menos una cámara RGB-D y preferiblemente 2 brazos.

El objetivo del sistema es que el robot sea capaz de reconocer una prenda de ropa cuando esta se encuentre en una pose inicial desconocida y aleatoria. El robot debe ser capaz de tomar la ropa, obtener distintas vistas utilizando percepción continua y clasificar la prenda de ropa en una de las 5 clases previamente definidas: *T-Shirt* (polera), *Shirt* (camisa), *Sweater* (sweater), *Pant* (pantalón) y *Hand-Towel* (toalla de mano). Las cuales fueron seleccionadas por su amplia utilización en el estado-del-arte. Finalmente, el robot debe depositar la prenda de ropa en un lugar previamente definido, dependiendo de la clase reconocida.

1.3. Objetivos

1.3.1. Objetivo General

Diseñar e implementar un sistema de reconocimiento de ropa para robots de servicio. La prenda de ropa a ser reconocida estará en una pose inicial desconocida y aleatoria, y luego de ser reconocida por el robot, deberá quedar en un lugar previamente definido, dependiendo de su clase. El robot deberá ser capaz de manipular la prenda para modificar su pose y facilitar su reconocimiento.

1.3.2. Objetivos Específicos

1. Diseñar e implementar métodos de extracción de características para su uso en clasificación de ropa.
2. Diseñar e implementar un método de detección de puntos de agarre para distintas clases de ropa.
3. Implementar métodos de manipulación de ropa basados en puntos de agarre utilizando uno o ambos brazos del robot.
4. Diseñar e implementar un modelo probabilístico para clasificación de ropa.
5. Construir una base de datos que permita la evaluación de distintos métodos de clasificación de ropa.
6. Implementar un sistema para robots de servicios para clasificar la ropa, integrando los módulos desarrollados anteriormente.
7. Validación del sistema bajo diferentes selecciones de ropa utilizando un robot humanoide.

1.3.3. Hipótesis

Esta tesis busca realizar aportes en el reconocimiento de ropa, realizando una integración entre la visión computacional y la manipulación sobre el objeto. En particular, uno de los aportes principales es la utilización de un modelo probabilístico que se va actualizando mediante la nueva información encontrada, considerando evidencia de distinta naturaleza para mejorar el reconocimiento.

1. El uso e integración de información proveniente de distintas deformaciones de la ropa puede mejorar el rendimiento y precisión del sistema de reconocimiento de ropa.
2. La integración de características de distinta naturaleza mejora el rendimiento del sistema de reconocimiento de ropa.

1.4. Aportes de la Tesis

En este trabajo de tesis se lograron tres aportes al estado arte. Se diseñó e implementó un método de reconocimiento de partes claves de ropa, tales como, el cuello de la clase *Shirt*, el cuello de la clase *T-Shirt* y la cintura de la clase *Pant* [44].

Además, se diseñó e implementó un método de reconocimiento y clasificación de ropa que mejoró los resultados del estado-del-arte. Este método, clasifica la ropa utilizando la información extraída de una secuencia de imágenes utilizando un enfoque de percepción continua basado en el uso de múltiples imágenes de profundidad de diferentes sensores, mientras que un brazo robótico mueve una prenda sobre y desde una superficie plana [43].

Otro aporte, fue la creación de la primera base de datos de secuencias de imágenes de ropa. La base de datos proporciona las imágenes RGB-D de dos sensores con distintas perspectivas y la cinemática del robot. Esto permite hacer evaluaciones de métodos que reconocen cuando la ropa se encuentra sobre una superficie, colgando por uno de los brazos del robot o utilizando percepción continua.

1.4.1. Publicaciones

Las publicaciones generadas que están directamente relacionadas con este trabajo de tesis, son las siguientes:

- Luz Martínez and Javier Ruiz-del-Solar and Li Sun and Jean Paul Siebert and Gerardo Aragon-Camarasa, “Continuous perception for deformable objects understanding”. In *Robotics and Autonomous Systems*, 118:220–230, 2019.
- Luz María Martínez, Javier Ruiz-del-Solar. “Recognition of Grasp Points for Clothes Manipulation under unconstrained Conditions”. In *Proceedings of 22th RoboCup International Symposium, Lecture Notes in Computer Science*, 2017.

Además, otras publicaciones fueron generadas durante el tiempo de desarrollo de esta tesis:

- Patricio Loncomilla, Javier Ruiz-del-Solar, Luz María Martínez. “Object recognition using local invariant features for robotic applications: A survey”. In *Pattern Recognition*, 60:499-514, 2016.
- María-Loreto Sánchez, Mauricio Correa, Luz María Martínez, Javier Ruiz-del-Solar. “An Episodic Long-Term Memory for Robots: The Bender Case”. In *Proceedings of 20th RoboCup International Symposium, Lecture Notes in Computer Science*, pp. 264-275, 2015.
- Luz María Martínez, Patricio Loncomilla, Javier Ruiz-del-Solar. “Object Recognition for Manipulation Tasks in Real Domestic Settings: A Comparative Study”. In *Proceedings of 19th RoboCup International Symposium, Lecture Notes in Computer Science*, pp. 207-219, 2014.

1.5. Estructura de la Tesis

Este trabajo de tesis fue estructurado de la siguiente forma. El Capítulo 2 presenta una revisión bibliográfica de los robots y sistemas utilizados en reconocimiento y clasificación de ropa y sistemas de detección de puntos de agarre de ropa. El Capítulo 3 describe la metodología propuesta de los sistemas de reconocimiento de partes claves de ropa y del sistema de reconocimiento y clasificación de ropa. El Capítulo 4 presenta los experimentos, base de datos utilizadas y los resultados de los sistemas propuestos en comparación con otros existentes en el estado del arte. Finalmente el Capítulo 5 presenta las conclusiones, discusiones y el trabajo futuro.

Capítulo 2

Revisión Bibliográfica

Este capítulo presenta una amplia revisión bibliográfica de los distintos tipos de trabajos realizados en reconocimiento y manipulación de ropa. Primero, se presentan los robots de servicio con los cuales se han realizado investigaciones en reconocimiento, estirado, manipulación y/o doblado de ropa. Luego, se informa el estado del arte de sistemas de reconocimiento y clasificación de ropa. Estos sistemas utilizan principalmente dos enfoques para realizar el reconocimiento. El primero se basa principalmente en extraer características de la ropa mientras se encuentra sobre una superficie sin realizar ninguna interacción sobre ella. El segundo enfoque es tomar la ropa y dejarla colgando de un brazo para ver más superficie de la ropa al momento de extraer las características. Finalmente, la sección tres presenta el estado del arte de sistemas de detección de puntos de manipulación, ya sea cuando la ropa se encuentra en una posición aleatoria sobre una superficie y se debe agarrar un punto para levantarla, o cuando ya se encuentra colgando de uno de los brazos del robot y se necesita detectar otro punto de manipulación para desplegar la ropa.

2.1. Robots Trabajando con Ropa

El sistema de reconocimiento de esta investigación busca ser utilizado en robots de servicio. Es por esto que es necesario conocer los robots que ya han sido utilizados en sistemas de reconocimiento y manipulación de ropa y las características que estos robots poseen. Actualmente, son cuatro los robots de servicio principales en los cuales se han realizado investigaciones en las áreas de reconocimiento, estirado y/o manipulación de ropa. El robot que tiene las primeras

investigaciones en ropa corresponde al robot PR2 de Willow Garage [19]. Este robot se encuentra en varios centros de investigación, y es la plataforma que posee más publicaciones relacionados al trabajo con ropa. Otro robot en el cual se han realizado bastantes investigaciones en esta área, es el robot del proyecto *Clothes Perception and Manipulation* (CloPeMa) [1]. Este robot fue creado exclusivamente para trabajar con ropa y realizar investigaciones en esta área. Finalmente, están los robots HRP-2 de KAWADA Industries [21] y Baxter de Rethink Robotics [54], los cuales han sido utilizados como plataformas de desarrollo para investigaciones en el área de reconocimiento y de doblado de ropa.

2.1.1. PR2

El *personal robot 2* (PR2) (ver Figura 2.1) construido por Willow Garage es una plataforma de manipulación móvil con el sistema de software trabajando completamente en *Robot Operating System* (ROS). El robot está equipado con dos brazos de *7 degree of freedom* (7-DOF) que cuentan con dos *grippers* que pueden levantar hasta 1.8 Kg. Los *grippers* cuentan con sensores de presión en sus pinzas, utilizados para determinar si la ropa ha sido tomada con éxito. Además posee 4 sensores en la cabeza y mono-cámaras en los antebrazos.

La cabeza del robot PR2 tiene tres cámaras y un *textured light projector*:

- *Wide Stereo Camera*: Esta cámara de color posee una resolución máxima de 752×480 píxeles a 15 fps y un campo de visión (FOV) de aproximadamente 90 grados.
- *Narrow Stereo Camera*: Esta cámara monocromática posee una resolución máxima de 752×480 píxeles a 15 fps y un FOV de aproximadamente 55 grados.
- *Gigabit Ethernet Camera*: Esta cámara tiene una resolución máxima de 2448×2050 píxeles a 15 fps.
- *Textured Light Projector*: El proyector tiene un FOV de aproximadamente 55 grados y una lente Kowa LM12JC de 5,6 mm F2.0.

El robot PR2 ha sido utilizado como plataforma en muchas investigaciones de ropa, tales como, el reconocimiento y manipulación de calcetines [75], reconocimiento de distintas clases de ropa

[11] y la detección de los puntos de manipulación [39], pero la mayor cantidad de investigaciones han sido en el doblado de ropa [8, 31, 33, 34, 39, 45, 46, 73].

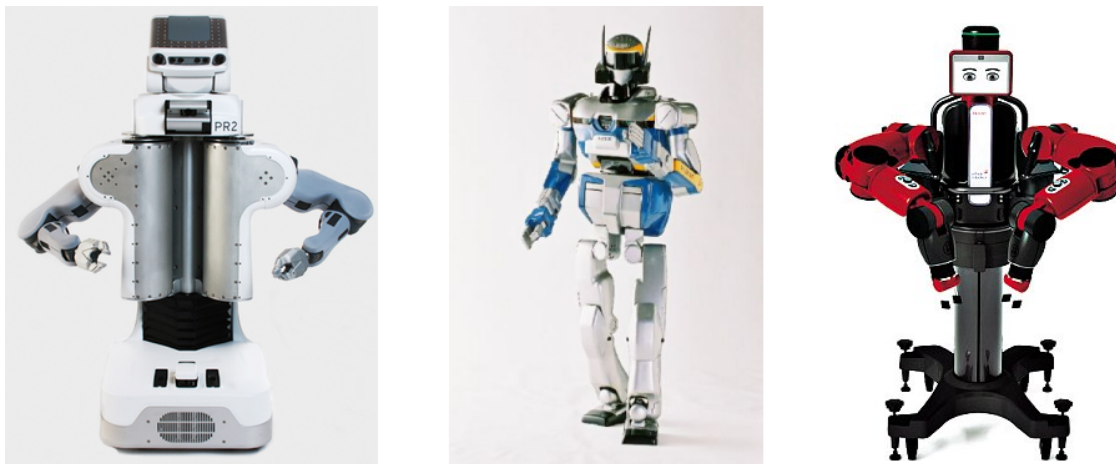


Figura 2.1: De izquierda a derecha se pueden observar los robots: PR2, HRP-2 y Baxter. Fuentes: Willow Garage [19], LAAS [30] y HTE Technologies [70].

2.1.2. HRP-2

El robot humanoide HRP-2 (ver Figura 2.1) diseñado e integrado por KAWADA Industries fue pensado para las áreas de “caminar sobre superficies irregulares”, “control de vuelco”, “levantarse de una posición caída” y “operaciones interactivas con humanos en espacios abiertos”. El robot posee dos brazos de 6-DOF con sensores de fuerza que pueden levantar hasta 2 kg. Para las investigaciones en ropa se anexo en el robot un sistema de visión estereoscópica trinocular, ya que el robot no posee incorporado un sistema de visión 3D.

Los trabajos realizados en el robot HRP-2 consisten en el reconocimiento de ropa [27], reconocimiento de puntos de agarre para desplegar ropa [24, 28] y en el doblado de ropa [25].

2.1.3. CloPeMa

El robot CloPeMa (Clothes Perception and Manipulation) [59] nació de un proyecto de investigación europeo que comenzó en el año 2012 y en el cual participaron 5 grandes centros de investigación de Europa. El objetivo del proyecto era diseñar el prototipo de un robot que fuera capaz de doblar ropa. Proporcionando al robot una variedad de componentes, tales como, cámaras

estéreo, cámaras 3D, sensores de fuerza y sensores fotométricos de corto alcance. El objetivo del proyecto se cumplió con la realización de 3 robots CloPeMa con todos los componentes antes mencionados.

El robot CloPeMa (ver Figura 2.2) posee dos brazos robóticos de soldadura industrial Yaskawa/Motoman MA1400. Equipados con *grippers* que tienen dos piezas asimétricas, diseñados especialmente para la manipulación de prendas de ropa. Los *grippers* poseen una pinza delgada se utiliza para deslizarse por debajo de la ropa, mientras que la segunda pinza incorpora varios sensores. El proyecto ha dado lugar a métodos para el reconocimiento de clase y de pose [13,14,42], de estirado [40, 41, 66, 67] y de doblado [48, 61, 62] de distintas clases de ropa, que han sido importantes contribuciones al estado del arte.



Figura 2.2: Robot del proyecto CloPeMa. Fuente: Doumanoglou et al., 2014 [13]

El robot posee los siguientes componentes que le permiten obtener información de distintas naturalezas de los objetos:

- Cámaras estéreo: El robot posee dos cámaras Nikon DSLR (D5100) que son capaces de capturar imágenes RGB y de profundidad de 16 megapíxeles.
- Cámaras 3D: El robot ha sido equipado con dos Asus Xtion Pro Live en las muñecas de cada brazo.
- *Grippers* con sensor de tacto: Los *grippers* poseen un sensor táctil entre las pinzas para analizar el material de la ropa.
- *Grippers* con sensor estéreo fotométrico: Sensor de cámara estéreo de pequeña escala que se encuentra en la “palma” de la pinza (altura: 39,5 mm, diámetro: 39 mm). El sensor obtiene

imágenes con una resolución de 1280 x 800 píxeles y cuentan con soporte de software para realizar la reconstrucción 3D de la superficie de la ropa.

- Sensores de fuerza y de torque en las articulaciones de los brazos.

2.1.4. Baxter

Finalmente, el robot Baxter (ver Figura 2.1) construido por la empresa Rethink Robotics está equipado con dos brazos de 7-DOF, sensores táctiles en las pinzas y con un sensor de profundidad en la parte superior de la cabeza. Las investigaciones realizadas en este robot son en reconocimiento de ropa [36,37], manipulación [18] y doblado [37].

2.2. Reconocimiento de Ropa

La percepción en 3D juega un papel significativo en la robótica, siendo una de las opciones más prometedoras para la amplia aplicación de soluciones robóticas a los problemas cotidianos. Existen varios sensores 3D, que son económicos y fiables que ofrecen una gran cantidad de nuevas oportunidades para desarrollar aplicaciones robóticas útiles para entornos domésticos. Asimismo, con la aparición y mejora de estos sensores 3D han aumentado las investigaciones realizadas en el área del reconocimiento de ropa.

Actualmente los métodos de reconocimiento de ropa han utilizado dos enfoques para resolver este problema. El primer enfoque que se utilizó consiste en detectar la ropa cuando se encuentra sobre una superficie, por ejemplo una mesa, y sin ningún otro objeto a su alrededor. Luego al intentar extraer características de una mayor superficie de la ropa, apareció el segundo enfoque que consiste en realizar el reconocimiento de la ropa cuando el robot la tiene colgando desde uno de sus brazos.

2.2.1. Caso de Ropa Sobre una Superficie

En general, la detección de ropa en su totalidad es una tarea difícil, ya que la ropa en su estado desdoblado no muestra una clara distinción sobre a que clase pertenece. Sin embargo, las clases de

ropa poseen secciones propias de cada clase (partes claves), que ayudan a la distinción entre clases. La Figura 2.3 muestra las partes claves de 6 clases de ropa que sirven como información para el reconocimiento, donde por cada una de las seis clases, se muestra un panel con cuatro imágenes: imagen de color con las anotaciones superpuestas de las partes claves (tipos de *collars*, *hip* y *hood*) y otras partes secundarias (tipos de *sleeves* y *hemline*), imagen original, máscara de segmentación y una representación de la información de profundidad. Al final de la sección, en la Tabla 2.1, se puede observar un resumen del estado-del-arte de los métodos de reconocimiento de partes claves y de los métodos de reconocimiento de ropa que se detallan en esta sección.



Figura 2.3: Ejemplos de detección de partes claves. En las imágenes se puede observar: anotaciones superpuestas (arriba), imagen original (parte inferior izquierda), máscara de segmentación (parte inferior central) y una representación de la información de profundidad (parte inferior derecho). Fuente: Ramisa et al., 2014 [53].

Considerando que por cada clase existen una gran cantidad de modelos en distintos colores, las investigaciones en reconocimiento de ropa y reconocimiento de parte claves de ropa se han enfocado principalmente en la utilización de descriptores 3D, tales como, *Geodesic-Depth Histogram* (GDH) [52], *Fast Point Feature Histogram* (FPFH) [78, 79], *Heat Kernel Signature* (HKS) [53] y *Fast Integral Normal 3D* (FINDDD) [50].

El descriptor GDH [38] captura la distribución conjunta de las distancias y profundidades geodésicas dentro de un parche. Es una adaptación al histograma de intensidad geodésica (GIH), propuesto originalmente para describir parches de imágenes deformables.

El descriptor FPFH [55], es una simplificación del descriptor PFH [57], está diseñado para caracterizar en una nube de puntos la geometría local alrededor de un punto. Donde dado un punto p_q para el cual se quiere calcular el descriptor, para cada uno de sus k -vecinos más cercanos, se determina un marco de coordenadas locales $\{u, v, w\}$ entre el punto de consulta y su vecino p_t , y se calculan las tres relaciones geométricas presentadas en la ecuación (2.1).

$$\begin{aligned}\alpha &= v \cdot n_t \\ \phi &= u \cdot \frac{(p_t - p_q)}{d} \\ \theta &= \arctan(w \cdot n_t, u \cdot n_t)\end{aligned}\tag{2.1}$$

donde d es la distancia euclidiana entre los puntos p_q y p_t y las normales de los dos puntos están representadas por n_q y n_t , respectivamente. Al tener las relaciones geométricas se calcula un descriptor similar para cada uno de los k -vecinos, en su propia vecindad k , y se realiza una suma ponderada de los descriptores simplificados para incorporar toda la información en el descriptor FPFH.

El descriptor HKS [63] se basa en el concepto de difusión de calor sobre una superficie desde un punto en específico. El descriptor captura la información sobre la vecindad de un punto al registrar la disipación de calor desde el punto hacia el resto. Debido a que el calor se difunde hacia vecindarios progresivamente más grandes, el parámetro de tiempo proporciona una noción natural de escala para describir la forma alrededor de un punto. Para cada punto de la nube de puntos, HKS define su vector de características que representa las propiedades geométricas del punto. Este descriptor está motivado por el hecho de que las deformaciones isométricas de la forma no cambian su topología, por lo tanto no cambiarán la forma en que se difunde el calor.

Dentro de los descriptores que han tenido un mejor desempeño en reconocimiento de ropa se encuentra FINDDD, el cual muestra una tasa de reconocimiento similar a los demás, pero al utilizar la imagen integral para calcular el vector normal de cada punto, el tiempo de procesamiento es más rápido en comparación con los otros descriptores 3D nombrados. El descriptor FINDDD [50] representa la distribución de las orientaciones de las normales 3D en una región alrededor de

un punto de interés en una nube de puntos estructurada. Para generar el descriptor FINDDD se calculan las normales de la nube de puntos, utilizando imágenes integrales para acelerar el proceso. Luego, la nube de puntos se divide en subregiones, calculando para cada una de ellas un descriptor mediante la construcción de histogramas de orientación normal. Sin embargo, en lugar de usar las coordenadas esféricas para representar los ángulos, el descriptor FINDDD se distribuye regularmente en toda la semiesfera utilizando coordenadas cartesianas. Esto evita la concentración alrededor del polo norte (elevación máxima) y el área desigual en las partes del histograma causada por la representación angular. En la Figura 2.4 se puede observar los resultados de detección de la parte clave *Neck* de la clase *Shirt*.

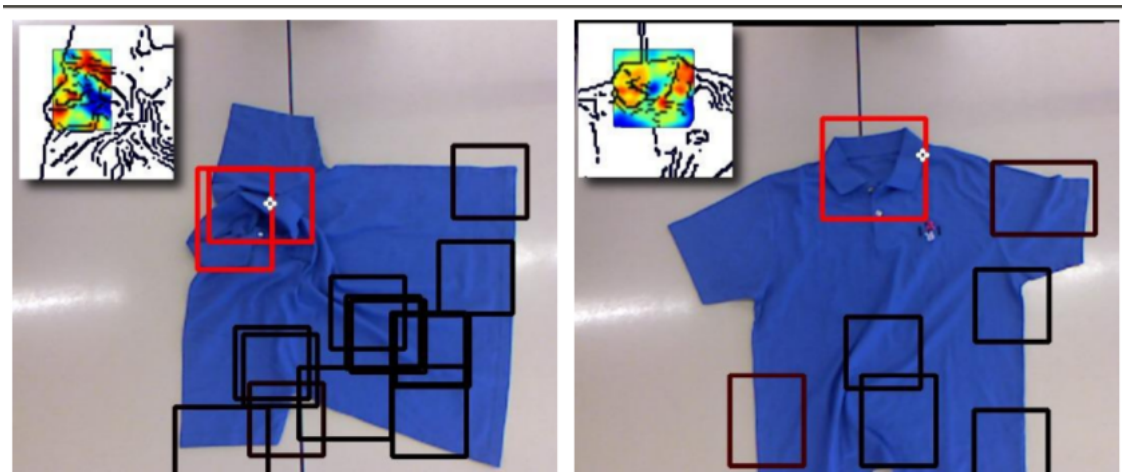


Figura 2.4: Resultados de la detección de partes claves utilizando FINDDD. Las dimensiones de las cajas corresponden a los cuellos de camisa detectados por FINDDD (normalizado de rojo a negro). Superpuesto en la parte superior izquierda de cada imagen, existe el mapa puntuación obtenida por el clasificador SVM solapa dentro del cuadro delimitador de más alto rango (el rojo significa más alta y más baja azul). El punto de agarre al final (puntuación más alta de la SVM) está marcado con una cruz blanca. Fuente: Ramisa et al., 2013 [50]

Para aplicaciones en tiempo real, el tiempo computacional de los descriptores 3D tiene gran relevancia. En Ramisa et al., 2014 [53] las implementaciones de los descriptores FPFH y FINDDD se encuentran en C++, y por lo tanto sus tiempos de cálculo son directamente comparables. El tiempo promedio del descriptor FINDDD es de 1 segundo y el del descriptor FPFH es de 322 segundos. Las implementaciones de los descriptores GDH y HKS se encuentran en Matlab y sus tiempos son considerablemente mayores que los implementados en C++. El descriptor GDH tiene un tiempo promedio de 6826 segundos y el descriptor HKS de 20703 segundos. En conclusión, los descriptores FINDDD y FPFH son los más recomendados para aplicaciones en tiempo real,

siendo FINDDD el que necesita menor tiempo computacional por una gran diferencia con los otros descriptores. En las implementaciones en C++ el descriptor FINDDD es cerca de 300 veces más rápido que el descriptor FPFH y en las implementaciones en Matlab el descriptor GDH es cerca de 3 veces más rápido que el descriptor HKS.

En los resultados de las partes claves mostrados en la Figura 2.5. Se concluye que FPFH obtuvo un rendimiento general más alto en términos de precisión, pero con un coste computacional mayor, lo que hace que el descriptor FINDDD, que alcanza resultados similares, sea más recomendable para aplicaciones de tiempo real. Además, los descriptores 3D fueron evaluados en unión con el descriptor 2D *Scale-invariant feature transform* (SIFT), sobre la imagen de profundidad, donde dicha combinación supera significativamente cualquiera de los descriptores individuales.

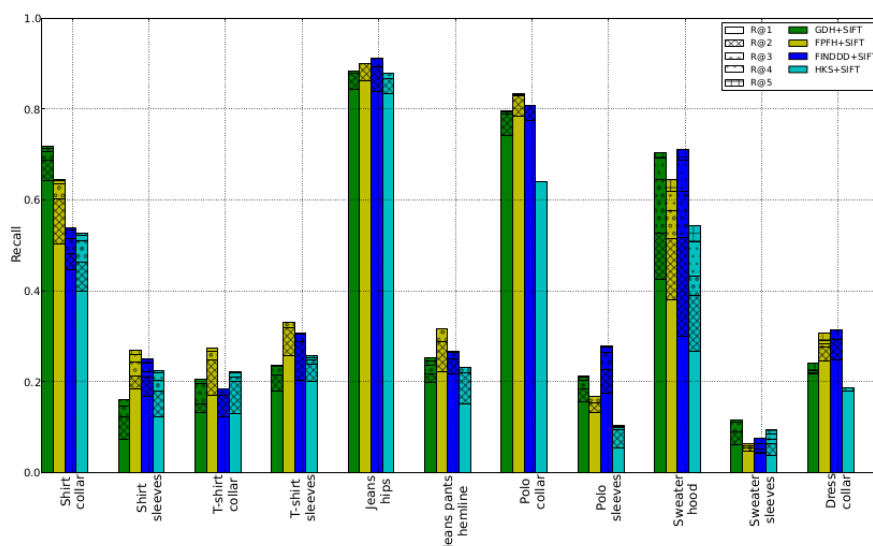


Figura 2.5: Recall de los descriptores 3D: GDH, FPFH, FINDDD y HKS, en unión con el descriptor SIFT en el reconocimiento de partes claves. El *recall* de cada de cada descriptor se separa en R@K, que indica el *recall* de detección en la imagen si se consideraran las *K* detecciones más altas. Fuente: Ramisa et al., 2014 [53].

Para realizar el reconocimiento completo de cada ropa, se proponen estrategias de clasificación, una de ellas es utilizar un clasificador *Support Vector Machine* (SVM) multi-capas y un modelo de *bag-of-words* [78, 79], utilizando características de bajo nivel, tales como: color, textura (SIFT), forma (FPFH) e información local y global de la silueta para cada clase de ropa. Otro enfoque es un método de clasificación, utilizando un conjunto de filtros de Gabor sobre una sola imagen [84], para generar características invariantes a la traslación, rotación y la escala. Se utiliza como información las arrugas y las superposiciones de tela en la ropa, detectados en base a las magnitudes y

orientaciones máximas, ofrecidos por los filtros de Gabor. Además, se ha utilizado la extracción de estructuras topológicas, es decir, los componentes de frontera mediante la extracción de contornos cerrados a partir de la imagen de profundidad [72].

El método más reciente realiza una clasificación multi-clase utilizando procesos gaussianos [69]. Esta técnica se basa en la percepción interactiva para el reconocimiento de ropa altamente arrugada, llevando a cabo diversas interacciones con la prenda para aumentar la confianza en la clasificación.

En la Tabla 2.1 se puede observar un resumen del estado-del arte de los reconocimientos en ropa sobre una superficie.

Tabla 2.1: Resumen del estado-del-arte de reconocimiento de ropa sobre una superficie. Fuente: Elaboración propia.

Publicación	Robot	Cant. brazos	Clases de ropa	Métodos	Desempeño
Wang11 [75]	PR2	2	<i>Sock</i>	LBP, HOG	R = 96.1 % P = 98.0 %
Ramisa12 [52]	WAM arm	1	<i>T-Shirt</i>	GDH y SIFT	TPR = 70 %
Ramisa13 [50]	WAM arm	1	<i>Shirt, T-Shirt, Jean, Sweater, Polo, Dress</i>	FINDDD	P = 78.8
Willimon13 [79]	-	0	<i>Shirt, Dress, Sock</i>	SVM, Información 2D/3D	TPR = 90 %
Yamazaki13 [84]	-	0	<i>Shirt, T-Shirt, Jean, Sweater, Polo, Dress, Hand-Towel, Turtleneck,</i>	Filtro de Gabor	ACC = 97.2
Ramisa14 [53]	-	0	<i>Shirt, T-Shirt, Jean, Sweater, Polo, Dress</i>	GDH, FPFH, FINDDD, HKS y SIFT	P = 82.4 ACC = 89.2
Twardon15 [72]	Mitsubishi Arms	2	<i>Shirt, Sock</i>	Reconocimiento por silueta	SR = 66.7 %
Sun16 [69]	Baxter	2	<i>Shirt, T-Shirt, Jean, Sweater, Towel</i>	Procesos Gaussianos	TPR = 70.8 %

2.2.2. Caso de Ropa Colgando

Cuando la ropa se encuentra colgando de un *gripper* se puede observar una mayor superficie que cuando se encuentra arrugada sobre la mesa; permitiendo realizar el reconocimiento de ropa con una representación de la silueta utilizando modelos geométricos [26, 27], métodos 2D que representan la silueta de la ropa y sus bordes [77] o realizando la extracción de características 3D volumétricas [36].

Además, dado que los brazos de los robots poseen la capacidad de girar la ‘muñeca’ en un mismo eje, existen métodos que aprovecharon esta habilidad y giraron la ropa para obtener la información en 3D desde múltiples direcciones. Una opción para mezclar la información de las múltiples vistas fue utilizar *Hidden Markov Model* (HMM), utilizando el detector de borde *Canny* para obtener la silueta de la ropa [7] o utilizando el contorno de los puntos de la imagen de profundidad [11]. Otro método desarrolló un *buffer* de la superficie aplanada de la silueta de la ropa con las distintas vistas y utilizó el *buffer* para el reconocimiento [27] (ver Figura 2.6).

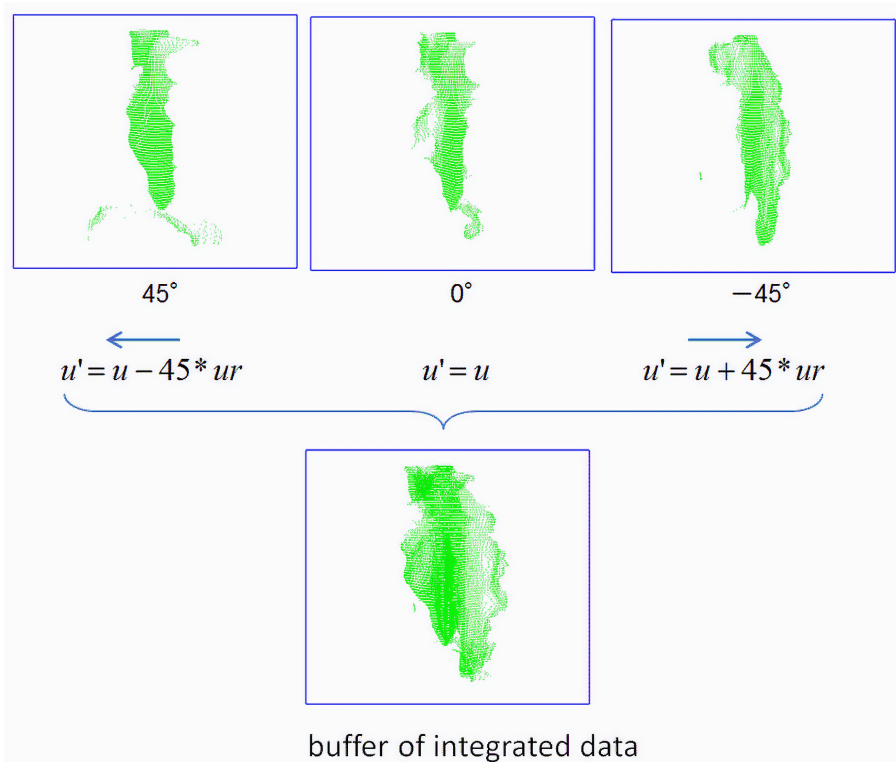


Figura 2.6: Buffer de imágenes de profundidad utilizada para obtener información de distintas vistas de la ropa mientras se encuentra colgando. Fuente: Kita et al., 2013 [27].

Considerando que la ropa se encuentra colgada, muchas investigaciones integraron el

reconocimiento de pose al reconocimiento de ropa para identificar los puntos óptimos a manipular dependiendo de la pose detectada y poder desplegar la ropa correctamente y de la forma más eficiente posible. En un comienzo para el reconocimiento de pose se realizó una simulación de los distintos estados de ropa, con el fin de ajustar los datos 3D y comprobar la consistencia entre las formas ajustadas y los datos observados para reconocer la pose [28]. Sin embargo, más adelante un método recurrente para realizar el reconocimiento de clase y de pose en conjunto, fue utilizar un clasificador jerárquico de dos etapas. Realizando en la primera etapa una clasificación para reconocer la clase de la ropa, y para la segunda etapa utilizar la salida de la primera etapa para seleccionar el clasificador apropiado.

En el enfoque de utilizar el clasificador jerárquico de dos etapas, uno de los métodos propuso utilizar el descriptor *Dense SIFT* sobre las imágenes de profundidad, para que sea invariante a la rotación y a la escala, y el clasificador *Basis Kernel function* (RBK) SVM y el clasificador *Linear Kernel function* SVM, para la primera y la segunda etapa respectivamente [35, 37] (ver Figura 2.7).

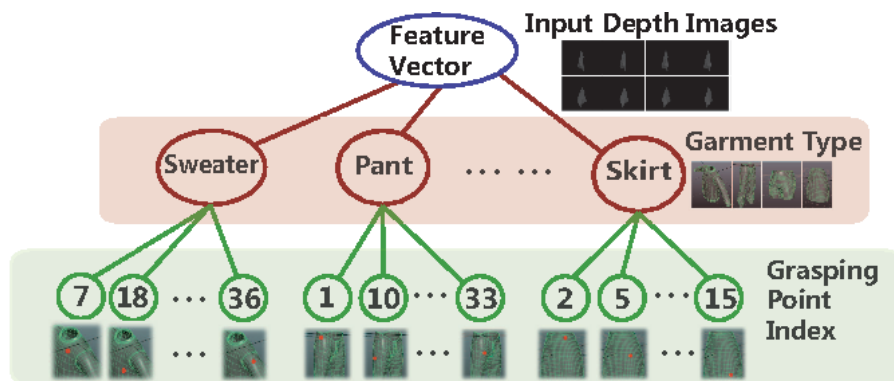


Figura 2.7: Reconocimiento de clase y de pose utilizando un clasificador jerárquico de dos etapas. En la primera etapa, se clasifica la ropa en una de las clases: *Sweater*, *Pant*, *Short* o *Shirt*, y en la segunda etapa se realiza un reconocedor de pose dependiendo de la clase reconocida. Fuente: Li et al., 2014 [35].

En Dumanoglou et al. [13, 14] se propone realizar el reconocimiento con *Random Forest*, el cual se encuentra entrenado con un vector de las imágenes de profundidad de cada estado del *framework partially observable Markov decision process* (POMDP) y girar la ropa actualizando el clasificador después de cada observación, y así decidir si se debe continuar girando la prenda para recoger más observaciones o tomar una decisión final. El *framework* POMDP se encuentra definido por la tupla $(S, A, O, T, P, R, \gamma, b_0)$, donde S es el conjunto de estados que corresponden a las 6 clases evaluadas $S = S_1, S_2, \dots, S_6$, A es el conjunto de acciones $A = A_{rotate}, A_1, A_2, \dots, A_6$,

A_{rotate} es el movimiento del gripper, mientras $A_1 - A_6$ es la decisión final de reconocimiento en el estado $S_1 - S_6$ respectivamente. Además, O es el conjunto de observaciones, T las probabilidades de transición condicionales, P la probabilidad de observación condicionales, R es la función de ganancia a las acciones y estados, γ el factor de descuento de recompensas a través del tiempo y b_0 es el estado creencia inicial. El diagrama general utilizado en este método se puede observar en la Figura 2.8.

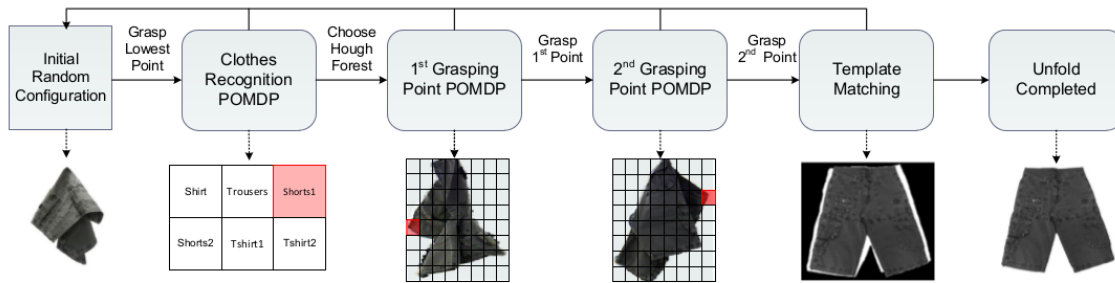


Figura 2.8: Diagrama de reconocimiento y desdoblado de ropa utilizando el framework POMDP. Este método realiza el reconocimiento utilizando *Random Forest*, el cual está entrenado con las imágenes de profundidad de cada estado del POMDP y gira la ropa actualizando el clasificador después de cada observación, y así puede decidir si se debe continuar girando la prenda para recoger más observaciones o se puede tomar una decisión final. Fuente: Doumanoglou et al., 2014 [13].

En el año 2015, con las mejoras de hardware y el avance en el área de Deep Learning, comenzaron investigaciones para el reconocimiento de ropa utilizando *Convolutional Neural Networks* (CNN) en ambas etapas [13]. Deep Learning posee un gran potencial para extraer atributos relevantes de los datos en bruto, donde cuanto más y mejores datos se tengan, mejor será el rendimiento del modelo. Para entrenar una red profunda desde cero, se recopila un conjunto de datos etiquetados muy amplio y se diseña una arquitectura de red que aprenda las características y el modelo. Las investigaciones que entrenaron la red desde cero utilizaron un motor de simulación de tela para aumentar la cantidad de datos [10,42] (ver Figura 2.9) o debieron disminuir la cantidad de clases y dejar un robot autónomamente manipulando ropa para obtener estos datos [17]. Sin embargo, otro método emplea el enfoque de transferencia del aprendizaje, un proceso que implica el ajuste de un modelo previamente entrenado. Se empieza con la red existente, y se le proporcionan datos nuevos que contienen clases previamente desconocidas [60].

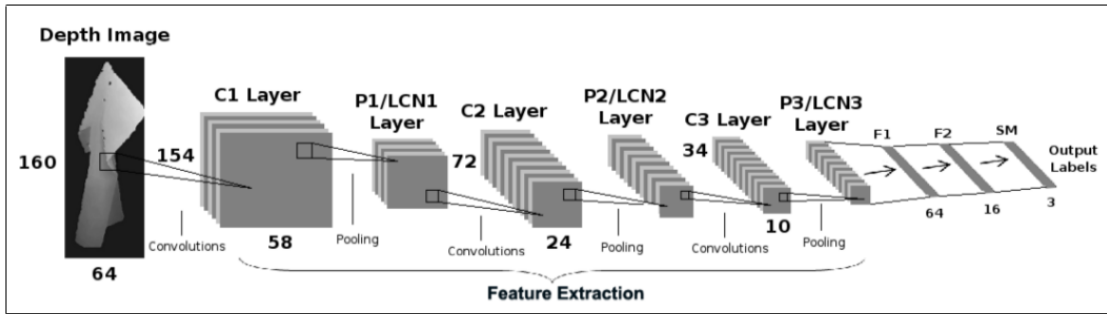


Figura 2.9: Arquitectura de la primera red profunda utilizada para el reconocimiento de ropa. Esta red fue entrenada desde cero utilizando un motor de simulación de tela para generar la cantidad de datos necesaria para el entrenamiento. Fuente: Mariolis et al., 2015 [42].

En Mariolis et al. [42] se realizó una comparación de algunos de los métodos nombrados y se encontró que los mejores resultados se obtuvieron utilizando una red profunda como se puede observar en la Tabla 2.2.

Tabla 2.2: Comparación de métodos realizada en Mariolis et al. [42] utilizando el enfoque de la ropa colgando. Fuente: Mariolis et al., 2015 [42]

<i>Método</i>	<i>Tasa de reconocimiento</i>
CNN	89.38 %
HOG-SVM	86.41 %
PT-RF	82.05 %

Finalmente, en la Tabla 2.3 se puede observar un resumen del estado-del-arte de los métodos realizando reconocimientos de ropa cuando se encuentra colgada por uno o dos brazos robóticos.

Tabla 2.3: Resumen del estado-del-arte de reconocimiento de ropa colgada. Fuente: Elaboración propia.

Publicación	Robot	Cant. brazos	Clases de ropa	Métodos	Desempeño
Kita02 [26]	-	0	<i>Shirt</i>	Reco. por silueta	No definido
Kita09 [28]	-	0	<i>Shirt</i>	Reco. por silueta	No definido
Cusumano11 [11]	PR2	2	<i>Shirt, T-Shirt, Jean, Skirt, Towel, Infant Clothing</i>	<i>Hidden Markov Model</i>	SR = 66
Willimon11 [77]	PUMA500 arm	1	<i>Shirt, T-Shirt, Jean, Short, Socks, Underwear</i>	Reco. por silueta, distancia Hausdorff y <i>k-NN</i>	$TPR \approx 72$
Bergstrom12 [7]	WAM arm	2	<i>T-Shirt</i>	<i>Hidden Markov Model, Canny edge detector</i>	TPR = 72
Kita13 [27]	HRP-2	2	<i>Shirt</i>	Representación cilíndrica, distancia geodesica	No definido
Doumanoglou14 [13, 14]	CloPeMa	2	<i>Shirt, T-Shirt, Jean, Short</i>	<i>Random Decision Forests</i>	ACC = 100
Li14IROS [36]	Baxter	2	<i>Jean, Short Sweater</i>	Reco. por silueta, <i>Hamming Distance</i> y <i>NN</i>	ACC = 81.9
Li14ICRA [35]	Baxter	2	<i>Jean, Short Sweater</i>	<i>dense SIFT, SVM</i>	ACC = 79.3
Mariolis15 [42]	CloPeMa	2	<i>Shirt, Jean, Hand-Towel</i>	<i>CNN</i>	TPR = 89
Gabas16 [17]	Barret's WAM arm	1	<i>Shirts, Trouser, Towel, Polo</i>	<i>CNN</i>	TPR = 92
Corona18 [10]	Barret's WAM arms	2	<i>T-Shirt, Jean, Jumper</i>	<i>CNN</i>	ACC = 96.8
StriaH18 [60]	Arm	1	<i>Shirt, T-Shirt, Pant, Short</i>	<i>CNN</i>	ACC = 92

2.3. Puntos de Manipulación

En el método de reconocimiento y clasificación de ropa es necesario determinar el punto de agarre para cuando la ropa se encuentra sobre una superficie y se desea levantarla con uno o ambos brazos del robot para separarla de la superficie.

Para tomar la ropa que se encuentra en una posición aleatoria sobre una superficie plana, se debe comenzar realizando una segmentación de la ropa, es decir, se debe eliminar de la imagen todo aquello que no sea la ropa a manipular, esto incluye: superficie, ambiente y otros objetos sobre la superficie, en caso de ser necesario. Al tener solo la información de la ropa a evaluar, se han utilizado distintos enfoques para detectar el punto de agarre: un punto central [39], el punto más alto [8] o el punto más arrugado [4, 51, 76]. Después de tener definido el punto de agarre se suele utilizar una orientación de agarre cercana a la perpendicular de la superficie.

El punto más arrugado [51] está normalmente asociado al punto que tiene la entropía más alta, es decir, aquel punto donde las orientaciones de las normales sean más diferentes en su vecindad. Para calcular este punto se realiza el histograma de los ángulos de las normales, utilizando los ángulos de inclinación y azimut definidos en la representación de coordenadas esféricas mostradas en la ecuación (2.2).

$$(\phi, \theta) = (\arccos(\frac{z}{r}), \arctan(\frac{y}{x})) \quad (2.2)$$

donde las variables x, y, z corresponden a las coordenadas 3D y la variable r es el radio de la representación esférica, definida como $r = \sqrt{x^2 + y^2 + z^2}$. Luego, para el modelo de distribución se crea un histograma bidimensional (64×64 bins) con los valores de inclinación y azimuth en una región local alrededor de cada punto, y su entropía se calcula con la ecuación (2.3).

$$H(X) = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (2.3)$$

En la Figura 2.10 se puede observar la imagen de “rugosidad” obtenida de dos prendas de ropa.

Otro método utilizado para calcular el punto más arrugado se basa en utilizar el descriptor SIFT, para capturar la distribución de las orientaciones de las normales dentro de un área alrededor de un píxel de interés (generalmente el punto más arrugado de la ropa). Este método funciona bastante bien para poder definir la orientación del *gripper* que sea más conveniente para el agarre [5].

Otro método realiza el reconocimiento de los cuellos de las camisas, utilizando un detector que combine las características de apariencia y geometría 3D (por ejemplo: SIFT y GDH

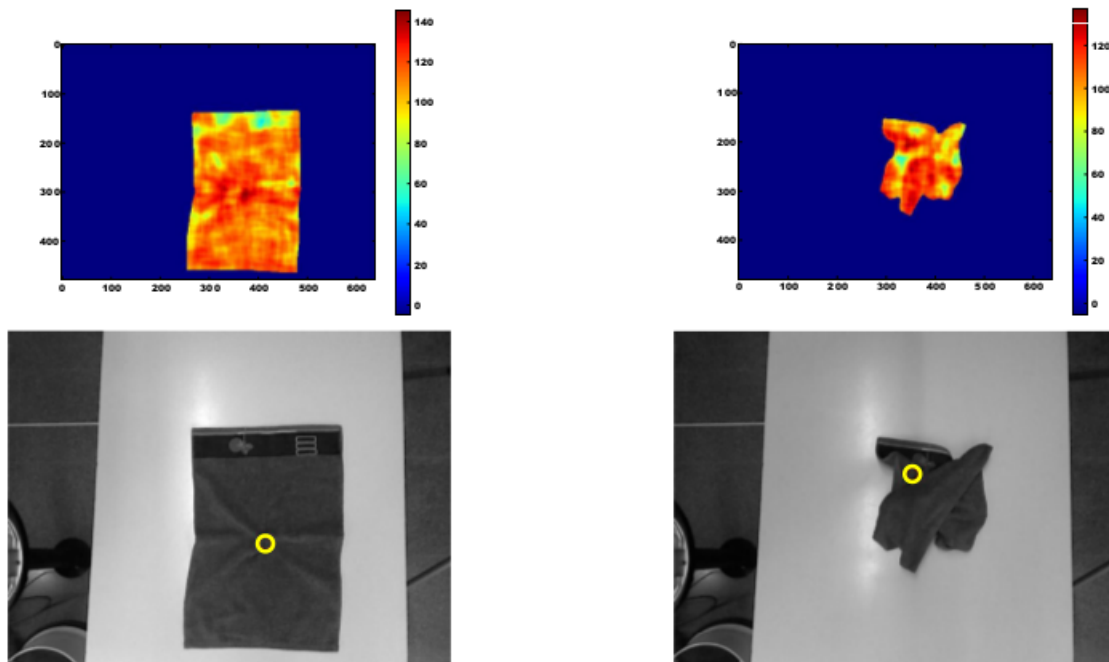


Figura 2.10: Ejemplo de imagen de rugosidad. En las imágenes superiores se puede observar el filtro de entropía aplicado sobre la ropa y en las imágenes inferiores el punto con el valor más alto de entropía, es decir, el punto más arrugado. Fuente: Alenya et al., 2012 [4]

respectivamente). Se utiliza un enfoque de ventana deslizante que cubre toda la imagen de profundidad, con ventanas de diferentes tamaños y formas, y utilizando un clasificador SVM se crea un mapa de probabilidad de la presencia del cuello de la camisa con todas las ventanas y se seleccionan los máximos locales. Para detectar el punto de agarre, se calcula el punto más arrugado donde exista mayor probabilidad de detección del cuello de camisa [52].

Un estudio reciente, desarrollo un método para seleccionar dos puntos de agarre de un trozo de tela, que permite a un robot de dos brazos estirar la tela directamente a un estado plano. El método propuesto utiliza un esquema de votación, el cual extrae un primer punto utilizando reconocimiento de esquina y realiza la búsqueda del segundo que tiene una alta probabilidad de estar conectado directamente con el primer punto [82, 83].

Con respecto a los distintos métodos que existen para la detección de puntos agarres, no se puede concluir que un método sea mejor que otro, ya que aún no existe una métrica para realizar una evaluación adecuada. Además, dependiendo de la aplicación a futuro, ya sea reconocimiento o doblado, es un área que aún se puede seguir investigando para detectar puntos que faciliten el objetivo.

2.4. Resumen y Análisis del Capítulo

Este capítulo presenta una amplia revisión bibliográfica de los distintos tipos de trabajos realizados en reconocimiento y manipulación de ropa. Primero, se presentan los robots de servicio con los cuales se han realizado investigaciones en reconocimiento, estirado, manipulación y/o doblado de ropa. Luego, se realiza una revisión bibliográfica de los sistemas de reconocimiento y clasificación de ropa, que se encuentran dividido en dos enfoques dependiendo de la posición de la ropa al realizar el reconocimiento. Finalmente, se presenta una revisión bibliográfica de los sistemas de detección de puntos de manipulación, principalmente cuando la ropa se encuentra en una posición aleatoria sobre una superficie y se debe agarrar un punto para levantarla.

Actualmente, son cuatro los robots de servicio principales en los cuales se han realizado investigaciones en las áreas de reconocimiento, estirado y/o manipulación de ropa; PR2 de Willow Garage, CloPeMa del proyecto *Clothes Perception and Manipulation*, HRP-2 de KAWADA Industries y Baxter de Rethink Robotics. Estos robots humanoides poseen diversos sensores y *grippers* que les permiten trabajar y manipular ropa precisamente.

En la revisión bibliográfica de los sistemas de reconocimiento y clasificación de ropa se han utilizado dos enfoques para resolver este problema. El primer enfoque consiste en detectar la ropa cuando se encuentra sobre una superficie. En este enfoque se utilizan los descriptores 3D para la detección de partes claves de cada clase. Estos descriptores muestran buenos resultados en partes claves que se diferencian claramente, tales como, la parte clave *Neck* de la clase *Shirt* y la parte clave *Waist* de la clase *Pant*. Sin embargo no es recomendable utilizarlo para partes claves que pueden ser confundidas hasta para el ser humano, como son la parte clave *Neck* de la clase *T-Shirt*.

Luego al intentar extraer características de una mayor superficie de la ropa, apareció el segundo enfoque que consiste en realizar el reconocimiento de la ropa cuando el robot la tiene colgando desde uno de sus brazos. Utilizando el despliegue de la ropa para realizar el reconocimiento de clase y de pose en conjunto. Realizando una clasificación para reconocer la clase de la ropa, y utilizando esta detección para seleccionar el clasificador de pose apropiado. Estos métodos obtuvieron mejores resultados que en el caso de detección sobre una superficie, pero no fue hasta el año 2015 cuando se comenzó a utilizar redes profundas que se obtuvieron los mejores resultados que se encuentran

actualmente en reconocimiento. Sin embargo, no existe una comparación clara entre los métodos analizados al no existir en muchos casos una base de datos.

En los sistemas de detección de puntos de manipulación cuando la ropa que se encuentra en una posición aleatoria sobre una superficie plana, se han utilizado distintos enfoques para detectar el punto de agarre: un punto central, el punto más alto, el punto más arrugado o el más cercano a la parte clave de la clase. En la actualidad no existe una métrica para comparar estos puntos, ya que están enfocados a lo que se desea realizar después de levantar la ropa, como es el caso de realizar una clasificación de ropa.

Capítulo 3

Metodología para Reconocimiento y Clasificación de Ropa

Para este trabajo de tesis se proponen dos métodos diferentes. El primer método permite reconocer los puntos de agarre para desplegar una ropa desde su posición inicial aleatoria sin información previamente almacenada. El segundo método consiste en un clasificador de ropa utilizando percepción continua. Si bien estos métodos no se utilizaron en conjunto, se utilizaron los conocimientos obtenidos del primer método para la definición del segundo.

3.1. Detección de Puntos de Agarre

Teniendo en cuenta que los puntos de agarre para desplegar una ropa suelen estar cerca de partes clave de su clase correspondiente (ver Figura 3.1), por ejemplo, la cintura de los pantalones (parte clave *Waist*) o el cuello en una camisa (parte clave *Neck*), se propone un método capaz de calcular los puntos de agarre para desplegar una prenda de ropa, que no requiere estos puntos almacenados previamente. Combinando los contornos de la región clave detectada y la información de rugosidad de la ropa para entregar dos puntos de agarre que puedan desplegar la ropa.



Figura 3.1: Ejemplo de detección de puntos de agarre para desplegar la ropa para las clases *T-Shirt*, *Pant* y *Shirt*, respectivamente. Estos puntos de agarre están cerca de las partes clave de cada una de estas clases; la parte clave *Waist* de la clase *Pant*, la parte clave *Neck* de la clase *Shirt* y la parte clave *Neck* de la clase *T-Shirt*. Fuente: Elaboración propia.

El método propuesto (ver Figura 3.2) se divide en tres partes: reconocimiento de partes claves (Sección 3.1.1), análisis de arrugas (Sección 3.1.2) y detección de puntos de agarre (Sección 3.1.3). En el reconocimiento de partes claves se detecta y reconoce la parte clave *Waist*, *Neck Shirt* o *Neck T-Shirt* a partir de la imagen de profundidad. En el análisis de arrugas se utiliza el filtro *Multiscale Vessel Enhancement Filter* (MVEF) para resaltar la estructura geométrica tubular de las arrugas. Finalmente, la detección de puntos de agarre unifica la detección de la parte clave con la información de las arrugas para entregar dos puntos de agarre que puedan desplegar la ropa.

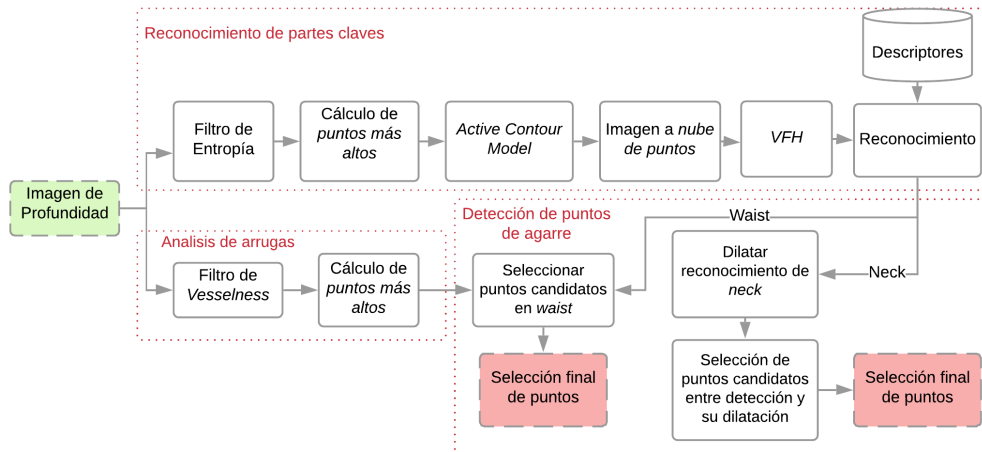


Figura 3.2: Diagrama de bloques del método de detección de puntos de agarre. Este método se divide en tres partes: reconocimiento de partes claves, análisis de arrugas y detección de puntos de agarre. Al realizar la detección y reconocimiento de la parte clave de la ropa en conjunto con la detección de los puntos representativos de las arrugas de la ropa, se utilizan estas dos informaciones para proponer puntos candidatos de agarre que sirvan para desplegar la ropa. Fuente: Elaboración propia.

3.1.1. Reconocimiento de Partes Claves

El reconocimiento de partes claves se realizó utilizando el descriptor *Viewpoint Feature Histogram* (VFH) en una región seleccionada por el método *active contours models* y realizando el calce con los k-vecinos más cercanos. Las partes que componen el reconocimiento de partes claves según el diagrama de la Figura 3.2 son: filtro de entropía, cálculo de puntos más altos, *active contour model*, imagen a nube de puntos, VFH y reconocimiento.

Filtro de entropía

Para tomar la ropa que se encuentra en una posición aleatoria sobre una superficie plana, se debe comenzar realizando una segmentación de la ropa, es decir, se debe eliminar de la imagen todo aquello que no sea la ropa a manipular, esto incluye: superficie, ambiente y otros objetos sobre la superficie, en caso de ser necesario.

Al tener sólo la información de la ropa a evaluar, se utilizó el filtro de entropía que al ser aplicado en la imagen de profundidad muestra donde las orientaciones de las normales sean más diferentes en su vecindad, es decir, muestra información de donde se encuentran las arrugas. La Figura 3.4 muestra el filtro de entropía aplicado en una imagen de profundidad de la clase *Shirt*.

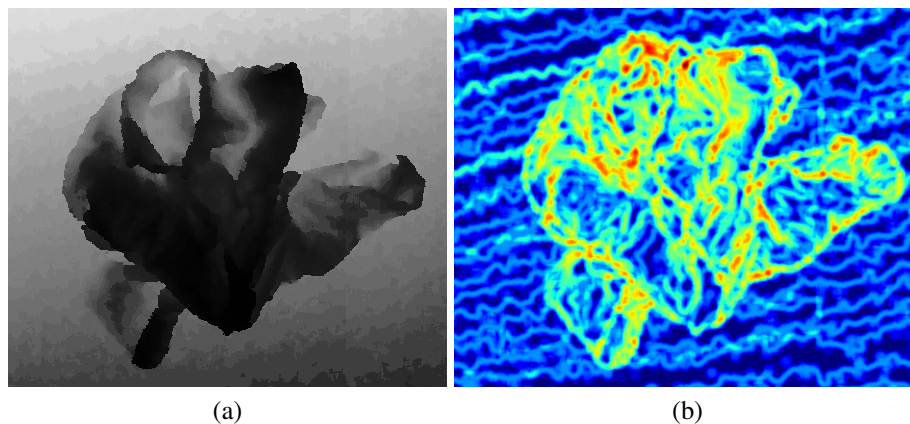


Figura 3.3: Filtro de entropía aplicado a la una imagen de la clase *Shirt*. Izquierda: Imagen de profundidad de la clase *Shirt*. Derecha: Filtro de entropía aplicado a la imagen de profundidad. Fuente: Elaboración propia.

Cálculo de puntos más altos

Al aplicar el filtro de entropía a la imagen de profundidad se obtiene la información de las arrugas de la ropa. Para obtener puntos candidatos para buscar las partes claves se buscaron los máximos locales de la imagen de entropía que se encontraran a una distancia mayor o igual de 50 píxeles.

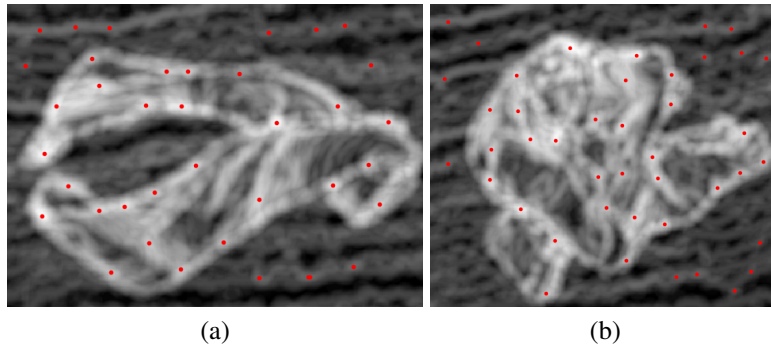


Figura 3.4: Cálculo de puntos más altos del filtro de entropía. a) clase *Pant*. b) clase *Shirt*. Fuente: Elaboración propia.

Active Contour Model

Para la evaluación del contorno, que corresponde a la selección del contorno de la parte clave de la ropa a detectar, se evaluó: detecciones de bordes y segmentación, tales como: *Felenszwabs's*, *SLIC*, *Quickshift*, *Compact watershed* y *Region Adjacency Graph*. Después de evaluar los distintos métodos, se seleccionó *active contours models* [23], por su adaptación a la parte clave de la ropa. Este método consiste en curvas definidas dentro de un dominio de la imagen que pueden moverse bajo la influencia de fuerzas internas provenientes de la propia curva y fuerzas externas calculadas a partir de los datos de la imagen. Las fuerzas internas y externas se definen de manera que el contorno se ajuste al límite del objeto u otras características deseadas dentro de una imagen. En la Figura 3.5 podemos ver dos ejemplos de cómo se adaptaron los *active contours models* a partir de un punto que estaba en la parte clave.

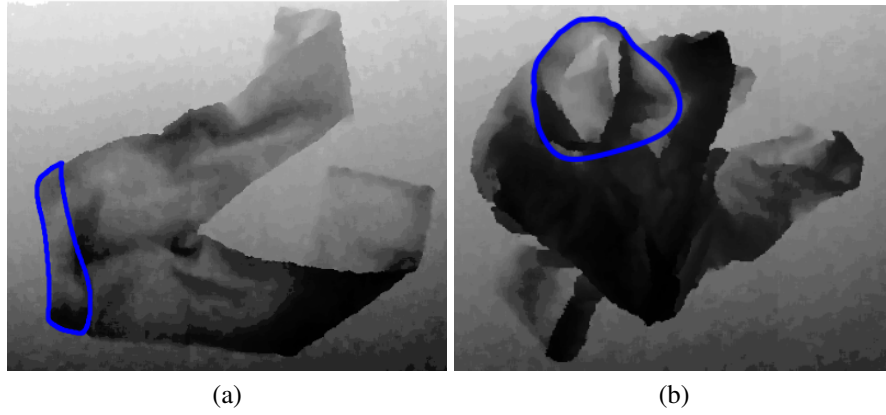


Figura 3.5: Detección de contornos de partes claves. Izquierda: Contorno de la parte clave *Waist* de la clase *Pant*. Derecha: Contorno de la parte clave *Neck* de la clase *Shirt*. Fuente: Elaboración propia.

Imagen a Nube de Puntos

Para poder ocupar la librería PCL, la cual proporciona la extracción de descriptores 3D, se debió transformar la imagen de profundidad a nube de puntos utilizando la ecuación (3.1).

$$\begin{aligned}
 X_{3D} &= \frac{X_{2D} - c_x}{f_x} \cdot D(X_{2D}, Y_{2D}) \\
 Y_{3D} &= \frac{Y_{2D} - c_y}{f_y} \cdot D(X_{2D}, Y_{2D}) \\
 Z_{3D} &= D(X_{2D}, Y_{2D})
 \end{aligned} \tag{3.1}$$

donde las variables X_{3D} , Y_{3D} y Z_{3D} corresponden a las coordenadas 3D de la nube de puntos, las variables X_{2D} , Y_{2D} corresponden a las coordenadas 2D de la imagen de profundidad D y los parámetros intrínsecos de la cámara se encuentran representados por $f_x = 525.0$, $f_y = 525.0$, $c_x = 319.5$ y $c_y = 239.5$. El resultado se puede observar en la Figura 3.6.

VFH

El descriptor *Viewpoint Feature Histogram* (VFH) [56] es un descriptor global, que está formado por un histograma de los componentes de las normales de la superficie del objeto. El histograma captura la forma del objeto y el punto de vista desde el cual se toma la nube de puntos. Primero, los

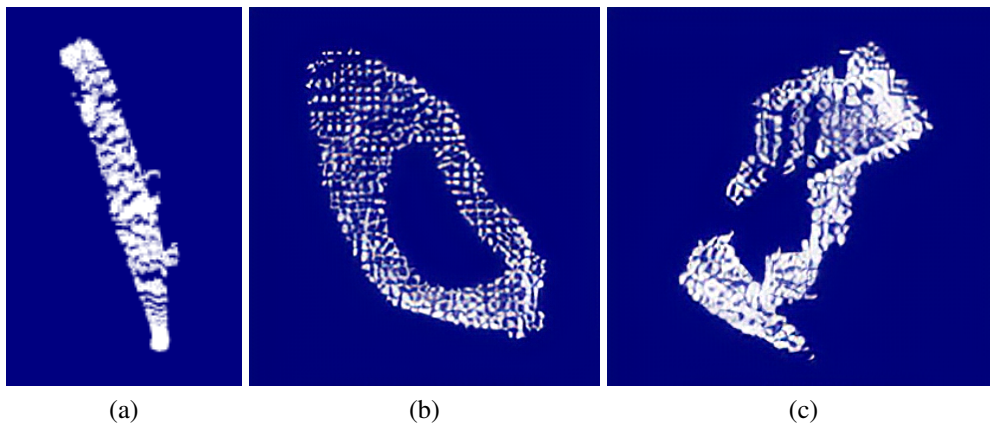


Figura 3.6: Partes claves en nube de puntos. a) Imagen de la parte clave *Waist*. b) Imagen de la parte clave *Neck Shirt*. c) Imagen de la parte clave *Neck T-Shirt*. Fuente: Elaboración propia.

ángulos α , ϕ y θ se calculan para cada punto según su normal y la normal del centroide de la nube de puntos c_i . El componente del descriptor que depende del punto de vista de la nube de puntos es un histograma de los ángulos entre el vector $p_c - p_v$ y cada normal. El otro componente es un descriptor FPFH [55] que está compuesto por un histograma de las distancias de los puntos en la nube al centroide de la nube y por la información del centroide de la nube de puntos. El descriptor VFH es un histograma compuesto que representa cuatro distribuciones angulares diferentes de las normales de superficie. En este trabajo, se usa la implementación de PCL [47], donde cada uno de esos cuatro histogramas tiene 45 *bins* y el componente del punto de vista tiene 128 *bins*, haciendo un total de 308 *bins*.

Reconocimiento

El reconocimiento de partes claves se divide en una fase de entrenamiento y una fase de prueba. La fase de entrenamiento guarda las características de las partes claves de las imágenes de entrenamiento en una base de datos y la fase de prueba utiliza las imágenes de prueba para reconocer las partes claves comparandolas con características que se encuentran en la base de datos. En la fase de entrenamiento (Algoritmo 1) se realizó la creación de la base de datos, que guardó los descriptores VFH de las partes claves de cada imagen del conjunto de entrenamiento. El algoritmo de la fase de entrenamiento tiene como entrada el conjunto de imágenes de entrenamiento con su respectivas máscaras (hechas a mano) que recalcan donde se encuentra la parte clave de cada imagen y la etiqueta de la clase correspondiente. Se seleccionó solo la parte clave de la imagen de

profundidad (línea 8) para transformarla a nube de puntos (línea 10) y calcular sus normales (línea 12). Con las normales se calculó el descriptor VFH de la parte clave (línea 14) para poder guardarlo en la base de datos (línea 21) que se utilizara en la fase de prueba.

Algoritmo 1: Algoritmo de reconocimiento de partes claves: Entrenamiento

```

1 function TrainGarmentParts (imgs, masks, labels);
   Input : Imágenes de profundidad, máscaras y labels del conjunto de entrenamiento
   Output: Descriptores del conjunto de entrenamiento
3  setVFH = [] ;
5  setLabels = [] ;
6  for img,mask,label in imgs, masks, labels do
8     region = img * mask ;
10    pcloud = img2pointcloud(region) ;
12    normals = getNormals(pcloud) ;
14    vfh = getVFHdescriptor(normals) ;
16    setVFH.append(vfh) ;
18    setLabels.append(label)
19 end
21 knn.train(setVFH, setLabels) ;

```

En la fase de prueba (Algoritmo 2) se reconocen las partes claves presentes en las imágenes del conjunto de prueba. En cada imagen se calcula el filtro de entropía (línea 2), y se calculan los máximos locales (línea 3) que serán utilizados para la evaluación del contorno utilizando el algoritmo *active contours models* (línea 6). Luego al tener los contornos, se utilizó k-NN (línea 10) sobre los descriptores VFH de los contornos (líneas 6-9), donde se seleccionó por un sistema de votación con los 10 vecinos más cercanos. En el caso de que dos o más clases tuvieran la misma votación, se agregaron las distancias de los vecinos que pertenecían a esas clases y se seleccionó la que tenía la distancia más corta.

3.1.2. Análisis de Arrugas

Se explota la estructura geométrica de los contornos de la ropa en imágenes de profundidad, mediante *Multiscale Vessel Enhancement Filter* (MVEF) propuesto por Frangi et al. [16]. Este filtro fue diseñado para detectar vasos sanguíneos en imágenes médicas, según su estructura geométrica tubular; pero cuando se utiliza en una imagen de profundidad de una ropa, resalta la forma tubular de las arrugas, lo que permite el cálculo del índice de rugosidad en la prenda.

Algoritmo 2: Algoritmo de reconocimiento de partes claves: Reconocimiento

```
1 Function RecognitionGarmentPart (img)
  Input : Imagen de profundidad
  Output: Clase de la parte clave reconocida
2  entropy = entropyFilter(img)
3  points = getPeaks(entropy)
4  listLabel = []
5  for p in points do
6    localRegion = getLocalRegion(p, img)
7    pcloud = img2pointcloud(localRegion)
8    normals = getNormals(pcloud)
9    vfh = getVFHdescriptor(normals)
10   neigh, dist = knn.findNearest(vfh, 10)
11   listLabel.append( selectLabel(neigh, dist))
12 end
13 return selectLabels(points, listLabel)
```

Filtro de *Vesselness*

El filtro se aplica a la imagen de profundidad que se encuentra en escala de grises y se calcula la métrica *vesselness*, que se obtiene mediante una aproximación de segundo orden entre el punto y su vecindad. Se utilizan los tres *Hessian eigenvalues* inferiores para modelar geoméricamente la región mediante un elipsoide, cuya forma y orientación están definidas por los valores propios y los vectores propios relacionados. Dependiendo de las magnitudes de valores propios, el filtro puede discriminar la región en las siguientes formas estructurales: aplanada, lineal y tubulares (ver Figura 3.12).

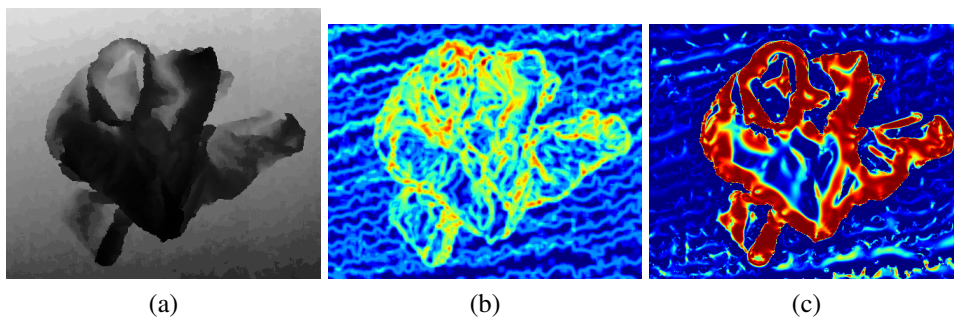


Figura 3.7: Resultados del filtro de entropía y de *vesselness*. Izquierda: Imagen de profundidad. Centro: Filtro de entropía. Derecha: Filtro de *vesselness*. Fuente: Elaboración propia.

Cálculo de puntos más altos

El filtro aplicado resalta la forma tubular de las arrugas y los pliegues generados por la ropa en la imagen de profundidad. Para considerar los patrones de diferentes tamaños, la estrategia del filtro es aplicar el proceso descrito en múltiples niveles de escala, eligiendo los puntos máximos de la imagen *vesselness* como los puntos candidatos de agarre.

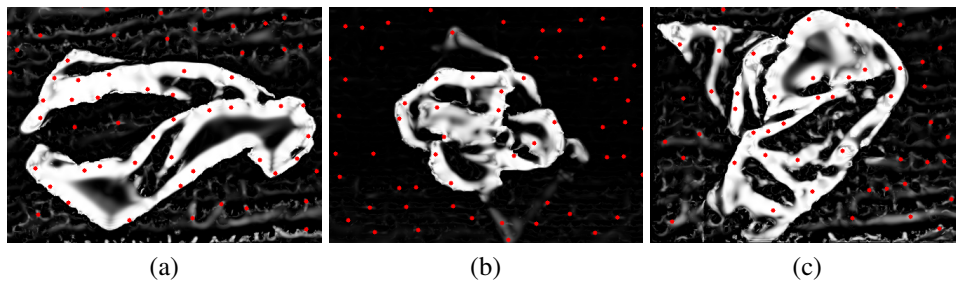


Figura 3.8: Cálculo de puntos más altos con el filtro de *vesselness*. a) Puntos más altos en la imagen del filtro de *vesselness* de la clase *Pant*. b) Puntos más altos en la imagen del filtro de *vesselness* de la clase *Shirt*. c) Puntos más altos en la imagen del filtro de *vesselness* de la clase *T-Shirt*. Fuente: Elaboración propia.

3.1.3. Detección de Puntos de Agarre

Tomando la información de la detección de la parte clave y los máximos locales de la imagen *vesselness*, los puntos de agarre se extraen dependiendo de la clase detectada. Para la clase *Neck* se realizan: dilatar reconocimiento de la clase *Neck*, selección de puntos candidatos entre detección y su dilatación y selección final de puntos. Para la clase *Waist* se realizan: seleccionar puntos candidatos en *Waist* y selección final de puntos.

Dilatar reconocimiento de la clase *Neck*

Se realiza la dilatación del contorno de la parte clave *Neck*, que consiste en aplicar una dilatación morfológica a la máscara que representa la parte clave *Neck* (Algoritmo 3, línea 3).

Selección de puntos candidatos entre detección y su dilatación (*Neck*)

Como entrada a esta parte del algoritmo, se obtienen como puntos candidatos los puntos máximos locales de la imagen del filtro de *vesselness* (Algoritmo 3, línea 2). Luego, con la dilatación de la máscara que representa la parte clave *Neck*, se seleccionan aquellos puntos que están dentro del contorno dilatado y fuera de la máscara original (Algoritmo 3, línea 4).

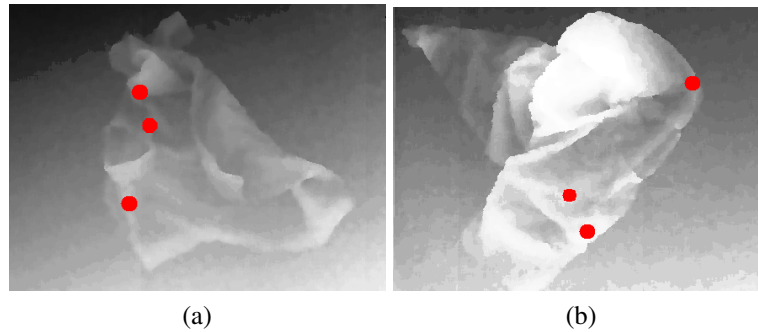


Figura 3.9: Selección de puntos candidatos de agarre de la clase *Shirt* y *T-Shirt*. Fuente: Elaboración propia.

Selección final de puntos (*Neck*)

Para seleccionar los puntos finales entre los puntos que fueron seleccionados, se crean líneas entre todos los pares de puntos candidatos, seleccionando el par de puntos en el cual su línea se encontraba más cerca del centro del contorno de la parte clave detectada (Algoritmo 3, líneas 6-15).



Figura 3.10: Selección final de puntos de agarre de la clase *Shirt* y *T-Shirt*. Fuente: Elaboración propia.

Seleccionar puntos candidatos en *Waist*

Para la selección de los puntos candidatos de la clase *Waist* se obtienen los puntos máximos locales de la imagen del filtro de *vesselness*. El algoritmo filtra todos los puntos que se encuentren fuera del contorno de la detección de la parte clave *Waist* y los puntos restantes serán puntos candidatos a ser puntos de agarre (Algoritmo 4, línea 2).

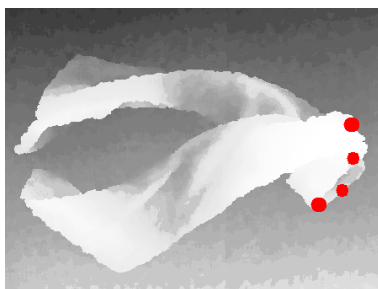


Figura 3.11: Selección de puntos candidatos de agarre de la clase *Waist*. Fuente: Elaboración propia.

Selección final de puntos (*Waist*)

Para seleccionar los puntos de agarre, se buscan los puntos candidatos más cercanos a los puntos extremos del contorno de la parte clave detectada (Algoritmo 4, líneas 3-15).

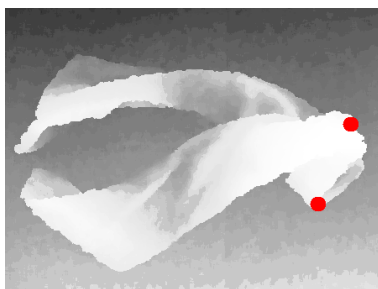


Figura 3.12: Selección final de puntos de agarre de la clase *Waist*. Fuente: Elaboración propia.

En la Sección 4.1 se presentan resultados de la aplicación del sistema de detección de puntos de agarre.

3.2. Clasificación de Ropa Mediante Percepción Continua

Este método tiene como objetivo clasificar la ropa utilizando la información extraída de una secuencia de imágenes. Para esto, se propone un enfoque de percepción continua basado en el

Algoritmo 3: Selección de puntos de agarre: Clases *T-Shirt* y *Shirt*

```
1 Function selectPointsNeck (mask, points)
2   center_point = getCenterPoint(mask) ;
3   mask2 = dilate(mask);
4   filteredPoints = points in mask2 and not in mask ;
5   min_dist = -1 ;
6   for p1 in filteredPoints do
7     for p2 in filteredPoints do
8       straight_line = getStraightLine(p1, p2);
9       dist = distancePoint2Line(center_point, straight_line) ;
10      if dist < min_dist or min_fist == -1 then
11        | pointA, pointB = p1, p2 ;
12      end
13    end
14  end
15  return pointA, pointB;
```

Algoritmo 4: Selección de puntos de agarre: Clase *Pant*

```
1 Function selectPointsWaist (mask, points)
2   filteredPoints = points in mask ;
3   pe1, pe2 = getExtremePoint(mask) ;
4   min_dist = -1 ;
5   for p1 in filteredPoints do
6     for p2 in filteredPoints do
7       d1 = getEuclideanDist(p1, pe1) + getEuclideanDist(p2, pe2);
8       d2 = getEuclideanDist(p1, pe2) + getEuclideanDist(p2, pe1);
9       dist = min(d1, d2) ;
10      if dist < min_dist or min_fist == -1 then
11        | pointA, pointB = p1, p2 ;
12      end
13    end
14  end
15  return pointA, pointB;
```

uso de múltiples imágenes de profundidad de diferentes sensores, mientras que un brazo robótico mueve una prenda sobre y desde una superficie plana. La Figura 3.13 representa el proceso de percepción continua propuesto dentro de una tarea de clasificación robótica. En este proceso, el robot captura la información RGB-D de la ropa de los dos sensores que tiene en sus brazos mientras manipula la ropa y la deja colgando, este procedimiento se realiza una sola vez. Luego se extraen las características de cada imagen capturada de la secuencia para realizar el reconocimiento y dejar la ropa en la caja correspondiente de la clase reconocida. Este diagrama incluye el proceso propuesto de reconocimiento de percepción continua que se puede ver en detalle en la Figura 3.14.

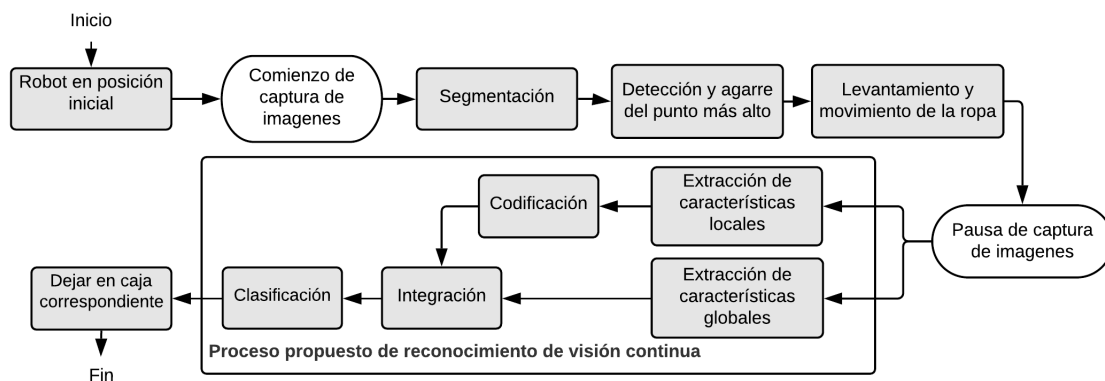


Figura 3.13: Diagrama de clasificación de ropa utilizando percepción continua. El robot captura la información RGB-D de la ropa mientras la agarra de la mesa, la manipula y la deja colgando. Luego se extraen las características de cada imagen capturada de la secuencia y se realiza el reconocimiento para clasificar la ropa en una de las siguientes 5 clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*. Finalmente, el robot deja la ropa en la caja correspondiente de la clase reconocida. Fuente: Elaboración propia.

El método propuesto de clasificación mostrado en la Figura 3.14 consta de cuatro módulos: (1) extracción de características locales y globales, (2) codificación de características locales utilizando LLC y GPLVM, (3) integración de las características y (4) clasificación con *Locality Constrained Group Sparse Representation* (LGSR). Las características locales entregan información distintiva sobre las arrugas de la ropa, mientras que las características globales capturan la forma de la prenda a medida que se mueve. Se asume que la prenda de ropa a reconocer es el único objeto sobre la mesa; por lo tanto, se utilizó una segmentación sobre la nube de puntos en función del umbral de la altura de la mesa, que separa la ropa del resto de la escena.

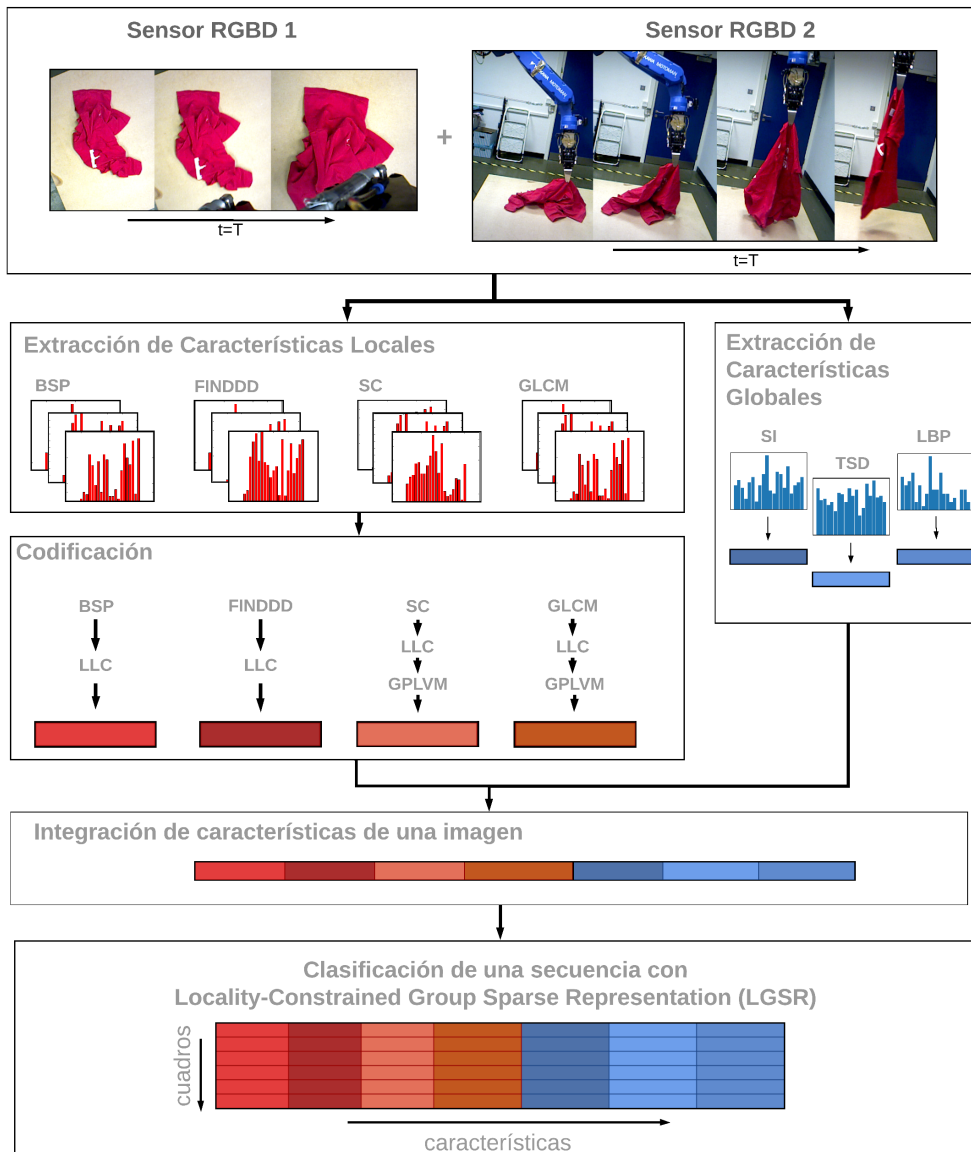


Figura 3.14: Diagrama de reconocimiento utilizando percepción continua. Las características locales y globales de una imagen se concatenan en un único vector de características y se unen con los vectores de la secuencia para ser utilizadas en la clasificación LGSR. Fuente: Elaboración propia.

En el primer módulo del proceso de percepción continua, las características locales (ver Sección 3.2.1) fueron seleccionadas por la información que contribuyen sobre las arrugas [53, 68], característica que poseen solo los objetos deformables. En cambio, las características globales (ver Sección 3.2.2) se seleccionaron por la información que aportan sobre la forma de la ropa y que han sido utilizadas en el estado-del-arte [51, 64, 68].

En el Capítulo 4 se analiza un estudio exhaustivo sobre las contribuciones de cada una de las características locales y globales, demostrando su contribución al clasificar las prendas (ver experimentos en la Sección 4.2.3).

3.2.1. Características Locales

Local B-Spline Patch

B-Spline Patches (BSP) es una extensión de *B-spline curve* en una superficie. BSP define una sección, en este caso llamado ‘parche’, como una función de dos variables con una matriz de puntos de control, que permite modelar distintos tipos de topología debido a su flexibilidad [68]. En este trabajo, se utilizó BSP [15] para describir las arrugas de la ropa mediante el muestreo de parches en la cresta de las arrugas detectadas utilizando *Shape Index* (ver Sección 3.2.2). BSP representa la superficie de un parche $P(x, y)$ calculado por la ecuación (3.2).

$$P(x, y) = \sum_{i=1}^{n+1} \sum_{j=1}^{m+1} \Omega_{i,j} \alpha_{i,k}(x) \beta_{j,l}(y) \quad (3.2)$$

donde $\Omega_{i,j}$ representa el punto de control en la fila i y la columna j . $\alpha_{i,k}(x)$ y $\beta_{j,l}(y)$ son las funciones básicas en el plano $x - y$. Se usaron los puntos de control Ω como la representación de la superficie local, que es una representación del subconjunto del conjunto total de puntos de la superficie P . Dado que los puntos de control se distribuyen uniformemente en el plano $x - y$, sólo se utilizaron 25 valores de profundidad para el descriptor BSP [68].

Fast Integral Normal 3D

El descriptor *Fast Integral Normal 3D* (FINDDD) [50] representa la distribución de las orientaciones de las normales 3D en una región alrededor de un punto de interés en una nube de puntos estructurada. Para generar el descriptor FINDDD se calculan las normales de la nube de puntos, utilizando imágenes integrales para acelerar el proceso. Luego, la nube de puntos se divide en subregiones, calculando para cada una de ellas un descriptor mediante la construcción de histogramas de orientación normal. Sin embargo, en lugar de usar las coordenadas esféricas para

representar los ángulos, el descriptor FINDDD se distribuye regularmente en toda la semiesfera utilizando coordenadas cartesianas. Esto evita la concentración alrededor del polo norte (elevación máxima) y el área desigual en las partes del histograma causada por la representación angular.

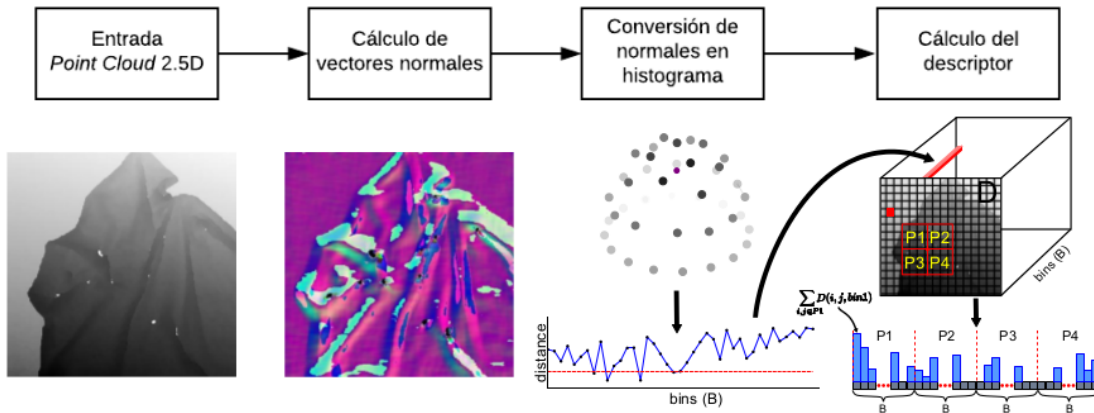


Figura 3.15: Diagrama del descriptor FINDDD. Fuente: Ramisa et al., 2013 [50]

Para calcular el descriptor FINDDD, la imagen integral I de la imagen de profundidad M es calculada utilizando la ecuación (3.3).

$$I_{i,j} = M_{i,j} + I_{i-1,j} + I_{i,j-1} - I_{i-1,j-1} \quad (3.3)$$

donde i y j son los sub-índices para recorrer las filas y columnas de la imagen. Para calcular el descriptor FINDDD se divide el área en subregiones y por cada subregión se calcula un histograma de orientación con las normales. A continuación, los histogramas de orientación de todas las subregiones en el área se concatenan para formar el vector descriptor FINDDD, como se muestra en la ecuación (3.4). El proceso completo del descriptor FINDDD puede ser observado en la Figura 3.15.

$$FINDDD = P_1 \cup P_2 \cup P_3 \cup P_4 \quad (3.4)$$

Shape Context

El descriptor *Shape Context* (SC) [6] describe la relación entre dos puntos que se encuentran en la misma forma del objeto. En este caso entrega información de dos puntos de una arruga

mediante una distancia logarítmica-polar y clasifica estos valores en un histograma de 12×5 bins. El histograma proporciona una caracterización discriminatoria global de la forma en un descriptor local, ya que las distancias se calculan con respecto a otros puntos de la forma. SC por lo tanto describe las estructuras con un descriptor invariante a la translación.

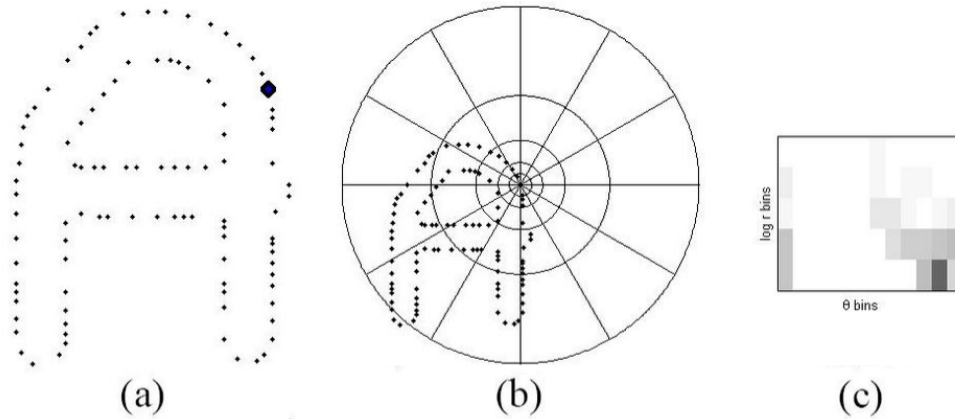


Figura 3.16: Cálculo del descriptor SC de una forma de ejemplo. a) son los puntos de borde de una forma. (b) es el diagrama de los contenedores log-polares utilizados para calcular el SC. (c) es el SC para el punto marcado con un círculo en (a). Fuente: Leen et al., 2001 [6]

Para un punto p_i en la forma, se calcula el histograma SC_i de las coordenadas relativas de los $n - 1$ puntos restantes. Se define como el contexto de forma de p_i . Los contenedores normalmente se consideran uniformes en el espacio log-polar. La definición del histograma como el descriptor SC se muestra en la ecuación (3.5). El proceso completo del descriptor SC puede ser observado en la Figura 3.16.

$$SC_i(k) = \{(q - p_i) \in bin(k)\} \quad (3.5)$$

Grey Level Co-occurrence Matrix + SVD

Grey level Co-occurrence Matrix (GLCM) [9,20] calcula la matriz de co-ocurrencia calculando la frecuencia con la aparece un píxel con determinado nivel de gris (intensidad de escala de grises) en cualquiera de las ocho direcciones definidas (0, 45, 90 y 135 grados).

GLCM determina la relación de píxeles en términos de distancia y ángulo. Aunque existen variantes de los métodos de análisis de textura, el algoritmo GLCM es uno de los más comúnmente adoptados para encontrar información de textura en imágenes de reconocimiento de objetos [58].

A partir de la matriz de co-ocurrencia se calcula la SVD (descomposición del valor singular), generando tres matrices (U , S y V). U y V representan los vectores singulares izquierdo y derecho de la matriz de la imagen, y S es una matriz diagonal con valores singulares. Luego, la normalización de L^1 se aplica en la matriz diagonal, utilizando este vector como el descriptor GLCM.

$$GLCM = \|S\|_{L^1} \tag{3.6}$$

3.2.2. Características Globales

Shape Index (SI)

El descriptor *Shape Index* (SI) [29] posee bastante relevancia al trabajar con ropa, ya que clasifica la región de cada arruga en una de 9 categorías (ver Figura. 3.17). Las arrugas son una característica única de los objetos deformables y que varían dependiendo del material y grosor del que esté hecha la ropa. En el reconocimiento de ropa, la forma “*ridge*” que entrega este descriptor es crítica en el análisis y descripción de arrugas.

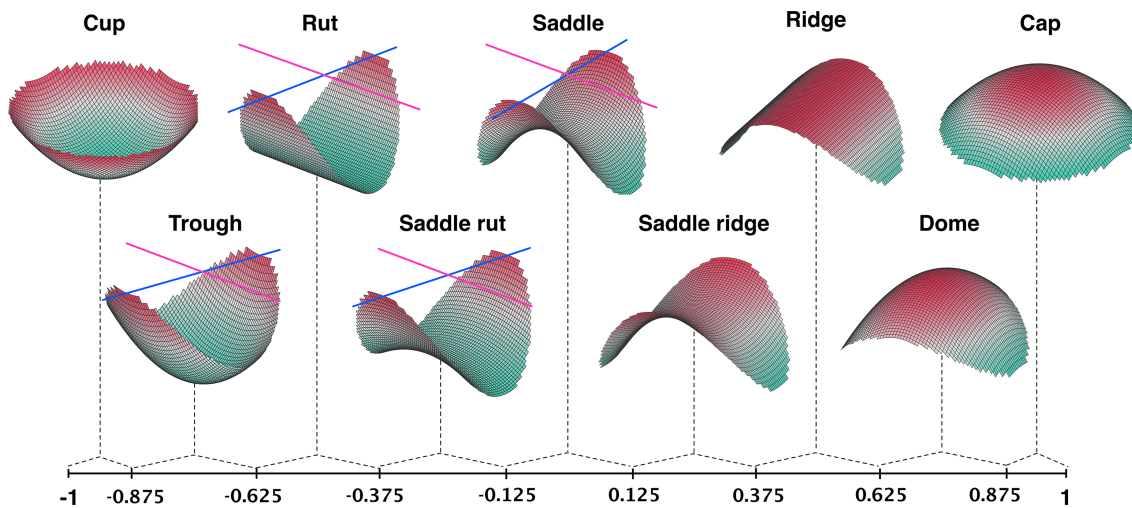


Figura 3.17: Ilustración de *Shape Index* dividida en las nueve categorías: *spherical cup*, *rough*, *rut*, *saddle rut*, *saddle*, *saddle ridge*, *ridge*, *dome* y *spherical cap*. Fuente: Thakku et al. [71].

El valor SI de un punto p se puede calcular utilizando la ecuación (3.7).

$$S^p = \frac{2}{\pi} \tan^{-1} \left[\frac{k_{min}^p + k_{max}^p}{k_{min}^p - k_{max}^p} \right] \quad (3.7)$$

donde k_{min} y k_{max} son las curvaturas mínima y máxima respectivamente del punto p . Para analizar la forma de la superficie, se calcula el SI del mapa de profundidad ajustado por B-Spline (Sección 3.2.1) y se aplica un filtro *majority rank*. Este filtrado no lineal elimina las clasificaciones de superficies atípicas y puede ajustarse para producir una clasificación relativamente limpia de los tipos de formas sobre la superficie. El descriptor SI corresponde a un histograma de 9 *bins*, donde cada *bin* representa cada uno de las categorías de *Shape Index*.

Topology Spatial Distances (TSD)

Topology Spatial Distances (TSD) [64] caracteriza las arrugas según sus atributos geométricos calculando la distancia entre cada “*ridge point*” ($R = \{r_1, \dots, r_{n_r}\}$) y sus puntos de contorno más cercanos ($W = w_1, \dots, w_{n_w}$). TSD es un histograma bidimensional compuesto por $\{TSD_1^{xy}, \dots, TSD_{n_r}^{xy}\}$ y $\{TSD_1^d, \dots, TSD_{n_r}^d\}$, que representan el ancho y altura de las arrugas. Estos componentes son calculados utilizando las ecuaciones (3.8) y (3.9).

$$TSD_i^{xy} = \min_{i \in n_w} \|r_i, w_j\|^2 \quad (3.8)$$

$$TSD_i^d = d_i^r - d_{arg_j \min \|r_i, w_j\|^2}^c \quad (3.9)$$

donde d_i^r y d_j^w son los valores de profundidad de r_i y w_j , respectivamente. Para TSD^{xy} y TSD^d , los *bins* se fijan en 10 intervalos uniformes que van desde 5 a 50 (la unidad en el plano $x - y$ es el píxel y en el eje de profundidad es el milímetro).

Después de aplicar la normalización L^2 , se obtiene un histograma unidimensional de 100 *bins*, que corresponde al descriptor global TSD, que se obtuvo al re-organizar el histograma bidimensional TSD de 10×10 *bins*.

Local Binary Patterns (LBP)

El operador *Local binary patterns* (LBP) [49] es una poderosa característica que se utiliza para describir la textura y la forma del contenido en una imagen. LBP etiqueta los píxeles de una imagen al utilizar un umbral en la vecindad de cada píxel. El umbral corresponde al valor de cada píxel en cada caso, donde al ser aplicado en toda la vecindad se tiene como resultado un número binario. El descriptor global LBP corresponde al histograma de los números binarios obtenidos.

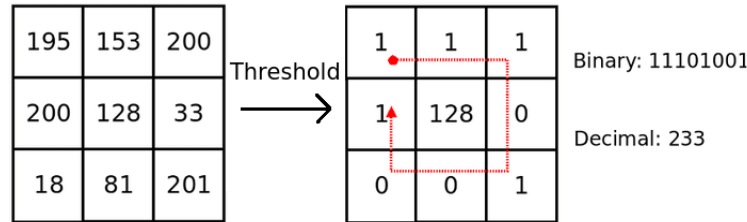


Figura 3.18: Explicación del descriptor LBP. Fuente: Elaboración propia

En esta implementación, se seleccionaron 58 patrones que fueron extraídos en múltiples escalas de la imagen de profundidad. Para la extracción de características a escala múltiple, se construyó una pirámide gaussiana de 3 capas utilizando un factor de subdivisión de 2 y un parámetro de suavizado gaussiano ($\sigma = 0.375$). Finalmente, se obtuvo un histograma unidimensional de tamaño 174 (58×3), correspondiente descriptor global LBP.

3.2.3. Codificación de Características Locales

El método propuesto de reconocimiento de ropa incluye un proceso que integra las características globales y locales, unificando la información continua de las imágenes de una secuencia. Siendo necesario transformar el conjunto de vectores de cada descriptor local en un descriptor global para que sea posible la integración. La codificación *Locality-constrained Linear Coding* (LLC) fue utilizado para la transformación, ya que ha demostrado que se desempeña de manera efectiva y eficiente en el reconocimiento de objetos y en reconocimiento de ropa según [68]. La codificación *Gaussian Process Latent Variable Model* (GPLVM) se utilizó para comprimir la información proporcionada por las características locales SC y GLCM, después de la técnica de codificación LLC, ya que demostró tener mejores resultados en compresión que la técnica de codificación LLC (ver Sección 4.2.3).

Locality-Constrained Linear Coding

Locality-constrained Linear Coding (LLC) [74] es un algoritmo de codificación que utiliza la restricción de localidad para proyectar cada descriptor en su sistema de coordenadas locales. LLC fue utilizado para transformar el conjunto de vectores de cada descriptor local en un descriptor global con elementos dispersos. Esta codificación ha demostrado que se desempeña de manera más efectiva y eficiente en el reconocimiento de objetos y en reconocimiento de ropa según [68]. Esta técnica de codificación se usó para cada una de las características locales (BSP, FINDDD, SC y GLCM), como se puede ver en las Figuras 3.13 y 3.14.

Sea $X = [x_1, x_2, \dots, x_N]$ un conjunto de características de un descriptor local extraídos de una imagen, donde N es el número total de características locales en la imagen, y $B = [b_1, b_2, \dots, b_M]$ un *codebook* con M entradas utilizando un *clustering K-means*. El coeficiente de codificación c_i del descriptor local x_i se obtiene al resolver el programa de optimización en la ecuación (3.10), con c_{ij} el coeficiente con respecto al *cluster* b_j .

$$c_i = \underset{c \in \mathbb{R}^n}{\operatorname{argmin}} \|x_i - Bc\|^2 + \lambda \|d_i \odot w \odot c\|^2 \quad (3.10)$$

$$s.t. c_1 = 1, w_1 = 1, \forall i$$

$$w_j = \frac{1}{1 + e^{-\sigma(n_j - N/K)}} \quad (3.11)$$

donde, d_i es la distancia euclidiana entre x_i y los átomos del *codebook*, $w = [w_1, \dots, w_K]$ el peso de los átomos, n_j el número de descriptores asignados al j -*cluster* y \odot el producto punto entre ambos elementos.

Con el coeficiente de codificación de todas las características locales en una imagen, se calculó *sum-pooling* para cada dimensión de los códigos de entrada $[c_1, \dots, c_N]$ para obtener una representación global L , calculada según la ecuación (3.12).

$$L_j = \sum_{i=1}^N c_{ij} \quad (3.12)$$

siendo $j = (1, \dots, K)$, con K el tamaño del cluster.

Gaussian Process Latent Variable Model

Gaussian Process Latent Variable Model (GPLVM) [32] es un algoritmo eficaz de reducción de dimensionalidad no lineal. Utilizado en este trabajo para comprimir la información proporcionada por las características locales. GPLVM es un modelo de variable latente que generaliza el análisis de componentes principales y proporciona un mapeo no lineal para reproducir las muestras transformadas del espacio de variable latente al espacio de observación mediante la imposición de un proceso gaussiano antes de la función de mapeo. Esta técnica de codificación se utilizó en las características locales: SC y GLCM, después de la técnica de codificación LLC, como se puede ver en las Figuras 3.13 y 3.14.

Sea $L = [l_1, \dots, l_N]$ la matriz que contiene un conjunto de N observaciones. Donde cada observación l_i se describe mediante un vector de características D -dimensionales $l_i \in R^D$. El objetivo de GPLVM es obtener las variables latentes/desconocidas $Z = [z_1, \dots, z_N]$ correspondientes a cada elemento de datos $l_i (i = 1, 2, \dots, N)$. GPLVM define un mapeo generativo de las variables latentes z_i a sus correspondientes variables observadas l_i , que se rigen por un grupo de *Gaussian Processes* (GPs) $l_i = g(z_i) + \epsilon$, donde $g = [g_1, \dots, g_D]^T$ se asume que es un grupo de D GPs, y ϵ un ruido gaussiano independiente con media cero y covarianza $\sigma^2 I$. Lo que significa que las observaciones llegan a tener una probabilidad gaussiana [22].

$$P(L|g, Z) = \prod_{n=1}^N \mathcal{N}(l_i | g(z_i), \sigma^2 I) \quad (3.13)$$

Supongamos que cada GP $g_i (i = 1, \dots, D)$ tiene la misma función de covarianza $k(\cdot, \cdot)$, entonces la probabilidad de datos definida por la ecuación (3.13) puede ser marginado con respecto al prior GP con respecto a todos los GP, dando lugar a la probabilidad general marginada de las observaciones L , mostradas en la ecuación (3.14).

$$P(L, Z) = \frac{1}{(2\pi)^{N/2} |K|^{1/2}} \exp\left(-\frac{1}{2} \text{tr}(K^{-1} L L^T)\right) \quad (3.14)$$

donde $K = K_{ZZ} + \sigma^2 I$ es el kernel sobre las variables latentes Z . El modelo de aprendizaje se implementa al maximizar la probabilidad de datos marginados anterior con respecto a las variables

latentes Z y los parámetros del kernel k .

Por lo general, GPLVM usa un modelo lineal, un modelo *kernel based regression* (KBR) o un modelo *multilayer perception* (MLP) para restringir los puntos latentes a una función uniforme de los puntos de datos correspondientes, lo que obliga a los puntos cercanos del espacio de datos a estar cerca en el espacio latente. En este caso, utilizamos un kernel gaussiano, que reduce la dimensionalidad del conjunto de datos L al número de dimensiones deseadas.

3.2.4. Clasificación

Integración de Características

El proceso de integración de características unifica la información continua de las imágenes de una secuencia. Primero, para cada imagen, se genera un vector de características compuestas (ver Figura 3.19). Luego, todos los vectores de la secuencia se integraron para crear la matriz de representación de $n \times F$, donde n es el número de imágenes en una secuencia y F el tamaño del vector de características compuestas v (ver Figura 3.14).

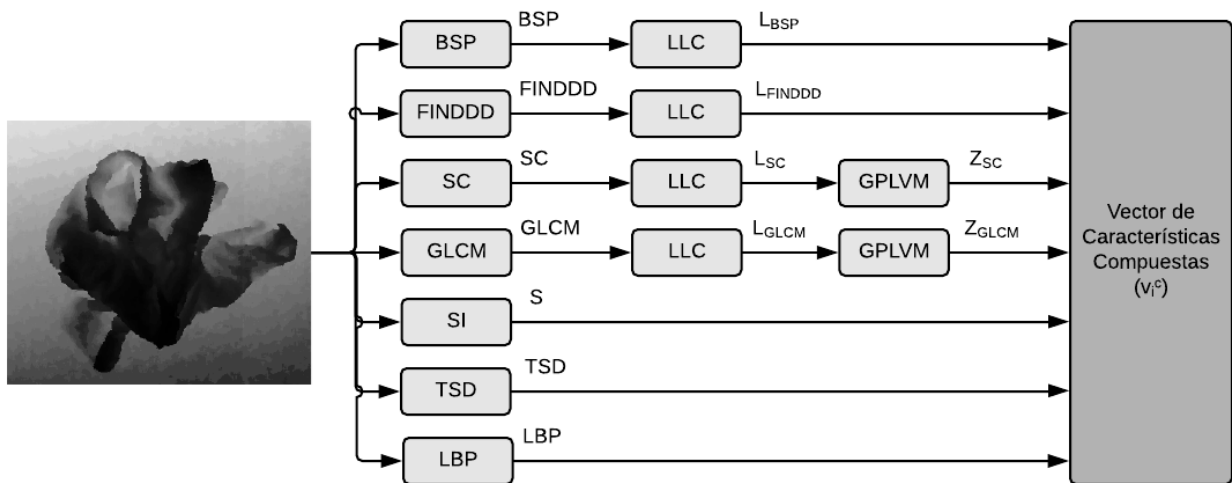


Figura 3.19: Composición del Vector de Características Compuestas de la ropa en una imagen (v_i^c). Este vector engloba las características locales codificadas y las características globales de la ropa de una imagen de la secuencia. Fuente: Elaboración propia.

Clasificación Mediante el Método *Locality Constrained Group Sparse Representation*

Locality Constrained Group Sparse Representation (LGSR) [81] es un método de clasificación comúnmente utilizado para el reconocimiento de acciones humanas, dónde se clasifica cada secuencia de entrada con la información de múltiples cuadros. Este método impone la penalización ponderada de normas mixtas en los coeficientes de reconstrucción para imponer restricciones de grupo dispersas como las restricciones locales suaves de dispersiones y, de este modo, utilizar efectivamente la información intrínseca de varias imágenes dentro de cada secuencia, tratando cada secuencia de entrenamiento/prueba como un grupo de características.

Sea $V = [V^1, V^2, \dots, V^M]$ y $V^i = [v_1^c, v_2^c, \dots, v_n^c]$ donde V^c es la secuencia c del conjunto de entrenamiento y $v_i^c = [L_{BSP}, L_{FINDDDD}, Z_{GLCM}, Z_{SC}, SI, LBP, TSD]$ el vector de características compuestas de la imagen i en la secuencia c ; n y M son el número total de imágenes y el número total de secuencias en el conjunto de entrenamiento, respectivamente. También se definió la secuencia de prueba $Y = [y_1, y_2, \dots, y_n]$, donde $y_i = [L_{BSP}, L_{FINDDDD}, Z_{SC}, Z_{GLCM}, SI, LBP, TSD]$ corresponde al vector de características compuestas de la imagen i de la secuencia de prueba. Siendo $S = [(S^1)^T, (S^2)^T, \dots, (S^M)^T]$ el coeficiente de reconstrucción, S^c es el coeficiente de reconstrucción para la secuencia de prueba con respecto a la secuencia c del conjunto de entrenamiento.

LGSR nos permite imponer restricciones de grupo dispersas y de dispersión local suave al minimizar el error de reconstrucción ponderado $l_{1,2}$ como es mostrados en la ecuación (3.15).

$$\begin{aligned} S^* &= \operatorname{argmin}_S G(S) \\ &= \operatorname{argmin}_S \frac{1}{2} \|Y - VS\|_F^2 + \lambda \sum_{c=1}^M \|D^c \odot S^c\|_F \end{aligned} \quad (3.15)$$

donde $R(S) = \frac{1}{2} \|Y - VS\|_F^2$ representa el error de reconstrucción de la secuencia de prueba Y con respecto al conjunto de entrenamiento V . El segundo término es el regularizador ponderado $l_{1,2}$ basado en normas mixtas del coeficiente de reconstrucción S , y $\lambda > 0$ es el parámetro de regularización para equilibrar estos dos términos. $D^c \in \mathbb{R}^{n_c \times n_p}$ es la matriz de distancia entre las imágenes de la secuencia c del conjunto de entrenamiento y la imágenes de la secuencia de prueba, calculada con la ecuación (3.16). Donde primero se calculó la distancia d_c entre la secuencia de

prueba y la secuencia c del conjunto de entrenamiento utilizando la distancia *single-level Earth mover's* [80], y d_{min} con la distancia mínima de $d_c|_{c=1}^M$.

$$D_{ij} = \exp[(d_c - d_{min})/\sigma]e_{ij} \quad (3.16)$$

Para las características compuestas de la imagen i de la secuencia c en el conjunto de entrenamiento $D_{ij}^c = \exp[(d_c - d_{min})/\sigma]e_{ij}$, donde e_{ij} es la distancia euclidiana entre v_i^c y y_j , y σ el parámetro banda ancha ($\sigma = 1/40, 1/8, 1/4$), que a menor valor, menor es el tiempo de convergencia.

Los valores de S^c se actualizan en la iteración $t + 1$ con la ecuación (3.17). Donde $\partial G(S)/\partial S^c$ es la actualización de la dirección (ecuación (3.18)) y β_t es el tamaño del paso determinado por un método de búsqueda de línea estándar.

$$S_{t+1}^c = S_t^c - \beta_t \left. \frac{\partial G(S)}{\partial S^c} \right|_{S=S_t} \quad (3.17)$$

$$\frac{\partial G(S)}{\partial S^c} = \frac{\partial R(S)}{\partial S^c} + \lambda \frac{\partial \|D^c \odot S^c\|_F}{\partial S^c} \quad (3.18)$$

donde,

$$\frac{\partial R(S)}{\partial S^c} = (V^c)^T (VS - Y) \quad (3.19)$$

$$\frac{\partial \|D^c \odot S^c\|_F}{\partial S^c} = \begin{cases} \frac{D^c \odot D^c \odot S^c}{\|D^c \odot S^c\|} & \text{si } S^c \neq 0 \\ Z^c & \text{si } S^c = 0 \end{cases} \quad (3.20)$$

Después de obtener el coeficiente de reconstrucción óptimo S^* , se utiliza el criterio de Error de Reconstrucción Mínima (minRE) para clasificar la secuencia de prueba. El error de reconstrucción para cada clase se calcula con la ecuación (3.21).

$$R_c((S^c)^*) = \frac{1}{2} \|Y - V^c(S^c)^*\|_F \quad (3.21)$$

donde el coeficiente de reconstrucción $(S^c)^*$ es la S^* correspondiente a la secuencia c del conjunto de entrenamiento. Luego, se clasifica la secuencia de prueba utilizando $c^* = \operatorname{argmin}_c R_c((S^c)^*)$.

El detalle del algoritmo se resume en el Algoritmo 5, en el cual se inicializa S como una matriz con todos sus elementos en cero, agregando las secuencias una por una al conjunto activo para actualizar los coeficientes de reconstrucción correspondientes.

Algoritmo 5: Algoritmo de LGSR

Input : Y : secuencia de prueba, V : conjunto de entrenamiento

- 1 Inicializar $t = 1, S_t = 0 \in \mathbb{R}^{n \times n_p}, A = \{\}$
- 2 Cálculo de D^c entre la secuencia c del conjunto de entrenamiento y la secuencia de prueba, $\forall c \in \{1, \dots, M\}$.
- 3 **while** $t < T_{Max}$ **do**
- 4 Cálculo de $L_c = \|\partial R(S)/\partial S^c\|_F|_{S=S_t} \quad \forall c \in \{c | S_t^c = 0\}$
- 5 **Find** $c^* = \operatorname{argmax}_c L_c$. **If** $L_{c^*} > \lambda \min(D^c)$ **then** $A = c^* \cup A$
- 6 **for each** c **in** A **do**
- 7 Actualizar S_{t+1}^c utilizando la eq. (3.17).
- 8 **if** $S_{t+1}^c = 0$ **then**
- 9 retirar c de A
- 10 **end**
- 11 **end**
- 12 **if** $\|S_{t+1} - S_t\|_F < \epsilon$ ($\epsilon = 0.001$) **then**
- 13 exit WHILE
- 14 **end**
- 15 $t = t+1$
- 16 **end**

Output: S

3.3. Resumen y Análisis del Capítulo

Este capítulo presenta los dos métodos propuestos en este trabajo de tesis. El primer método permite reconocer los puntos de agarre para desplegar una ropa desde su posición inicial aleatoria sin información previamente almacenada. El segundo método consiste en un clasificador de ropa de 5 clases utilizando percepción continua desde múltiples sensores.

El método propuesto de reconocimiento de los puntos de agarre, se basa en la hipótesis de que los puntos de agarre para desplegar una ropa suelen estar cerca de las partes clave de cada clase. El método se divide en tres partes: reconocimiento de partes claves, análisis de arrugas y detección de puntos de agarre. En el reconocimiento de partes claves se reconoce las partes claves *Waist*,

Neck Shirt o *Neck T-Shirt*. En el análisis de arrugas se realiza una detección de las arrugas de la ropa. Finalmente, la detección de puntos de agarre unifica la detección de la parte clave con la información de las arrugas para entregar dos puntos de agarre que puedan desplegar la ropa.

El método de clasificación de ropa mediante percepción continua se basa en la hipótesis de que el uso e integración de información proveniente de distintas deformaciones de la ropa puede mejorar el rendimiento y precisión del sistema de reconocimiento de ropa. El método propuesto propone un enfoque de percepción continua basado en el uso de múltiples imágenes de profundidad de múltiples sensores, mientras que un brazo robótico mueve una prenda sobre y desde una superficie plana. El robot captura la información RGB-D de la ropa de los dos sensores que tiene en cada uno de sus brazos mientras manipula y deja colgando la ropa. Luego se extraen las características de cada imagen capturada de la secuencia para realizar el reconocimiento y dejar la ropa en la caja correspondiente de la clase reconocida.

Capítulo 4

Experimentos y Resultados

Este capítulo muestra los experimentos, y análisis de los resultados del sistema de detección de puntos de agarre y del sistema de clasificación de ropa utilizando percepción continua, propuestos en esta tesis (ver Sección 3). En la evaluación del sistema de puntos de agarre (Sección 4.1) y del subsistema de reconocimiento de partes claves, se evaluaron las clases *Jean*, *T-Shirt* y *Shirt* con la base de datos *Clothing Part Dataset*. En la evaluación del sistema de clasificación de ropa (Sección 4.2) se evaluaron las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*, utilizando las bases de datos *Continuous Clothing Dataset* y *Free-Configuration Clothing Dataset*. Además, se realizó un *ablation study* (Sección 4.2.3) para examinar la efectividad y las contribuciones de los componentes utilizados en el sistema de clasificación.

4.1. Detección de Puntos de Agarre

El sistema de puntos de agarre se basó en la hipótesis de que los puntos de agarre para desplegar una ropa se encuentran cerca o sobre la parte clave de dicha clase, por ejemplo, en la clase *Shirt* los puntos de agarre para desplegar la ropa, se encuentran en los hombros, que se encuentran al lado de la parte clave *Neck*. Teniendo en consideración la importancia de la detección de la parte clave en el sistema, se evaluaron tanto el reconocimiento de partes claves (Sección 4.1.2) como la detección de los puntos de agarre (Sección 4.1.3). Para evaluar ambas partes del sistema se utilizaron las clases: *Jean*, *T-Shirt* y *Shirt*, de la base de datos *Clothing Part Dataset* (Sección 4.1.1). Estas clases fueron seleccionadas por la forma característica de sus partes claves y los buenos resultados que

se encuentran en el estado-del-arte (ver resultado de Ramisa et al., 2014 en la Figura 2.5). En la evaluación del reconocimiento de partes claves se intentó reconocer las partes claves de estas tres clases; cintura (*Waist*) de la clase *Jean*, cuello (*Neck*) de la clase *T-Shirt* y cuello (*Neck*) de la clase *Shirt*. Para la evaluación de la detección de los puntos de agarre, se intentó reconocer aquellos puntos de agarre que permiten extender la ropa y facilitan su doblado. Para esto se marcaron los puntos en una imagen *ground truth*: en la clase *Jean* los puntos se encuentran en los extremos de la parte clave *Waist* y en las clases *T-Shirt* y *Shirt* los puntos se encuentran en los hombros de la ropa, a los lados de la parte clave *Neck*.

4.1.1. *Clothing Part Dataset*

La base de datos *Clothing Part Dataset* [2] comprende 776 escenas de ropa que pertenecen a seis clases: *Polo*, *Jean*, *T-Shirt*, *Dress*, *Shirt* y *Sweater*. Cada escena tiene los siguientes datos: imagen de color, nube de puntos, máscara de segmentación y anotaciones. Para cada prenda, una o dos partes se encuentran anotadas manualmente con polígonos, haciendo un total de 225 anotaciones en toda la base de datos. Las imágenes fueron tomadas utilizando una cámara Kinect que se instaló en la parte superior de una mesa a aproximadamente setenta centímetros. Las imágenes RGB-D adquiridas con la cámara tienen una resolución de 640×480 , y se encuentran en formato PNG y en formato de archivo de texto plano PCD v.7 (parte de profundidad).

En la evaluación del proceso de reconocimiento de partes clave, se utilizaron 300 imágenes de las clases: *Shirt*, *T-Shirt* y *Pant* (ver Figura 4.1). Las cuales fueron divididas en el conjunto de entrenamiento y el conjunto de prueba, en un 60 % y 40 % respectivamente. En cambio, para la evaluación del proceso de detección de los puntos de agarre, se utilizaron 120 imágenes de cada clase. Imágenes en las cuales las partes clave fueron identificadas correctamente. En estas imágenes un punto de agarre se encuentra visible y el segundo punto se encuentra visible o completamente ocluido. Para estos experimentos se evaluaron las partes claves: *Neck* de la clase *Shirt* (NS), *Neck* de la clase *T-Shirt* (NTS) y *Waist* de la clase *Pant* (NP).

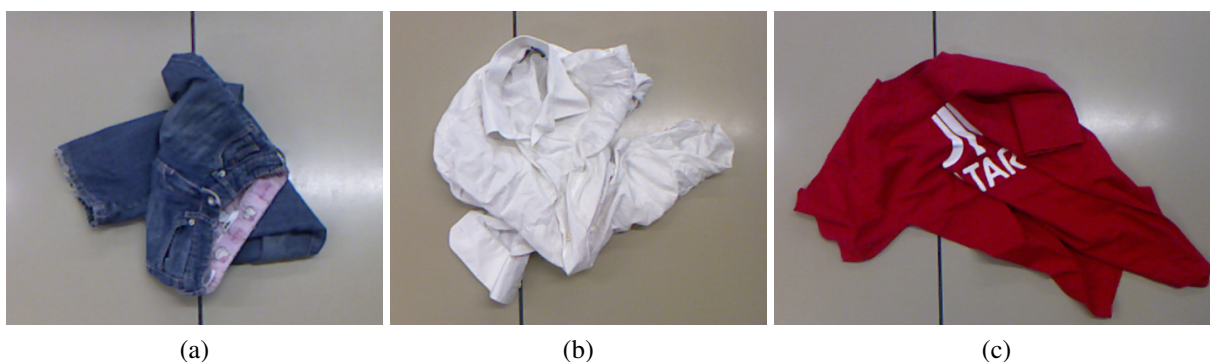


Figura 4.1: Imágenes de las clases evaluadas de la base de datos *Clothing Part Dataset*. a) *Pant*. b) *Shirt*. c) *T-Shirt*.

4.1.2. Resultados del Reconocimiento de Partes Claves

Los resultados del reconocimiento de las partes clave de la ropa se pueden observar en la matriz de confusión en la Tabla 4.1. La parte clave *Neck* de la clase *Shirt*, que físicamente es la que más se diferencia de las otras dos clases por sus dobleces, obtuvo la mayor tasa de reconocimiento con un 74 %. Le sigue la parte clave *Waist* de la clase *Pant* con un 60 % y finalmente la parte clave *Neck* de la clase *T-Shirt* con un 29 %. La baja tasa de reconocimiento de la parte clave *Neck* de *T-Shirt* se debe a que el 40 % se mantiene sin detección. Esto se debe a que se utilizó un filtro Voxel Grid a la nube de puntos, para reducir la cantidad de puntos y por lo tanto disminuir el costo computacional. Al utilizar esta técnica, en algunos casos los parámetros seleccionados eliminaban demasiados puntos para seguir adelante con el algoritmo y el resultado quedaba inconcluso, pero con parámetros que dejaban una mayor cantidad de puntos en otros casos fallaban subpartes del algoritmo. Los resultados de la parte clave *Neck* de la clase *Shirt* y de la parte clave *Waist* de la clase *Pant* son ligeramente superiores a los resultados de Ramisa et al., 2014 [53], los cuales se pueden apreciar en la Figura 2.3.

Tabla 4.1: Matriz de confusión del reconocimiento de partes claves. Clases: *Neck Shirt* (NS), *Neck T-Shirt* (NTS), *Waist Pant* (W), *No detection* (ND).

	NS	NTS	W	ND
NS	74 %	2 %	0 %	24 %
NTS	5 %	29 %	26 %	40 %
W	5 %	25 %	60 %	10 %

4.1.3. Resultados de la Detección de Puntos de Agarre

En la detección de puntos de agarre para medir la precisión de las detecciones se utilizó como métrica la intersección sobre la unión (IoU), que se puede observar en la ecuación (4.1). En la implementación de esta métrica, se generó una región rectangular de largo 51 sobre los puntos detectados y los puntos marcados en el *ground truth*.

$$IoU = Intersection/Union \quad (4.1)$$

La Tabla 4.2 muestra el IoU promedio de las 120 imágenes. Para el mejor punto localizado, se tiene una tasa de detección similar en todas las clases. Sin embargo, al considerar la detección de ambos puntos, la tasa de detección disminuye considerablemente. En el caso de la clase *T-Shirt* la tasa de detección de los 2 puntos disminuye un 47 % con respecto al mejor punto localizado. Esto puede significar, que la hipótesis de que se pueden encontrar los puntos de agarre basándose en la detección de la parte clave no es correcta para la clase *T-Shirt*.

Tabla 4.2: Resultados promedios de la detección de puntos de agarre utilizando la métrica IoU

	1 punto	2 puntos
<i>Pant</i>	0.65	0.34
<i>Shirt</i>	0.57	0.28
<i>T-Shirt</i>	0.54	0.07

En los resultados de 1 punto, el cálculo está considerando sólo el punto detectado más preciso de ambos puntos, ya que el promedio de los resultados supera los 0.5 de IoU, se confirmó que al menos uno de los puntos de agarre podría servir para desplegar la ropa. En los resultados particulares de la detección de partes claves, se pueden observar las mismas proporciones de detección, es decir, los mejores resultados de detección de los puntos de agarre se encuentran en la clase *Pant*, luego en *Shirt* y finalmente en *T-Shirt*.

Para detectar además la precisión de los puntos detectados se utilizó una segunda métrica, la cual es considerar cómo correctas sólo aquellas detecciones con un IoU mayor o igual al 50 %. Con esto se obtuvo el *recall* para uno y para dos puntos de agarre. Los resultados se pueden observar en la Tabla 4.3.

Tabla 4.3: *Recall* de las detecciones de puntos de agarre

	1 punto	2 puntos
<i>Pant</i>	63 %	43 %
<i>Shirt</i>	55 %	33 %
<i>T-Shirt</i>	48 %	29 %

En la Tabla 4.3 se puede observar en todas las clases que uno de los puntos de agarre tiene una detección cercana o superior al 50 %. Esto puede deberse a que la parte clave se encuentra visible en las imágenes utilizadas, y por lo tanto al menos uno de los puntos de agarre está visible la mayoría de los casos. El segundo punto de agarre estuvo oculto en varias imágenes y en estos casos es difícil para el sistema encontrar un punto similar al marcado en el *ground truth*.

En las Tablas 4.2 y 4.3, los resultados de detección de un punto de agarre utilizando ambas métricas son bastante cercanas. Por lo que se puede concluir que al utilizar este método de detección de puntos de agarre, el 55 % de las veces al menos uno de los puntos sería correctamente detectado y el segundo serviría para desplegar la ropa y que el robot pueda observar más superficie de la ropa.

Debido a la inexistencia de investigaciones en este problema de detección de puntos de despliegue de ropa no puede realizar una comparación con el estado-del-arte. Sin embargo, en la Figura 4.2 es posible observar ejemplos de los resultados de la detección de los puntos de agarre y las imágenes de todas las clases en las diferentes fases del algoritmo.

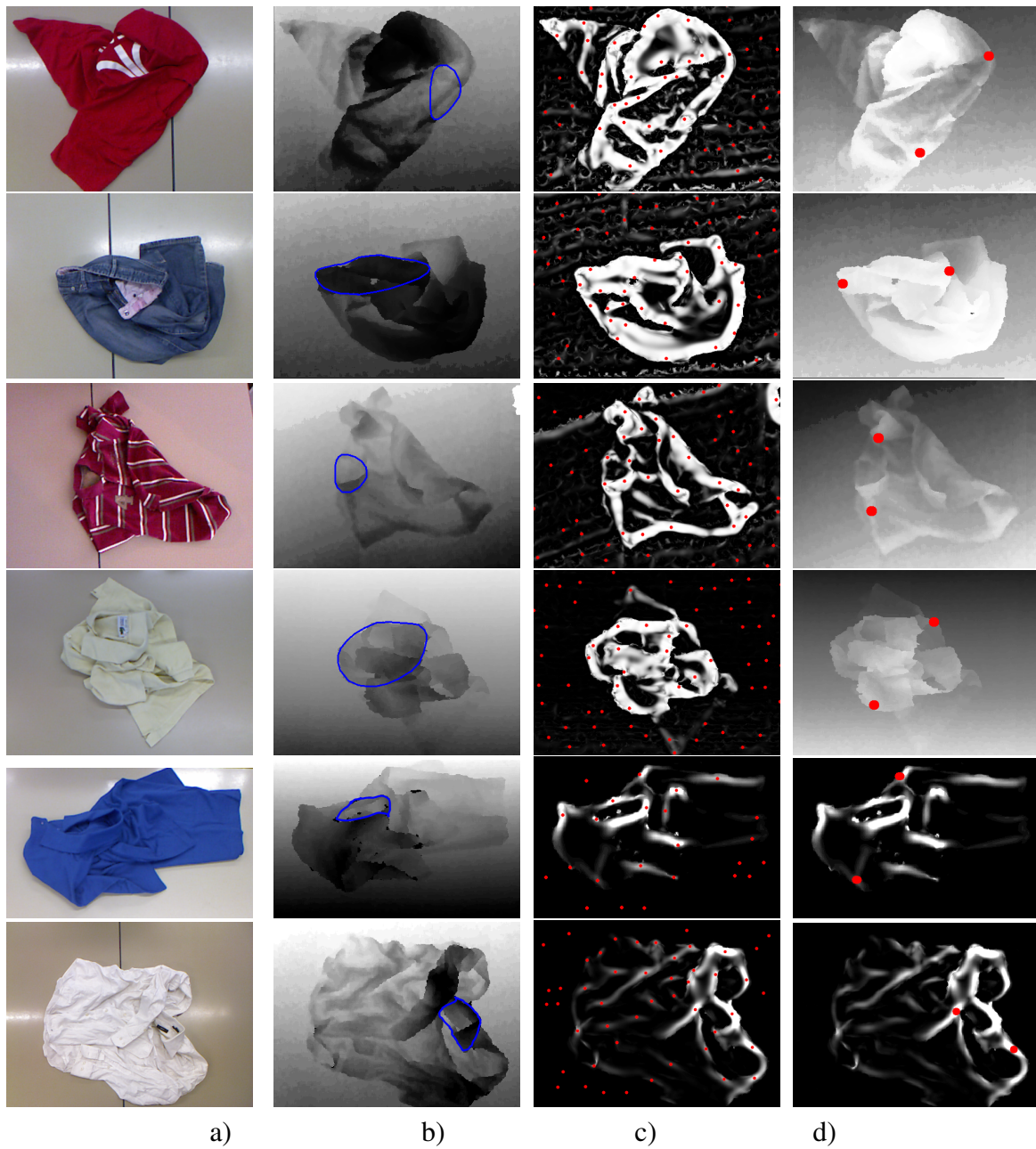


Figura 4.2: Resultados del algoritmo de detección de puntos de agarre. Las primeras tres filas corresponden a casos con detecciones correctas y los siguientes tres casos corresponden a casos fallidos. a) Imagen RGB, b) Imagen de profundidad con el contorno de la parte clave detectada marcada en azul, c) Análisis de arrugas con el filtro vessel y los puntos máximos locales marcados en rojo, y d) Puntos de agarre detectados marcados en rojo.

4.2. Clasificación con Percepción Continua

Para demostrar la hipótesis de que al utilizar percepción continua se obtiene mayor información visual al trabajar con objetos deformable, se realizó un diseño experimental que consiste en clasificar ropa con dos bases de datos diferentes. Con estas bases de datos, se pudo evaluar y comparar el rendimiento del sistema con otros enfoques de la literatura, utilizando el enfoque de percepción continua, en el cual se utiliza la información obtenida de una secuencia al manipular una prenda de ropa. Para examinar la efectividad y las contribuciones de las diferentes características visuales (Secciones 3.2.1 y 3.2.2) y técnicas de codificación (Sección 3.2.3) utilizadas en el sistema, se realizó un ablation study (Sección 4.2.3).

4.2.1. Bases de Datos

Para la evaluación del sistema de clasificación de ropa, se recopiló la primera base de datos de secuencias de vídeo RGB-D, llamada *Continuous Clothing Dataset*. Esta base de datos contiene las secuencias de vídeo que muestran cómo se manipulan distintos tipos de ropa. Las secuencias fueron grabadas utilizando dos sensores Asus Xtion Pro Live ubicados en los grippers de un robot industrial con dos brazos, llamado CloPeMa [1]. Además, para evaluar el sistema con otros métodos del estado-del-arte, se utilizó la base de datos *Free-Configuration Clothing Dataset* para hacer clasificaciones *single-shot* (clasificaciones utilizando una sola imagen). La base de datos posee imágenes con diferentes resoluciones, lo que permitió comparar el rendimiento del sistema utilizando un sensor estéreo de alta resolución y una cámara Asus Xtion Pro Live.

Continuous Clothing Dataset

Al utilizar percepción continua en este sistema, fue necesaria la recopilación de una base de datos de secuencias de vídeo RGB-D. La cual para facilitar la integración con el *framework* ROS, se creó la base de datos como una colección de ‘rosbags’ [3], que contienen las imágenes de color y de profundidad, nubes de puntos, información de los sensores y todas las transformaciones cinemáticas del robot, de ambos sensores a 15 Hz de cada secuencia de vídeo. La base de datos guarda las secuencias de 5 modelos de ropa de cada una de las siguientes 5 clases: *T-Shirt*,

Shirt, Sweater, Pant y Hand-Towel. En la Figura 4.3 se puede observar una imagen de ejemplo de cada clase de la base de datos, tomada desde el sensor que se encuentra en el brazo que manipula la prenda. Cada modelo de ropa se capturó partiendo de 10 configuraciones aleatorias distintas, resultando un total de 150 secuencias. En las secuencias la prenda de ropa comienza en una configuración aleatoria y se muestra como es manipulado por el robot. La importante contribución de la creación de esta base de datos permite realizar comparaciones de los distintos enfoques vistos en la Sección 2; cuando la ropa se encuentra sobre una superficie plana, colgando o realizando un movimiento continuo. Esta base de datos se encuentra disponible gratuitamente en <http://dx.doi.org/10.5525/gla.researchdata.669>. Más información sobre esta base de datos se puede encontrar en el Anexo A.

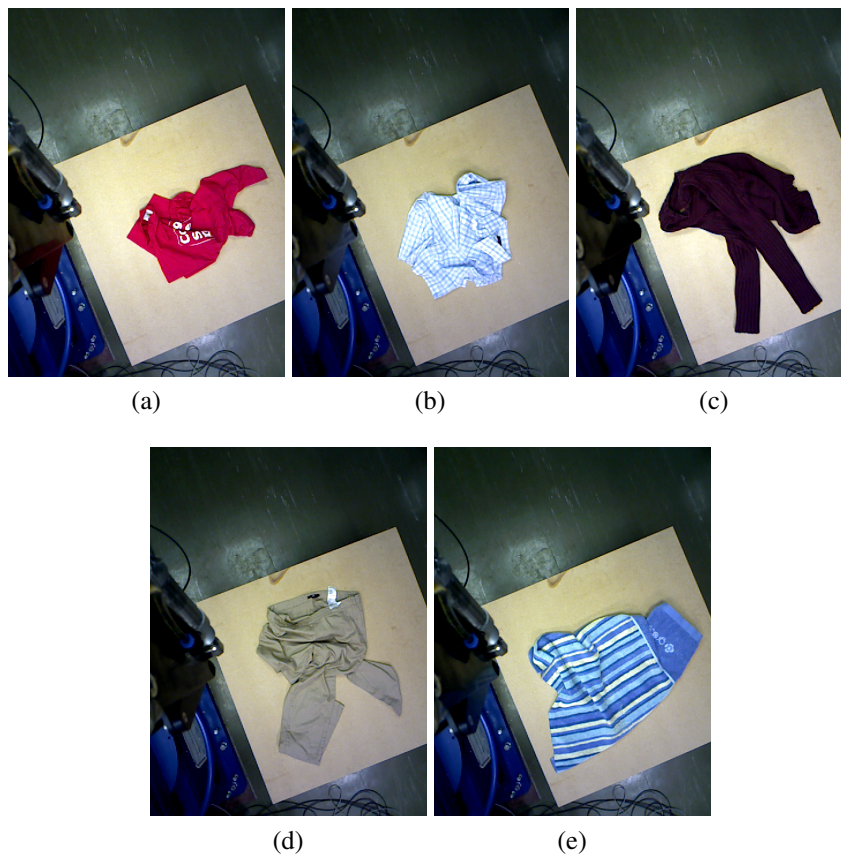


Figura 4.3: Imágenes de las clases de la base de datos *Continuous Clothing Dataset*, tomada desde el sensor que se encuentra en el brazo que manipula la prenda. a) clase *T-Shirt*, b) clase *Shirt*, c) clase *Sweater*, d) clase *Pant* y e) clase *Hand-Towel*. Fuente: Elaboración propia.

Free-Configuration Clothing Dataset

Para los experimentos *single-shot*, se usó la base de datos *Free-Configuration Clothing Dataset* [65]. Esta base de datos comprende 10 modelos de ropa de cada una de las siguientes 5 clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*. Cada modelo de ropa se capturó en 21 configuraciones aleatorias diferentes con un sistema estéreo de alta resolución y un sensor Asus Xtion Pro Live. Ambos sensores proporcionan por cada configuración: una imagen RGB, una imagen de profundidad y una máscara segmentada. Las imágenes capturadas con el sistema estéreo tienen una resolución de 16 megapíxeles (4928×3264) y las imágenes capturadas con el sensor Asus Xtion Pro Live tienen una resolución VGA (640×480). En total, la base de datos está compuesta por 1,050 imágenes de prendas en configuraciones aleatorias por cada uno de estos sensores.



Figura 4.4: Imágenes de las clases en su estado desdoblado de la base de datos *Free-Configuration Clothing Dataset*, tomada desde el sensor Asus Xtion Pro Live. a) clase *T-Shirt*, b) clase *Shirt*, c) clase *Sweater*, d) clase *Pant* y e) clase *Hand-Towel*. Fuente: Sun et al. [65].

4.2.2. Experimentos de Percepción Continua

El método propuesto se comparó con dos métodos del estado-del-arte: percepción interactiva [64] y percepción *single-shot* [68], utilizando la base de datos *Continuous Clothing Dataset*. Dado que estos dos métodos evalúan utilizando una imagen, se seleccionaron tres imágenes

representativas de cada secuencia: la primera imagen (cuando el objeto está sobre la mesa), la última imagen (cuando el objeto está colgando de un *gripper*) y la imagen con el mejor resultado. Para el tercer caso, los métodos evaluaron cada imagen de la secuencia y se seleccionó el resultado con el mejor rendimiento. La Tabla 4.4 muestra la comparación entre el método propuesto y los dos métodos del estado-del-arte utilizando las tres imágenes representativas.

Tabla 4.4: Comparación de rendimiento entre el método propuesto con dos métodos del estado-del-arte. Primero, el método de percepción interactivo [64], con las características LBP, SI y TSD (L-S-T) utilizando el proceso gaussiano (GP). Segundo, la percepción *single-shot* [68], con las características LBP, SI, TSD y BSP (L-S-T-B) usando SVM.

Algoritmo	Precisión
L-S-T con GP (primera imagen)	35.6 %
L-S-T con GP (ultima imagen)	37.47 %
L-S-T con GP (mejor resultado)	38.00 %
L-S-T-B con SVM (primera imagen)	38.93 %
L-S-T-B con SVM (ultima imagen)	37.67 %
L-S-T-B con SVM (mejor resultado)	39.40 %
Método propuesto	66.7 %

Los resultados de la clasificación están representados en la Tabla 4.4 y la Figura 4.5. En el método propuesto se observa una precisión de clasificación promedio del 66.7 %, con una precisión específica por clase de 58.00 %, 41.58 %, 83.84 %, 67.00 %, 83.84 % para las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*, respectivamente. De los resultados, se puede apreciar que las clases de *Sweater* y *Hand-Towel* representan los mejores rendimientos de clasificación debido a las diferencias que existen entre las clases, en forma y superficie entre las diferencias más notorias a simple vista. Si bien la clase *Sweater* obtiene los mejores rendimientos en la clasificación, también posee los valores más altos de falsos positivos, lo que afecta mayormente a la clase *Shirt*. Esto se puede deber a la similitud que existe entre estas clases, ya que ambos se utilizan para la parte superior del cuerpo, poseen mangas largas y tienen un volumen bastante similar. Además, al utilizarse distintos modelos de ropa para la fase de entrenamiento y la fase de prueba, de los resultados se puede concluir que los modelos utilizados en la fase prueba de la clase *Shirt* al estar sobre la mesa y durante el movimiento tienen más similitud a la clase *Sweater*.

Además, otra reducción en el rendimiento se debe al hecho de que las arrugas durante la secuencia de las clases *Sweater* y *Shirt* son bastante similares, ya que ambas clases están hechas del

mismo material, es decir, ambos tipos de ropa están hechos a base de algodón. En el mejor resultado del método de percepción interactiva [64], se observa una precisión de clasificación promedio del 38.00 %, con una precisión específica por clase de 41.67 %, 17.33 %, 31.67 %, 32.00 %, 51.33 % para las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel* respectivamente. En el mejor resultado del método de percepción *single-shot* [68], se observa una precisión de clasificación promedio del 39.4 %, con una precisión específica por clase de 48.00 %, 16.33 %, 25.00 %, 55.00 %, 52.67 % para las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel* respectivamente.



Figura 4.5: Matriz de confusión del método propuesto utilizando la base de datos *Continuous Clothing Dataset*.

Al considerar la precisión promedio y la precisión individual de cada clase, se puede confirmar que el método propuesto de percepción continua supera el método de percepción interactiva por un 31.9 %, y al método de percepción *single-shot* por un 29.23 %. Por lo tanto, se concluye que el método propuesto mejora la tarea robótica de clasificación de ropa con respecto a los métodos del estado-del-arte, ya que el espacio de objetos ya no se describe basándose en la estructura 3D de su superficie visible, sino observando la prenda a lo largo de un tiempo determinado.

4.2.3. Ablation Study

Ablation study se define como el procedimiento en el que se eliminan algunos componentes de un método en estudio y se analiza el efecto al eliminar estos componentes. En este estudio se desea ver la contribución particular de cada uno de los componentes del método propuesto, al evaluar el desempeño con y sin estos componentes. En este *ablation study* se investiga cómo las diferentes

características visuales y los algoritmos de codificación contribuyen y se desempeñan en el método propuesto de percepción continua.

Los experimentos de este estudio están indicados en las Tablas 4.5 y 4.6. Los experimentos en la Tabla 4.5 tratan de evaluar la contribución de cada una de las características locales y globales para la tarea de clasificación de la ropa con percepción continua. Los experimentos fueron divididos de la siguiente manera: método propuesto (ID 1.1), características locales (ID 1.2-1.5), características globales (ID 1.6-1.8), sólo características globales (ID 1.9) y sólo características locales (ID 1.10). En los experimentos de 1.2 a 1.8, se desactivo la contribución de una característica visual, mientras se dejaba el resto sin cambios. De manera similar, la Tabla 4.6 muestra los experimentos que evalúan la contribución de los algoritmos de codificación LLC y GPLVM para la tarea de clasificación. Estos algoritmos de codificación se aplican sólo en las características locales. Los experimentos se distribuyen de la siguiente manera: el método propuesto (ID 2.1), desactivando LLC para las características BSP, FINDDD, SC y GLCM (ID 2.2-2.5) y desactivando GPLVM para las características SC y GLCM (ID 2.6-2.7).

Tabla 4.5: *Ablation study* de las características del método propuesto. Experimentos que desactivan la contribución de una o un grupo características visuales, mientras se deja el resto sin cambios. Los valores de la tabla indican si característica está activada (valor ✓) o cuando está desactivada (valor ✗).

ID	BSP	FINDDD	SC	GLCM	SI	LBP	TSD
1.1	✓	✓	✓	✓	✓	✓	✓
1.2	✗	✓	✓	✓	✓	✓	✓
1.3	✓	✗	✓	✓	✓	✓	✓
1.4	✓	✓	✗	✓	✓	✓	✓
1.5	✓	✓	✓	✗	✓	✓	✓
1.6	✓	✓	✓	✓	✗	✓	✓
1.7	✓	✓	✓	✓	✓	✗	✓
1.8	✓	✓	✓	✓	✓	✓	✗
1.9	✗	✗	✗	✗	✓	✓	✓
1.10	✓	✓	✓	✓	✗	✗	✗

Tabla 4.6: *Ablation study* de los algoritmos de codificación del método propuesto. Experimentos que desactivan la contribución de un algoritmo de codificación, mientras se deja el resto sin cambios. Los valores de la tabla indican si el algoritmo de codificación está activada (valor ✓) o cuando está desactivada (valor ✗).

ID	LLC				GPLVM	
	BSP	FINDDD	SC	GLCM	SC	GLCM
2.1	✓	✓	✓	✓	✓	✓
2.2	✗	✓	✓	✓	✓	✓
2.3	✓	✗	✓	✓	✓	✓
2.4	✓	✓	✗	✓	✓	✓
2.5	✓	✓	✓	✗	✓	✓
2.6	✓	✓	✓	✓	✗	✓
2.7	✓	✓	✓	✓	✓	✗
2.8	✓	✓	✓	✓	✗	✗

Las Figuras 4.6 y 4.7 muestran los resultados de los experimentos descritos en las Tablas 4.5 y 4.6, respectivamente. En la Figura 4.6 se puede observar como las características locales (ID 1.2 - 1.5) tienen una alta contribución al método propuesto, ya que los puntajes de clasificación están cerca o por debajo del 50 % en el puntaje de clasificación, menor que cuando no se considera una característica global.

En las características globales se observan contribuciones mínimas (ver resultados de clasificación de los experimentos 1.9 y 1.10). Aunque la combinación de características locales y globales representa la mejor puntuación de clasificación, las características globales sólo contribuyen aproximadamente el 3 % de la puntuación total.

Además, al observar la Figura 4.7 se puede concluir que el algoritmo de codificación GPLVM (ID 2.6 y 2.7) no contribuye considerablemente al método propuesto con respecto a LLC, por lo que se considera al codificación GPLVM una técnica opcional en caso de querer disminuir el costo computacional.

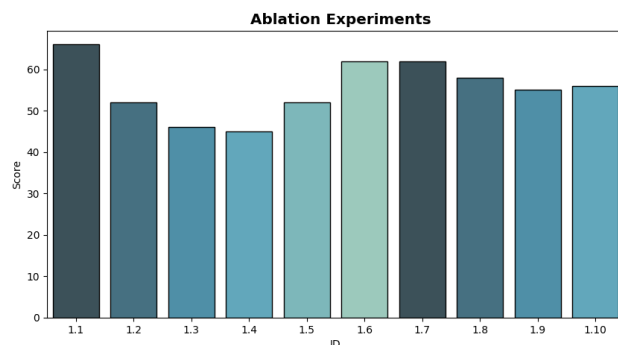


Figura 4.6: Resultados del *ablation study* sobre las características de los experimentos mostrados en la Tabla 4.5.

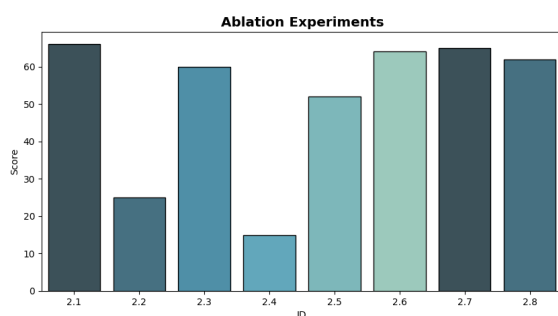


Figura 4.7: Resultados del *ablation study* sobre los algoritmos de coding de los experimentos mostrados en la Tabla 4.6.

4.2.4. Estrategia en Percepción Continua

Siendo el sensor 1 la cámara RGB-D que se encuentra en el brazo que manipula la prenda y captura las primeras imágenes de la prenda sobre la mesa, es decir, una vista egocéntrica y el sensor 2 la cámara RGB-D que captura imágenes a distancia, mientras que el otro brazo recoge la prenda de la mesa, es decir, la vista exocéntrica, son muchas las imágenes que se pueden considerar en una secuencia. Por lo que para evaluar cuántas imágenes de cada sensor debían considerarse en una secuencia, se evaluó con 1 a 4 imágenes del sensor 1 y con 5, 15 y 20 imágenes del sensor 2 (ver Figura 3.14). Estos resultados se pueden observar en la Tabla 4.7.

Tabla 4.7: Resultados de los experimentos al utilizar distintas estrategias en el algoritmo, cambiando la cantidad de imágenes del Sensor RGBD 1 (S1) y el Sensor RGBD 2 (S2) en la secuencia.

S2 \ S1	1	2	3	4
5	55.0 %	55.0 %	55.0 %	50.0 %
15	66.7 %	66.7 %	63.3 %	53.3 %
25	58.3 %	61.7 %	61.7 %	51.7 %

La Tabla 4.7 muestra que es mejor utilizar 1 o 2 imágenes del sensor 1. Además, se descubrió que el método propuesto logra un mejor rendimiento al usar las últimas quince imágenes del sensor 2; esto se debe a que se observa más superficie de la prenda cuando está colgando.

4.2.5. Experimentos *single-shot*

Con el fin de comparar el rendimiento de las características y las técnicas de codificación utilizadas en el método propuesto con el estado-del-arte, se validó el método propuesto utilizando la base de datos *Free-Configuration Clothing Dataset*. Esta base de datos permite evaluar la clasificación *single-shot*, pero no contiene secuencias de video para explorar completamente el enfoque de percepción continua. Para adaptar el método propuesto que utiliza percepción continua a *single-shot*, se reemplazó el algoritmo de clasificación LGSR (Sección 3.2.4) con un clasificador SVM, pero con las mismas características locales (BSP, FINDDD, SC y GLCM) y características globales (SI, LBP y TSD). Con el fin de aumentar la precisión del sistema, se evaluó la integración de características distintivas, específicamente la información sobre las partes claves *Neck* y *Waist* de las clases *Shirt* y *Pant*, respectivamente, como en el sistema de la Sección 3.1.

Para estos experimentos, se utilizaron sólo las imágenes de alta resolución de la base de datos (4928×3264), obteniendo un puntaje promedio de clasificación de 84.8 % (ver Tabla 4.8) usando el método propuesto sin LGSR. Como se muestra en la matriz de confusión de la Figura 4.8, (a) el puntaje de clasificación por clase es 91 %, 67.1 %, 83.8 %, 90.9 % y 91.1 % para las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*, respectivamente. Al utilizar la misma base de datos, estos valores pueden compararse con el resultado del estado-del-arte de Sun et. al. [68], que entregan un puntaje promedio de clasificación de 83.2 % (ver Tabla 4.8) y con puntuaciones individuales de: 89.2 %, 70.0 %, 80.8 %, 87.0 % y 88.8 % (Figura 6 (f) en [68]) para las clases: *T-Shirt*, *Shirt*,

Sweater, Pant y Hand-Towel, respectivamente. Se debe tener en cuenta que la mejora es sólo marginal, pero permite confirmar que el enfoque del método propuesto es comparable con los métodos actuales para la clasificación de ropa al realizar el reconocimiento *single-shot*.

Para los experimentos donde se incluyeron las características de *Neck* y *Waist*, se obtuvo un puntaje de clasificación promedio de 87.7 % (ver Tabla 4.8 y Figura 4.8 (b)), con los puntajes de clasificación 88.3 %, 79 %, 87.4 %, 93.1 % y 90.5 % para las clases: *T-Shirt, Shirt, Sweater, Pant* and *Hand-Towel*, respectivamente. Al integrar información de características visuales más distintivas dentro del sistema, se puede observar un aumento en el rendimiento en los puntajes de clasificación de aproximadamente un 3 %. Por lo tanto, se puede especular que las características visuales tales como los botones y mangas en las prendas, darían lugar a menos similitudes entre las clases y, por consiguiente, aumentarán los puntajes de clasificación específicos de cada clase.

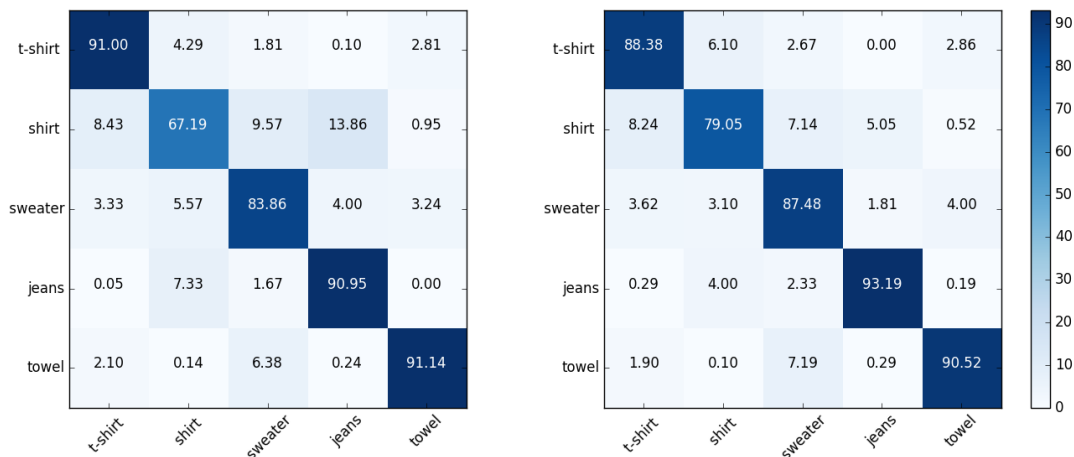


Figura 4.8: Matriz de confusión del método propuesto utilizando las imágenes de alta resolución de la base de datos *Free-Configuration Clothing Dataset*. **a)** muestra los resultado del método propuesto y **b)** los resultados del método propuesto integrando las características distintivas.

La Tabla 4.8 también muestra una comparación entre los resultados con distintas resoluciones de imagen. Esto nos permitió demostrar que el sistema mejora la precisión de reconocimiento al mejorar la resolución de la imagen de profundidad. Además, se puede observar que en las imágenes de baja resolución el método propuesto aumenta el 1.1 % el resultado del estado-del-arte de Sun et al., [68] y el 9.4 % agregando características distintivas. Al considerar las imágenes de alta resolución se observa que el método propuesto aumenta el 1.6 % el resultado del estado-del-arte y 4.5 % al agregar características distintivas. El aumento en el rendimiento al usar las características distintivas se debe a que la parte clave *Neck* de la clase *Shirt* y la parte clave *Waist* de la clase *Pant*

mejoran la precisión en estas clases y disminuyen los falsos positivos con respecto a otras clases.

Se debe tener en cuenta que estas características distintivas dieron como resultado un aumento en el rendimiento de las tareas de reconocimiento / clasificación en *single-shot*. La razón de esto es que mientras percibimos continuamente las prendas de ropa a medida que el robot las recoge, las características distintivas de la ropa desaparecen y aparecen aleatoriamente entre los marcos; haciendo que LGSR pierda precisión. Por lo tanto, no incluimos estas características distintivas en los experimentos de percepción continua mostrados en las Secciones 4.2.2 y 4.2.3.

Tabla 4.8: Comparación de rendimiento entre el método propuesto con y sin características distintivas contra dos métodos del estado-del-arte. Primero, el método de percepción interactivo [64], con las características LBP, SI y TSD (L-S-T) utilizando GP. Segundo, la percepción *single-shot* [68], con las características LBP, SI, TSD y BSP (L-S-T-B) utilizando el clasificador SVM.

Algoritmo	Baja resolución	Alta resolución
L-S-T with GP [64]	58.5 %	70.8 %
L-S-T-B with SVM [68]	64.2 %	83.2 %
Método propuesto	65.3 %	84.8 %
Método propuesto + información distintiva	73.6 %	87.7 %

4.3. Resumen y Análisis del Capítulo

Este capítulo muestra los experimentos, y análisis de los resultados del método propuesto de detección de puntos de agarre y del método propuesto de clasificación de ropa utilizando percepción continua. En la evaluación del sistema de puntos de agarre y del subsistema de reconocimiento de partes claves, se evaluaron las clases *Jean*, *T-Shirt* y *Shirt*. En la evaluación del sistema de clasificación de ropa se evaluaron las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*, utilizando el enfoque de percepción continua, *single-shot* y *ablation study*.

En el método propuesto de detección de puntos de agarre se evaluaron tanto el reconocimiento de partes claves como la detección de los puntos de agarre. En la evaluación del reconocimiento de partes claves se intentó reconocer las partes claves de estas tres clases; cintura (*Waist*) de la clase *Jean*, cuello (*Neck*) de la clase *T-Shirt* y cuello (*Neck*) de la clase *Shirt*. Para la evaluación de

la detección de los puntos de agarre, se intentó reconocer aquellos puntos de agarre que permiten extender la ropa y facilitan su doblado. Para esto se marcaron los puntos en una imagen *ground truth*: en la clase *Jean* los puntos se encuentran en los extremos de la parte clave *Waist* y en las clases *T-Shirt* y *Shirt* los puntos se encuentran en los hombros de la ropa, a los lados de la parte clave *Neck*.

En la evaluación del reconocimiento de partes claves, la parte clave *Neck* de la clase *Shirt* obtuvo una tasa de reconocimiento del 74 %, la parte clave *Waist* de la clase *Pant* un 60 % y la parte clave *Neck* de la clase *T-Shirt* un 29 %. Se observa que los resultados de la parte clave *Neck* de la clase *Shirt* y de la parte clave *Waist* de la clase *Pant* son ligeramente superiores a los resultados de Ramisa et al., 2014 [53], los cuales se pueden apreciar en la Figura 2.3. Sin embargo, es difícil realizar una comparación directa dado que en Ramisa et al., 2014 [53] se realiza el reconocimiento de 11 clases distintas.

En la evaluación de la detección de los puntos de agarre se analizó el *recall* (aquellos puntos con un IoU mayor o igual al 50 %) para uno y dos puntos de agarre. Donde en promedio se obtuvo un *recall* del 55 % del mejor punto detectado y del 35 % de ambos puntos. De estos resultados se puede concluir que al ser la ropa un objeto uniforme, en muchas situaciones la hipótesis no se cumple, ya que puede ocurrir que una manga u otra sección de la ropa se encuentre tapando los puntos de agarre que podrían desplegar la ropa, aun cuando la parte clave se encuentre visible. En estos casos se podría añadir como trabajo futuro, que al detectar estas situaciones el punto para desplegar se encuentre en la parte clave. Sin embargo, se considera satisfactorio el experimento dado que estos puntos si bien no serían los finales para desplegar la ropa completamente, pueden servir para levantar la ropa y permitir al robot desde esta posición realizar un reconocimiento.

En el método propuesto de clasificación de ropa utilizando percepción continua se realizaron experimentos utilizando el enfoque de percepción continua y single-shot.

El método propuesto utilizando percepción continua se comparó con dos métodos del estado-del-arte, pero dado que estos dos métodos utilizan una imagen para realizar el reconocimiento, la comparación se realizó seleccionando tres imágenes representativas de cada secuencia: la primera imagen (cuando el objeto está sobre la mesa), la última imagen (cuando el objeto está colgando de un *gripper*) y la imagen con el mejor resultado. Para el tercer caso,

los métodos evaluaron cada imagen de la secuencia y se seleccionó el resultado con el mejor rendimiento.

En los resultado se observa que el método propuesto tiene una precisión de clasificación promedio del 66.7%, y los métodos del estado-del-arte una precisión de clasificación promedio del 38.00% y del 39.4% utilizando la base de datos *Continuous Clothing Dataset*. Por lo que al considerar que los métodos utilizados del estado-del-arte obtienen mejores resultados utilizando la base de datos *Free-Configuration Clothing Dataset*, la cual posee imágenes sobre una superficie donde el sensor se encuentra más cerca de la ropa y los sensores son de mejor resolución, se puede concluir que que los métodos del estado-del-arte fueron afectados por la baja resolución de las imágenes de la base de datos y que el método propuesto que utiliza percepción continua es más robusto en imágenes de baja resolución.

Con el fin de comparar el rendimiento de las características y las técnicas de codificación utilizadas en el método propuesto con el estado-del-arte, se utilizó una base de datos que permite evaluar la clasificación *single-shot*. Para adaptar el método propuesto de percepción continua a *single-shot*, se reemplazó el algoritmo de clasificación LGSR con un clasificador SVM. Además, con el fin de aumentar la precisión del sistema, se evaluó la integración de características distintivas, es decir, se integró la información de reconocimiento de las partes claves *Neck* y *Waist* que se desarrolló en el primer método propuesto. En los experimentos se obtuvo un puntaje promedio de clasificación de 84.8% y 87.7%, sin y con características distintivas respectivamente. Estos valores pueden compararse con el 83.2% de Sun et. al. [68], que si bien es una mejora marginal, permite confirmar que el enfoque del método propuesto es comparable con los métodos actuales para la clasificación de ropa al realizar el reconocimiento *single-shot*. Además, al observar una mejora al utilizar características distintivas se puede especular que las características visuales tales como los botones y mangas en las prendas, darían lugar a menos similitudes entre las clases y, por consiguiente, aumentarían los puntajes de clasificación específicos de cada clase.

Capítulo 5

Conclusiones y Trabajo Futuro

5.1. Conclusión

En la actualidad, el reconocimiento de ropa a partir de una pose aleatoria, es considerada una tarea difícil incluso para un ser humano, ya que existen una gran cantidad de clases, modelos, tamaños y diseños. Además, la incalculable cantidad de poses que puede llegar a tener cada prenda de ropa al tratarse de un objeto deformable, lo convierte en una de las tareas más difíciles en el área de la visión computacional. Para resolver este problema, este trabajo de tesis propone una metodología que permite que el robot sea capaz de reconocer una ropa, cuando esta se encuentre en una pose inicial desconocida y aleatoria, tomando la ropa y obteniendo distintas vistas utilizando percepción continua y clasificar la prenda de ropa en una de la 5 clases previamente definidas: *T-Shirt* (polera), *Shirt* (camisa), *Sweater* (sweater), *Pant* (pantalón) y *Hand-Towel* (toalla de mano). El objetivo de esta metodología propuesta es que el robot sea capaz de depositar la prenda de ropa en un lugar previamente definido, dependiendo de la clase reconocida.

Este trabajo de tesis logró tres aportes considerables al estado-del-arte. El primero corresponde al diseño e implementación de un sistema de detección de puntos agarre sin utilizar información previa. El segundo aporte, corresponde al diseño e implementación de un sistema de reconocimiento y clasificación de ropa que mejoró los resultados del estado-del-arte en experimentos *single-shot* y que propone un método de evaluación utilizando percepción continua. El último aporte, corresponde a la creación de la primera base de datos de secuencias de ropa que permite la evaluación del sistema de reconocimiento propuesto. Esta base de datos permite hacer

evaluaciones de métodos que evalúan cuando la ropa se encuentra sobre la mesa, colgando por uno de los brazos del robot o utilizando percepción continua, que permite unificar las investigaciones de este problema. La base de datos proporciona las imágenes RGB-D de dos sensores con distintas perspectivas y la cinemática del robot en el *framework* ROS.

Considerando la hipótesis de que los puntos de agarre para desplegar una ropa se encuentran cerca o en la parte clave de dicha clase, se propuso un sistema capaz de calcular los puntos de agarre para desplegar una prenda de ropa, que no requiere estos puntos almacenados previamente. Este sistema combina los contornos de la región clave detectada y la información de rugosidad de la ropa obtenida. Para la evaluación de la detección de los puntos de agarre, se intentó reconocer aquellos puntos de agarre que extienden la ropa y facilitan su doblado, para esto se marcaron los puntos en una imagen *ground truth*: en la clase *Pant* los puntos se encuentran en los extremos de la parte clave *Waist* y en las clases *Shirt* y *T-Shirt* se encuentran en los hombros de la ropa, al lado de la parte clave *Neck*.

En la evaluación del sistema de detección puntos de agarre, el mejor punto localizado obtuvo un 65 %, 57 % y 54 % para las clases *Pant*, *Shirt* y *T-Shirt* respectivamente, utilizando la métrica IoU. Al considerar cómo detectados correctamente solo aquellos puntos que tienen al menos un 0.5 de IoU, se calculó la tasa de reconocimiento, obteniendo un 64 %, 55 % y 48 % para las clases *Pant*, *Shirt* y *T-Shirt*, respectivamente. Al ser bastante similares los resultados de IoU con la tasa de reconocimiento se puede concluir que en la mayoría de los casos el mejor punto localizado tiene un muy buen calce con respecto al marcado en el *ground truth*. Al considerar la detección de los dos puntos localizados utilizando la métrica IoU, se obtuvo un 34 %, 28 % y 7 % para las clases *Pant*, *Shirt* y *T-Shirt*, respectivamente. Al calcular la tasa de reconocimiento se obtuvo un 43 %, 33 % y 29 % para las clases *Pant*, *Shirt* y *T-Shirt*, respectivamente. Al ver que la tasa de reconocimiento mejora con respecto a la métrica IoU se puede concluir que el segundo punto localizado tiene un mal calce con respecto al *ground truth*, pero que para la clase *Pant* se podría desplegar satisfactoriamente la ropa un 43 % de las veces, la clase *Shirt* un 33 % de las veces y la clase *T-Shirt* un 29 % de las veces. Concluyendo para este sistema, que se logró el objetivo al lograr detectar los puntos de agarre que podrían desplegar la ropa de estas clases, sin guardar conocimiento previo del modelo o la ropa en específico.

Considerando el sistema de reconocimiento y clasificación de ropa utilizando percepción

continua, que tiene como objetivo clasificar la ropa utilizando la información extraída de una secuencia de imágenes de profundidad proveniente de diferentes sensores, mientras que uno de los brazos del robot mueve la prenda sobre una superficie plana. Se llevó a cabo el diseño y la construcción de la primera base de datos de percepción continua de ropa, *Continuous Clothing Dataset*. En las secuencias que componen la base de datos, la prenda de ropa comienza en una configuración aleatoria y se muestra como es manipulado por el robot. Utilizando esta base de datos se llevó a cabo la evaluación del sistema de reconocimiento, donde se obtuvo una precisión de clasificación promedio del 66.7 %, con una precisión específica por clase de 58.00 %, 41.58 %, 83.84 %, 67.00 %, 83.84 % para las clases: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel*, respectivamente. De los resultados, se puede apreciar que las clases *Sweater* y *Hand-Towel* obtienen los mejores resultados de reconocimiento. Esto se puede deber mayormente a las diferencias físicas que existen entre estas clases y las demás, las cuales poseen diferencias notorias a simple vista en cuanto tamaño y forma. Si bien la clase *Sweater* obtiene los mejores rendimientos en la clasificación, también posee los valores más altos de falsos positivos, lo que afecta mayormente a la clase *T-Shirt*. Además, se puede apreciar que existe una confusión entre las clases *Sweater* y *Shirt*, esto se debe al hecho de que las arrugas producidas al manipular la ropa durante la secuencia poseen propiedades bastante similares, dado a que ambas clases están hechas del mismo material (algodón).

Para concluir, si bien en el área de investigación de reconocimiento y clasificación de ropa existen diversos trabajos relacionados con el reconocimiento de ropa en distintas posiciones, estos trabajos están enfocados en sólo una parte del sistema, es decir, reconocen cuando la ropa se encuentra colgada o cuando esta se encuentra sobre la mesa. Sin embargo, gracias a este trabajo de tesis, se logró el diseño e implementación de un sistema que está enfocado en complementar la información obtenida a lo largo de una secuencia de ropa en movimiento, considerando que la información obtenida en todo el proceso tiene su grado de aporte para mejorar el reconocimiento de la clase. Lo cual confirma las hipótesis en este trabajo de tesis de que “el uso e integración de información proveniente de distintas deformaciones de la ropa puede mejorar el rendimiento y precisión del sistema de reconocimiento de ropa”.

5.2. Trabajo Futuro

Como trabajo futuro, la base de datos presentada tiene gran potencial para poder ser utilizada para entrenar una red profunda, aunque ahora contiene pocas secuencias y modelos para poder realizar satisfactoriamente el entrenamiento. Se podría automatizar un brazo robótico y dejarlo tomando posiciones aleatorias de la ropa durante un tiempo considerable para aumentar la cantidad de secuencias por modelo de ropa, aumentando además la cantidad de modelos. Esto daría una base para realizar una afinación de una red profunda ya existente o en caso de conseguir suficientes imágenes, entrenar una red desde cero.

Además, para el sistema de detección de puntos de agarre se propone realizar una detección de pose de la parte clave e utilizar esta información para mejorar la detección de los puntos. Esto además podría mejorar específicamente la detección de los puntos de la clase *T-Shirt*, ya que se podría mitigar la flexibilidad de la parte clave de esta clase. Finalmente, se propone realizar pruebas del sistema de detección de puntos de agarre con una plataforma robótica con una mayor cantidad de modelos y realizar la integración del sistema con un sistema anexo de doblado de ropa, para la validación del sistema en robots de servicio.

Capítulo 6

Bibliografía

- [1] CloPeMa clothes perception and manipulation. <http://www.clopema.eu>. Accessed: 2019-02-30.
- [2] Clothing part dataset. <http://www.iri.upc.edu/groups/perception/clothingDataset>. Accessed: 2019-02-30.
- [3] ROS rosbags. <http://wiki.ros.org/rosbag>. Accessed: 2019-02-30.
- [4] G. Alenya, F. Moreno-Noguer, A. Ramia, and C. Torras. Active perception of deformable objects using 3d cameras. In *2011 Workshop de Robótica Experimental*, 434-440, 2011.
- [5] G. Alenya, A. Ramisa, F. Moreno-Noguer, and C. Torras. Characterization of textile grasping experiments. In *ICRA Workshop on Conditions for Replicable Experiments and Performance Comparison in Robotics Research*, pages 1–6, 2012.
- [6] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. *Advances in Neural Information Processing Systems 13*, pages 831–837, 2001.
- [7] N. Bergström, C. H. Ek, D. Kragic, Y. Yamakawa, T. Senoo, and M. Ishikawa. On-line learning of temporal state models for flexible objects. In *Humanoids*, pages 712–718. IEEE, 2012.
- [8] C. Bersch, B. Pitzer, and S. Kammel. Bimanual robotic cloth manipulation for laundry

- folding. In *IROS*, pages 1413–1419. IEEE, 2011.
- [9] R. W. Conners, M. M. Trivedi, and C. A. Harlow. Segmentation of a high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing*, 25:273–310, 1984.
- [10] E. Corona, G. Alenyà, A. Gabas, and C. Torras. Active garment recognition and target grasping point detection using deep learning. *Pattern Recognition*, 74:629–641, 2018.
- [11] M. Cusumano-Towner, A. Singh, S. Miller, J. F. O’Brien, and P. Abbeel. Bringing clothing into desired configurations with limited perception. In *Proceedings of IEEE International Conference on Robotics and Automation (ICRA) 2011*, pages 1–8, May 2011.
- [12] C. I. N. de Estadísticas Chile. Encuesta nacional sobre uso del tiempo. In *ENUT*, 2018.
- [13] A. Doumanoglou, A. Kargakos, T. Kim, and S. Malassiotis. Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning. In *2014 IEEE International Conference on Robotics and Automation, ICRA 2014, Hong Kong, China, May 31 - June 7, 2014*, pages 987–993, 2014.
- [14] A. Doumanoglou, T.-K. Kim, X. Zhao, and S. Malassiotis. Active random forests: An application to autonomous unfolding of clothes. In D. J. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *ECCV (5)*, volume 8693 of *Lecture Notes in Computer Science*, pages 644–658. Springer, 2014.
- [15] M. Eck and H. Hoppe. Automatic reconstruction of b-spline surfaces of arbitrary topological type. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH ’96*, pages 325–334, New York, NY, USA, 1996. ACM.
- [16] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. *Multiscale vessel enhancement filtering*, pages 130–137. Springer Berlin Heidelberg, Berlin, Heidelberg, 1998.
- [17] A. Gabas, E. Corona, G. Alenyà, and C. Torras. Robot-aided cloth classification using depth information and cnns. In *Articulated Motion and Deformable Objects - 9th International Conference, AMDO 2016, Palma de Mallorca, Spain, July 13-15, 2016, Proceedings*, pages

16–23, 2016.

- [18] A. Gabas and Y. Kita. Physical edge detection in clothing items for robotic manipulation. In *18th International Conference on Advanced Robotics, ICAR 2017, Hong Kong, China, July 10-12, 2017*, pages 524–529, 2017.
- [19] W. Garage. Pr2 robot.
- [20] R. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6), 1973.
- [21] K. Industries. Hrp-2 robot.
- [22] X. Jiang, J. Gao, X. Hong, and Z. Cai. Gaussian processes autoencoder for dimensionality reduction. In V. S. Tseng, T. B. Ho, Z.-H. Zhou, A. L. P. Chen, and H.-Y. Kao, editors, *Advances in Knowledge Discovery and Data Mining*, pages 62–73, Cham, 2014. Springer International Publishing.
- [23] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1988.
- [24] Y. Kita, F. Kanehiro, T. Ueshiba, and N. Kita. Clothes handling based on recognition by strategic observation. In *Humanoids*, pages 53–58. IEEE, 2011.
- [25] Y. Kita, F. Kanehiro, T. Ueshiba, and N. Kita. Strategy for folding clothing on the basis of deformable models. In A. J. C. Campilho and M. S. Kamel, editors, *ICIAR (2)*, volume 8815 of *Lecture Notes in Computer Science*, pages 442–452. Springer, 2014.
- [26] Y. Kita and N. Kita. A model-driven method of estimating the state of clothes for manipulating it. In *WACV*, pages 63–69. IEEE Computer Society, 2002.
- [27] Y. Kita, T. Ueshiba, F. Kanehiro, and N. Kita. Recognizing clothing states using 3d data observed from multiple directions. In *13th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2013, Atlanta, GA, USA, October 15-17, 2013*, pages 227–233, 2013.
- [28] Y. Kita, T. Ueshiba, E. S. Neo, and N. Kita. A method for handling a specific part of clothing

- by dual arms. In *IROS*, pages 4180–4185. IEEE, 2009.
- [29] J. J. Koenderink and A. J. van Doorn. Surface shape and curvature scales. *Image Vision Comput.*, 10(8):557–565, Oct. 1992.
- [30] LAAS. Hrp-2 robot.
- [31] K. Lakshmanan, A. Sachdev, Z. Xie, D. Berenson, K. Goldberg, and P. Abbeel. A constraint-aware motion planning algorithm for robotic folding of clothes. In J. P. Desai, G. Dudek, O. Khatib, and V. Kumar, editors, *ISER*, volume 88 of *Springer Tracts in Advanced Robotics*, pages 547–562. Springer, 2012.
- [32] N. D. Lawrence. Gaussian process latent variable models for visualisation of high dimensional data. In *Proceedings of the 16th International Conference on Neural Information Processing Systems*, NIPS’03, pages 329–336, Cambridge, MA, USA, 2003. MIT Press.
- [33] A. Lee, A. Gupta, H. Lu, S. Levine, and P. Abbeel. Learning from multiple demonstrations using trajectory-aware non-rigid registration with applications to deformable object manipulation. In *Proceedings of the 28th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015.
- [34] A. X. Lee, H. Lu, A. Gupta, S. Levine, and P. Abbeel. Learning force-based manipulation of deformable objects from multiple demonstrations. In *ICRA*, pages 177–184. IEEE, 2015.
- [35] Y. Li, C.-F. Chen, and P. K. Allen. Recognition of deformable object category and pose. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2014.
- [36] Y. Li, Y. Wang, M. Case, S.-F. Chang, and P. K. Allen. Real-time pose estimation of deformable objects using a volumetric approach. In *IROS*, pages 1046–1052. IEEE, 2014.
- [37] Y. Li, Y. Yue, D. Xu, E. Grinspun, and P. K. Allen. Folding deformable objects using predictive simulation and trajectory optimization. In *IEEE/RSJ IROS 2015*. IEEE/RSJ, 2015.
- [38] H. Ling and D. W. Jacobs. Deformation invariant image matching. In *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2, ICCV ’05*, pages 1466–1473,

Washington, DC, USA, 2005. IEEE Computer Society.

- [39] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel. Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In *IEEE International Conference on Robotics and Automation, ICRA 2010, Anchorage, Alaska, USA, 3-7 May 2010*, pages 2308–2315, 2010.
- [40] I. Mariolis and S. Malassiotis. *Matching folded garments to unfolded templates using robust shape analysis techniques.*, pages 193–200. Berlin: Springer, 2013.
- [41] I. Mariolis and S. Malassiotis. Modelling folded garments by fitting foldable templates. *Mach. Vis. Appl.*, 26(4):549–560, 2015.
- [42] I. Mariolis, G. Peleka, A. Kargakos, and S. Malassiotis. Pose and category recognition of highly deformable objects using deep learning. In *ICAR*, pages 655–662. IEEE, 2015.
- [43] L. Martínez, J. Ruiz-del-Solar, L. Sun, J. P. Siebert, and G. Aragon-Camarasa. Continuous perception for deformable objects understanding. *Robotics and Autonomous Systems*, 118:220–230, 2019.
- [44] L. María Martínez and J. Ruiz-del Solar. Recognition of grasp points for clothes manipulation under unconstrained conditions. In *Proceedings of 22th RoboCup International Symposium, Lecture Notes in Computer Science.*, july 2017.
- [45] S. Miller, M. Fritz, T. Darrell, and P. Abbeel. Parametrized shape models for clothing. In *IEEE International Conference on Robotics and Automation (ICRA-11)*, pages 4861–4868, Shanghai, China, 2011. IEEE.
- [46] S. Miller, J. van den Berg, M. Fritz, T. Darrell, K. Y. Goldberg, and P. Abbeel. A geometric approach to robotic laundry folding. *I. J. Robotic Res.*, 31(2):249–267, 2012.
- [47] PCL. Point cloud library 1.7.
- [48] V. Petřík, V. Smutny, P. Krsek, and V. Hlavác. Robotic garment folding: Precision improvement and workspace enlargement. In C. Dixon and K. Tuyls, editors, *TAROS*, volume 9287 of *Lecture Notes in Computer Science*, pages 204–215. Springer, 2015.

- [49] M. Pietikainen, M. Pietikainen, T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29:51–59, 1996.
- [50] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras. Finddd: A fast 3d descriptor to characterize textiles for robot manipulation. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)*, pages 824–830, 2013.
- [51] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras. Determining where to grasp cloth using depth information. In C. Fernández, H. Geffner, and F. Manyà, editors, *CCIA*, volume 232 of *Frontiers in Artificial Intelligence and Applications*, pages 199–207. IOS Press, 2011.
- [52] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras. Using depth and appearance features for informed robot grasping of highly wrinkled clothes. In *ICRA*, pages 1703–1708. IEEE, 2012.
- [53] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras. Learning rgb-d descriptors of garment parts for informed robot grasping. *Engineering Applications of Artificial Intelligence*, 35(Complete):246–258, 2014.
- [54] R. Robotics. Baxter robot.
- [55] R. B. Rusu, N. Blodow, and M. Beetz. Fast point feature histograms (fpfh) for 3d registration. In *Proceedings of the 2009 IEEE International Conference on Robotics and Automation, ICRA'09*, pages 1848–1853, Piscataway, NJ, USA, 2009. IEEE Press.
- [56] R. B. Rusu, G. R. Bradski, R. Thibaux, and J. M. Hsu. Fast 3d recognition and pose using the viewpoint feature histogram. In *IROS*, pages 2155–2162. IEEE, 2010.
- [57] R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz. Persistent Point Feature Histograms for 3D Point Clouds. In *Proceedings of the 10th International Conference on Intelligent Autonomous Systems (IAS-10), Baden-Baden, Germany*, 2008.
- [58] M. Saraswat, A. K. Goswami, and A. Tiwari. Object recognition using texture based analysis. *International Journal of Computer Science and Information Technologies*, 4:775–782, 2013.

- [59] J. Šindler and V. Hlaváč. World representation for a dual-arm robot manipulating with clothes. Research Report CTU–CMP–2013–05, Center for Machine Perception, K13133 FEE Czech Technical University, Prague, Czech Republic, January 2013.
- [60] J. Stria and V. Hlavác. Classification of hanging garments using learned features extracted from 3d point clouds. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2018, Madrid, Spain, October 1-5, 2018*, pages 5307–5312, 2018.
- [61] J. Stria, D. Prusa, V. Hlavac, L. Wagner, V. Petrik, P. Krsek, and V. Smutny. Garment perception and its folding using a dual-arm robot. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 61–67, Sept 2014.
- [62] J. Stria, D. Průša, and V. Hlaváč. Polygonal models for clothing. In M. Mistry, A. Leonardis, M. Witkowski, and C. Melhuish, editors, *Advances in Autonomous Robotics Systems*, volume 8717 of *Lecture Notes in Computer Science*, pages 173–184. Springer International Publishing, 2014.
- [63] J. Sun, M. Ovsjanikov, and L. Guibas. A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. *Computer Graphics Forum*, 2009.
- [64] K. Sun, S. Rogers, G. Aragon-Camarasa, and J. Siebert. Recognising the clothing categories from free-configuration using gaussian-process-based interactive perception. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2464–2470, 2016.
- [65] L. Sun and G. Aragon-Camarasa. Free-configuration clothing dataset. <http://researchdata.gla.ac.uk/270/>. DOI:10.5525/gla.researchdata.270.
- [66] L. Sun, G. Aragon-Camarasa, W. P. Cockshott, S. Rogers, and J. P. Siebert. A heuristic-based approach for flattening wrinkled clothes. In A. Natraj, S. Cameron, C. Melhuish, and M. Witkowski, editors, *TAROS*, volume 8069 of *Lecture Notes in Computer Science*, pages 148–160. Springer, 2013.
- [67] L. Sun, G. Aragon-Camarasa, S. Rogers, and J. P. Siebert. Accurate garment surface analysis using an active stereo robot head with application to dual-arm flattening. In *ICRA*, pages 185–192. IEEE, 2015.

- [68] L. Sun, G. Aragon-Camarasa, S. Rogers, R. Stolkin, and J. P. Siebert. Single-shot clothing category recognition in free-configurations with application to autonomous clothes sorting. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2017, Vancouver, BC, Canada, September 24-28, 2017*, pages 6699–6706, 2017.
- [69] L. Sun, S. Rogers, G. Aragon-Camarasa, and J. P. Siebert. Recognising the clothing categories from free-configuration using gaussian-process-based interactive perception. In D. Kragic, A. Bicchi, and A. D. Luca, editors, *ICRA*, pages 2464–2470. IEEE, 2016.
- [70] H. Technologies. Baxter robot.
- [71] S. G. Thakku, Y.-C. Tham, M. Baskaran, J.-M. Mari, N. G. Strouthidis, T. Aung, C.-Y. Cheng, and M. J. A. Girard. A global shape index to characterize anterior lamina cribrosa morphology and its determinants in healthy indian eyes. *Investigative Ophthalmology and Visual Science*, 56(6):3604, 2015.
- [72] L. Twardon and H. J. Ritter. Interaction skills for a coat-check robot: Identifying and handling the boundary components of clothes. In *IEEE International Conference on Robotics and Automation, ICRA 2015, Seattle, WA, USA, 26-30 May, 2015*, pages 3682–3688, 2015.
- [73] J. van den Berg, S. Miller, K. Y. Goldberg, and P. Abbeel. Gravity-based robotic cloth folding. In D. Hsu, V. Isler, J.-C. Latombe, and M. C. Lin, editors, *WAFR*, volume 68 of *Springer Tracts in Advanced Robotics*, pages 409–424. Springer, 2010.
- [74] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *IEEE Conference on Computer Vision and Pattern Classification*, 2010.
- [75] P. C. Wang, S. Miller, M. Fritz, T. Darrell, and P. Abbeel. Perception for the manipulation of socks. In *IROS*, pages 4877–4884. IEEE, 2011.
- [76] X. Wang, X. Jiang, J. Zhao, S. Wang, T. Yang, and Y. Liu. Picking towels in point clouds. *Sensors*, 19(3):713, 2019.
- [77] B. Willimon, S. Birchfield, and I. D. Walker. Classification of clothing using interactive

- perception. In *ICRA*, pages 1862–1868. IEEE, 2011.
- [78] B. Willimon, I. D. Walker, and S. Birchfield. Classification of clothing using midlevel layers. In *ISRN Robotics*, 2013.
- [79] B. Willimon, I. D. Walker, and S. Birchfield. A new approach to clothing classification using mid-level layers. In *ICRA*, pages 4271–4278. IEEE, 2013.
- [80] D. Xu and S. F. Chang. Video event recognition using kernel methods with multilevel temporal alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):1985–1997, Nov 2008.
- [81] D. Xu, Y. Huang, Z. Zeng, and X. Xu. Human gait recognition using patch distribution feature and locality-constrained group sparse representation. *IEEE Trans. Image Processing*, 21(1):316–326, 2012.
- [82] K. Yamazaki. Grasping point selection on an item of crumpled clothing based on relational shape description. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14-18, 2014*, pages 3123–3128, 2014.
- [83] K. Yamazaki. A method of grasp point selection from an item of clothing using hem element relations. *Advanced Robotics*, 29(1):13–24, 2015.
- [84] K. Yamazaki and M. Inaba. Clothing classification using image features derived from clothing fabrics, wrinkles and cloth overlaps. In *IROS*, pages 2710–2717. IEEE, 2013.

Anexo A

Continuous Clothing Dataset

Durante el desarrollo de este trabajo de tesis se creó la primera base de datos de ropa utilizando percepción continua. La base de datos contiene secuencias de video RGB-D de 15 modelos de ropa distintos. Son 5 modelos de cada una de las siguientes clases ropa: *T-Shirt*, *Shirt*, *Sweater*, *Pant* y *Hand-Towel* (Figuras A.2, A.3, A.4, A.5 y A.6). Cada modelo se capturó iniciando desde 10 configuraciones iniciales aleatorias distintas y fue levantado por el robot, resultando un total de 150 secuencias. Cada secuencia fue grabada desde dos sensores Asus Xtion pro live ubicados en las ‘muñecas’ de los brazos del robot (ver Figura A.1). Esta base de datos se encuentra disponible gratuitamente en: <http://dx.doi.org/10.5525/gla.researchdata.669>. Además, para facilitar el trabajo con esta base de datos se creó un repositorio público disponible en: https://github.com/lmmartine/cp_clothes_db.

La base de datos está compuesta por una colección de rosbags, donde cada uno de ellos guarda la información de una secuencia. Los rosbag contienen las imágenes de color, imágenes de profundidad, nubes de puntos, información de los sensores y las transformaciones cinemáticas del robot, de ambos sensores a 15 Hz de cada secuencia. Los tópicos guardados en cada rosbag con su respectiva información y *rate* se puede observar en la Tabla A.1

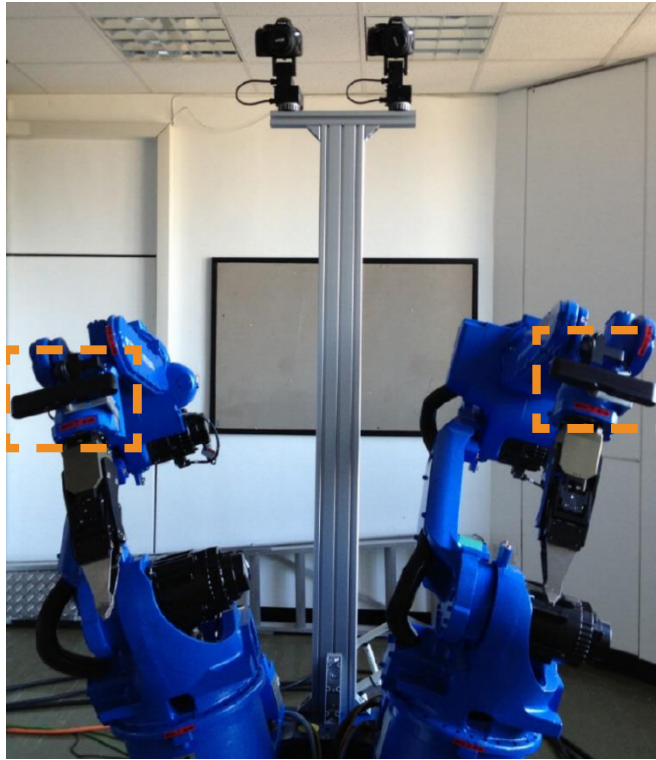


Figura A.1: Robot CloPeMa con las cámaras RGB-D marcadas en un rectángulo naranja.

Tabla A.1: Información y *rate* promedio de los tópicos de cada rosbag

Topico	Información	Rate
<i>/tf</i>	Transformaciones cinemáticas del robot	128.9
<i>/xtion1/depth/camera_info</i>	Información del sensor de profundidad xtion1	14.5
<i>/xtion1/depth/image_raw</i>	Imagen de profundidad del sensor xtion1	15.0
<i>/xtion1/depth/points</i>	Nube de puntos del sensor xtion1	14.5
<i>/xtion1/rgb/camera_info</i>	Información del sensor rgb xtion1	15.0
<i>/xtion1/rgb/image_raw</i>	Imagen de color del sensor xtion1	14.9
<i>/xtion2/depth/camera_info</i>	Información del sensor de profundidad xtion2	14.9
<i>/xtion2/depth/image_raw</i>	Imagen de profundidad del sensor xtion2	15.0
<i>/xtion2/depth/points</i>	Nube de puntos del sensor xtion2	15.0
<i>/xtion2/rgb/camera_info</i>	Información del sensor rgb xtion2	15.0
<i>/xtion2/rgb/image_raw</i>	Imagen de color del sensor xtion2	15.0

Como se observó en la revisión bibliográfica (Sección 2) existen dos enfoques altamente utilizados en robótica para el reconocimiento de ropa. El primer enfoque, es cuando la ropa se encuentra sobre la mesa y sin realizar ninguna interacción previa, el robot toma una imagen, extrae características y clasifica la ropa. El segundo enfoque, es cuando la ropa se encuentra sobre la

mesa, el robot manipula un punto de la ropa y la deja colgando desde uno de los *grippers*. El robot analiza una o varias imágenes de la ropa colgando y clasifica la ropa. Esta base de datos intenta proporcionar el conjunto de prueba para ambos enfoques de reconocimiento y para el propuesto en este trabajo de tesis, realizar reconocimiento utilizando percepción continua.



Figura A.2: Modelos de la clase *Pant* de la base de datos



Figura A.3: Modelos de la clase *Shirt* de la base de datos



Figura A.4: Modelos de la clase *Sweater* de la base de datos



Figura A.5: Modelos de la clase *Hand-Towel* de la base de datos



Figura A.6: Modelos de la clase *T-Shirt* de la base de datos