



Type independent hierarchical analysis for the recognition of folded garments' configuration

Dimitra Triantafyllou¹ · Panagiotis Koustoumpardis¹ · Nikolaos Aspragathos¹

Received: 16 November 2020 / Accepted: 29 March 2021 / Published online: 19 May 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

This paper proposes a hierarchical visual architecture for perceiving garments' configuration independently from their type for the robotic unfolding task. Special focus is given on the decomposition of folded configurations into low- and high-level features. The low-level features comprise junctions of edges, which act as localized indicators of the clothing article's state, while the high-level components refer to its layers and the axis that unites them. The proposed methodology extracts and classifies the low-level components into indicators of folds, overlaps, garment's edges and corners and through their combination reconstructs the axis and the layers of the garment. The methodology is independent from the garment's shape while it uses depth sensors so that it can deal with garments of various colours, patterns and decorative features. Experiments showed the effectiveness of the method in scenarios with onefold or twofold and in different datasets, proving the extensibility of the approach.

Keywords Garments · Unfolding · Layer detection · Axis detection · Depth images

1 Introduction

The intense contemporary lifestyle and the increasing ageing of population have led the scientific world to research on the relief of people from tedious, tiring and time-consuming housework. In this scope, service robots were introduced and their potential to handle housekeeping has been widely explored.

One of the most challenging tasks among the household chores is the robotic manipulation of laundry and, in particular, garment unfolding, due to the highly non-rigid nature of garments. The infinite configurations that a garment can change during manipulations, create difficulties to the recognition of its state and the detection of the desirable grasp points to handle it. In the robotic literature, two main approaches are explored for the robotic unfolding. In the first

one, all the manipulations of the garment take place in the air while, in the second one, partial unfolding is achieved in the air and the rest of the procedure is completed on a working table.

This paper presents a hierarchical visual architecture for perceiving a garment's folded configuration on a working table, independently of its type (towel, shirt, shorts etc.) for the unfolding task. Through conceptual analysis, the garment's configuration is broken down to its constituent components that are divided into high-level and low-level features. From a high-level point of view, these components comprise the garment's layers and the folding axes that connect them, and from a lower point of view, they refer to features such as junctions of the garment's edges that indicate the garment's configuration at localized areas. Using a bottom-up approach, these junctions are recognized, classified and interconnected leading to the recognition of the garment's axes and layers. The proposed method is independent from the garment's shape, while input from depth sensors, such as kinect, is utilized so that it can deal with clothes with various colours, patterns and decorative features (e.g. pockets).

The proposed method is integrated in a garment unfolding procedure introduced in previous works [1, 2]. The task is divided in three phases that include: (1) the robotic

✉ Dimitra Triantafyllou
dimtria@upatras.gr
Panagiotis Koustoumpardis
koust@upatras.gr
Nikolaos Aspragathos
aspragat@upatras.gr

¹ Department of Mechanical Engineering and Aeronautics, University of Patras, 26504 Rio, Greece

manipulation in the air to transform the garment from a random, crumpled state to a half-folded, planar configuration and its placement on a working table, (2) the recognition of the half-folded garment's upper layer that needs to be unfolded, (3) the robotic unfolding of the detected layer. Since the first phase has been extensively explored and tested in previous work [1], in this paper, the description of the method focuses on the second stage, while the third phase is an independent future work.

The main advantage of this method is its independency from garments types (e.g. skirt, T-shirt, shorts) since it is based on features that apply in all garments and do not relate to their shapes. Furthermore, utilizing only depth data, it can deal with garments with a variety of colours and patterns. Inspired by the use of junctions for the extraction of 3D information from 2D images [3], a similar approach adjusted to the planar data and small depth differences on the surface of garments is proposed. Therefore, a “junctions’ dictionary” is created that detects and translates primitive features, i.e. junctions, to localized configurations. Based on this “dictionary”, a new method for folding axis extraction and the detection of the garment's upper layer is presented. Finally, both axis and upper layer detection methods are evaluated, apart from the half-folded configurations, in cases with two-fold to explore the generalization ability of the methods.

To sum up, the main contributions of this paper are enumerated as follows:

1. A new method for the perception of the folded configuration of a garment on a working table based on hierarchical analysis of its components is introduced. The method integrates human knowledge on the semantics of visual features that indicate the garment's configuration and uses machine learning approaches to classify them and combine them.
2. Low-level features of garments, i.e. junctions of edges occurring from depth differences, are attributed to semantic information on localized configuration information, leading to a “dictionary” that indicates folds, overlaps, edges or corners of the cloth while it can also handle wrinkles.
3. The proposed methodology is independent from the garment's type while it can deal with garments with various colours, patterns and decorative features.
4. The experiments tested the effectiveness of the methodology to scenarios with one or two folds, considering the number of folds either a priori known or unknown.

The remainder of the paper is organized as follows. Section 2 presents related work, while Sect. 3 introduces the hierarchical analysis of the folded configuration into its high and low-level components. Sects. 4 and 5 analyse the extraction of the low- and high-level components of the folded garment

respectively. Finally, experimental results are presented in Sect. 6, while the paper concludes in Sect. 7.

2 Related work

The robotic unfolding has attracted the interest of many scientists that wanted to explore this complicated and challenging task. The main approaches met in the literature can be divided in two major categories. In the first category, the whole unfolding procedure takes place in the air, while, in the second one, the task is separated in two stages: (1) the robot handles the garment in the air so as to transform it into a planar, half folded configuration, (2) the half-folded garment is lied down on a working table so as to determine its state and proceed to the completion of unfolding. An interesting observation is that the first category requires the classification of the garment to specific types of clothing (e.g. skirt, shorts, T-shirt, long-sleeved shirt, towel) while in the second one there are certain approaches that are independent of the garment type. In all the cases bi-manual robots are used to achieve handling.

2.1 Unfolding completed entirely in the air

The methodologies that complete the unfolding task in the air usually involve in their approach; the classification of the garment to predefined types [4–8] or its type is a priori known [9–12]. The garment's configuration is usually recognized while hanging and, then, appropriate points according to its type are detected and selected as grasp points that lead to complete unfolding (e.g. the shoulder of a shirt).

The approaches that require previous knowledge of the garment's type are either type specific, for example in [9] only towels are handled, or can generalize to more types as long as they are provided as input by the user [10–12]. Therefore, the authors of [9] address the robotic unfolding of towels by extracting geometric cues from stereo images, depicting the towel hanging from a random point, in order to detect corner points. The selected points are grasped in order to bring the manipulated piece of clothing into a spread-out state. In [10], two grasp points are selected simultaneously aiming at the natural unfolding of garments randomly placed on a table. The hem lines of the garment are detected using a range image while the selection of the grasp points is based on global shape similarity with a list of training data. Furthermore, the authors of [11] simulate garments hanging from various points through mass-spring models. These models are compared to images from stereo cameras depicting real hanging garments in order to predict their configuration. If the prediction is not robust, “recognition aid” actions, such as rotation or spreading, are programmed to facilitate the evaluation of the garment's state. In a similar

approach [12], a database with simulated models of hanging garments is created. The 3D model of the real garments acquired by a depth sensor is compared through a feature extraction and matching scheme algorithm to the simulated models in order to predict their pose and select appropriate grasping points for unfolding.

Contrary to the aforementioned methods that considered the garment type known in advance, the following approaches [4–8] incorporated a type classification step into their flow. To achieve it, training datasets are utilized that comprised of specific types of garments, real or simulated, hanging while being grasped from different points. Therefore, they do not require user input, while they can handle only garment types that are included in their datasets. The authors of [4] applied the Random Forests method for the classification of hanging garments to their types while Hough Forests suggest grasp points that lead to their unfolding. The grasp point detection is further improved in [5] using Active Random Forests to deal with ambiguous situations. Furthermore, in [6] the classification task and the two grasp points detection required for the unfolding are achieved through an hierarchy of three levels of Convolutional Neural Networks (CNNs) trained in both simulated and real data. In a similar approach, using CNNs with different architecture, the whole unfolding procedure is executed in [7] while in [8] convolutional neural networks are used for the classification and pose estimation of the garment, nevertheless, without concluding the unfolding chain by finding grasp points.

2.2 Two-stage unfolding

In the two-stage unfolding, at the beginning, the garment is partially unfolded in the air and, then, it is placed on a working table for further unfolding manipulations. Although some researchers explored both of these stages, others focused only in one of the two of them. In this scope, the analysis of the related papers is divided in these two parts.

2.2.1 Partial unfolding in the air

In this group of methodologies, the garment is grasped by a random point and held up in the air. The goal is to detect points that are located at the outline of the garment (as it is defined when the garment is in an unfolded state) and, by consecutive regrasping, to transform it to a planar half folded state. To achieve this task, a heuristic approach is proposed in [13]. The lowest point of the hanging garment is detected and grasped in iterations leading in this way to the desirable planar folded configuration. Following a similar logic, the lowest hanging procedure is repeated in [14], where a Hidden Markov Model is utilized in order to track the clothing article's configuration during handling by matching

its outline with existing templates. Although the heuristic approach utilized in the aforementioned approaches provides an easy solution, it could create problems to the handling of real size garments due to the robots' workspace limitations. Therefore, the authors of [15] proposed the detection of outline points based on the appearance of shadows created by folds. Nevertheless, as it is reported in that work, the approach suffers from changes of the environmental conditions. Moreover, a method that detects through geometric features folds and hemlines of the hanging garment, i.e. parts of its outline, is presented in [1]. The method provides the option to eliminate the proposed grasp candidates to the robot's workspace while it utilizes a depth camera remaining independent from environmental conditions.

2.2.2 Manipulations onto a working table

At the second phase of unfolding, the robot has grasped the garment from two outline points and laid it in a folded configuration onto a working table. The goal is to transform the garment from a folded planar state to a spread-out configuration. The utilized methods either try to re-configure the garment based on their type (type dependent) [1, 14] or by detecting the upper layer formulated by the unfolding manipulations [2, 16] (type independent).

In this scope, referring to the first category, [1, 14] match the garments to their types and configurations using template matching techniques. Based on the garment type, the appropriate grasp points, that are defined a priori by the user for each template, are selected in order to lead it to an unfolded state.

On the other hand, towards a type-independent procedure, a specialized installation illuminating the folded garment from three different orientations is utilized in [16]. The goal of this installation is to accentuate the edges of the overlapping parts of the garments so as to extract the upper layer of the fold. Furthermore, fold detection is explored as well in [17] and [18], although in these cases the fold formulation is not specifically connected to previous robotic manipulations. Both approaches are based on depth information to create clusters that indicate the upper layer of the fold while as a next step they proceed to the detection of the fold's axis. In [17], the axis detection is achieved by selecting the axis that provides the shortest contour once the garment is unfolded. Moreover, in [18], the axis detection is defined by a "bumpiness" criterion that evaluates all the edges of the outline of the garment and selects the one that is more smoothly connected with the detected layer.

The last three methodologies [16–18] are the most highly related to the approach presented in this paper. Comparing to [16], both methodologies focus on the edges created by the formulation of two layers. Nevertheless, [16] requires a specialized illumination installation while it handles only

single-colour garments. Moreover, apart from the fold, their work does not explore other features of a folded configuration such as axes. Regarding the methods described in [17, 18], the similarity with the proposed methodology lies in their goal to extract the upper layer of a folded configuration and its axis. Nevertheless, [18] does not focus on the extraction of the whole upper layer and compromises on detecting a part of it so that it can be grasped. Although this strategy can be sufficient for thick garments, which are more stiff, and for small folds, it cannot generalize to thinner garments and various folds. On the other hand, [17] proposes a method that extracts the whole area of the upper layer using a variety of garments with different shapes, fabrics and thicknesses. The approach is not related to manipulations that transform the garment into a planar folded configuration; therefore the way the folds were formulated is unknown. The method maps the garment to two classes referring to the upper and lower layer; nevertheless, it is not clear if it can cope with cases that the visible part of the lower layer is separated in more segments by the upper layer or even in cases that the lower layer is completely covered by the upper layer. Nevertheless, these cases are common during the robotic unfolding task. Moreover, its experimental results refer only to configurations with one fold. Contrary to this work, our proposed method is focusing on the edges that define the area of the layers and their interrelations and not on the clusters formed by the layers. This provides us the capability to handle the aforementioned cases occurring during robotic unfolding. Furthermore, the extensibility of our method was tested with experiments on cases of configurations with two folds.

3 Hierarchical analysis of the planar folded configuration

The method presented in this paper is a part of an unfolding procedure that, through robotic manipulations in the air, transforms the garment from a random, crumbled configuration in a half-folded, planar state and lays it on a working table so that the two layers in which the garment is divided are extracted. Once the upper layer is known, the garment can be completely unfolded. Since the robotic manipulations that lead the garment to a half folded state are analysed in detail in [1], the present paper focuses on the analysis of the planar, folded configuration based on a hierarchical architecture. Primitive, low-level features that derive from junctions of edges and are indicative of the localized configurations are extracted in order to lead to the perception of high-level features, such as the axis and the layers of the garment. The outline of the unfolding procedure and the schema of the hierarchical architecture are presented in Fig. 1 while Fig. 2 depicts the overview of the proposed method's steps in order to extract the low- and high-level features.

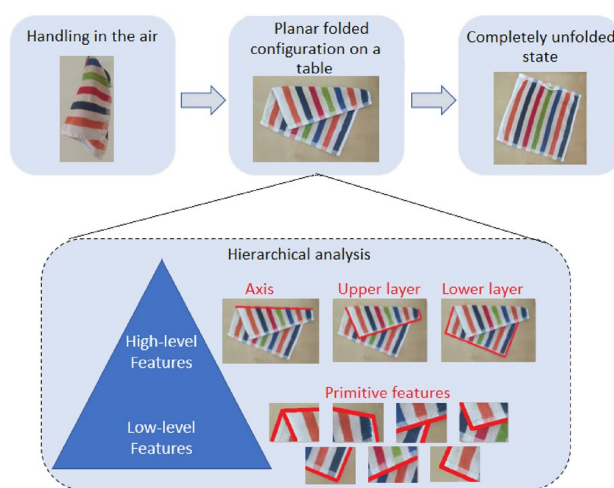


Fig. 1 Outline of the unfolding task focusing on the analysis of the planar folded configuration of the garment

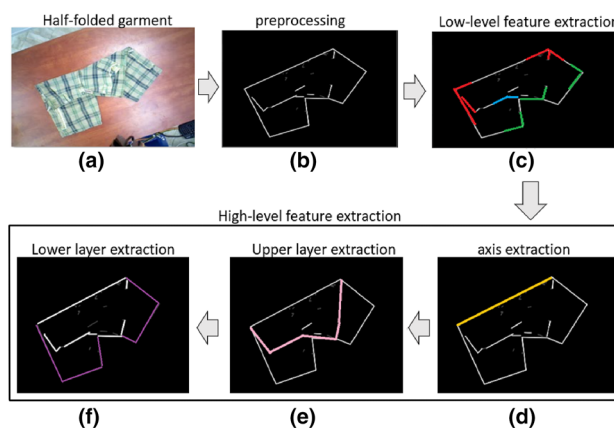


Fig. 2 Outline of the proposed method to extract the high and low-level features: **a** the garment lied onto the working table, **b** preprocessing for edge detection, **c** extraction of low-level features/examples of junctions are marked in different colours, **d** axis extraction, **e** upper level extraction, **f** lower level extraction (colour figure online)

This paper introduces a new method that relates with two previous works of the authors and expands them under a new perspective. In the first one [2], an edge-oriented approach based on the depth differences on a half folded garment, extracts the edge sequence that connects the two ends of the a priori known folding axis and forms the upper layer. The search of the edge sequence is based on criteria like the proximity, continuity and collinearity of the edges while a neural network makes the final decision on the choice of the edges that participate on the sequence of the upper layer outline. In another work [19], applied on colour images, junctions of edges are detected and classified based on their schema (i.e. “arrow”, “L”, “T” and “T” junctions) to different indicators of localized garment

configuration, such as folds and overlaps, creating in this way a junctions’ “dictionary”.

The work presented in this paper enhances and expands these methods while it proposes a new hierarchical approach for the comprehension of a half folded garment’s configuration. In particular, it applies the “dictionary” on depth data, instead of colour, and expands it to more categories utilizing depth information while it also incorporates cases with edges created from noise. Furthermore, the junctions’ classification is achieved by means of Random Forests and not simple geometrical criteria, coping, in this way, with more complex situations (e.g. noise). This new, updated “dictionary” is integrated in an upper layer edge sequence search procedure inspired from [2] but enriched with new features and methods that express the hierarchy between the low-level and high-level components of the garment’s configuration. Therefore, in the current paper, the sequence of the upper layer edges is determined by the detection of junctions and the application of the introduced “dictionary”. Moreover, other features of the garment, such as the axis and the lower layer, are explored providing a better description of the garments’ configuration while more scenarios like twofold are instigated.

3.1 Features of the hierarchical analysis

Observing garments of various types being half-folded on a table, several features that facilitate the comprehension of the cloth’s configuration are detected (Fig. 3). From a high-level point of view, the garment can be divided in two main layers: (1) the upper layer and, (2) the lower layer, while the common edge that unites the two layers is the folding axis

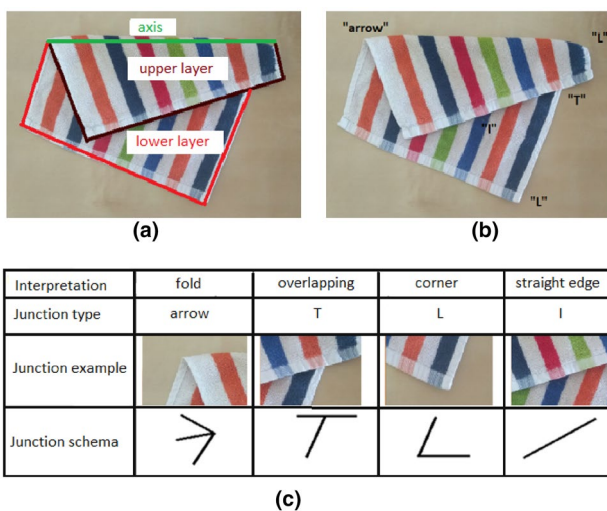


Fig. 3 High and low-level features: **a** The high-level features: upper layer, lower layer and axis, **b** examples of low-level features, **c** analysis of the low-level features

(Fig. 3a). At the same time, focusing on a lower-level point of view, it is observed that the layers and their interrelation are defined by junctions of edges, i.e. edges that intersect (Fig. 3b). These junctions have various schemas, each one with different semantics, creating a junctions’ “dictionary” [19]. Using this “dictionary”, an overlap, a fold, a corner or just an edge of the garment can be identified (Fig. 3c) while their combination leads to the higher-level features, i.e. the two layers and the folding axis. In the next paragraphs these low- and high-level features are described in detail.

Junctions The groups of edges, i.e. junctions, studied in this paper can be divided according to their schemas to features with different and specific semantics [19]. The main types of junction that were observed can be classified based on their 2D geometry to “arrow”, “T”, “L” and “I” junctions. Each junction type symbolizes a different indicator of the possible garment’s configuration. In particular, the “arrow” depicts a fold, “T” indicates an overlap, and “I” refers to an edge, while “L” may have two explanations: (1) a garment’s corner, or (2) a fold with the folding part not visible to the observer. In the case of interest, where the garment’s configuration occurs from manipulations in the air, it is a safe approach to treat “L” junctions located on the folding axis as folds while “L” junctions at the rest parts of the garments are more likely to be real corners of the garment’s outline (Fig. 3b).

Apart from the 2D schemas, another important factor, introduced in this paper, that indicates the configuration of a garment around a junction is the depth differences of the areas formed between the junction’s edges. Based on them, the garment’s layers (upper and lower) can be mapped. Therefore, sub-categories for each type of junction are defined and depicted in Fig. 4, expanding in this way the dictionary proposed in [19]. Our goal is not to exhaustively enumerate all possible combinations of depth levels and angles between the junction edges but to focus on the main cases met on the garment once it is led in a half folded configuration. In Fig. 4, while in both cases of “arrow” junctions there is a fold, according to the depth levels the configuration is different. Similar examples for

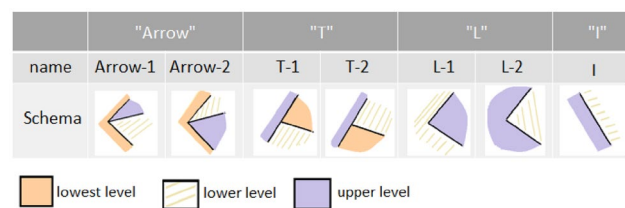


Fig. 4 Junction types according to the 2D geometry and the depth levels between the junctions’ edges. The three types of levels show the relative relation between the depths of a junction and do not refer to specific layers

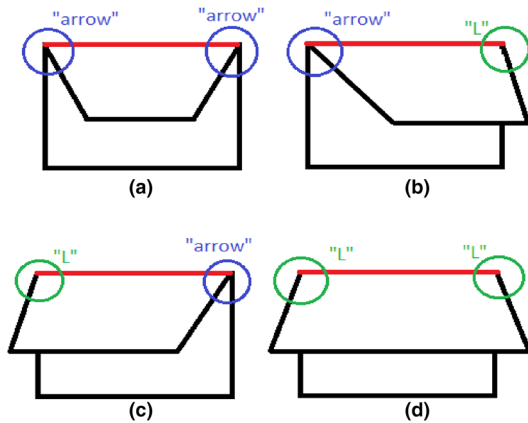


Fig. 5 Axis types based on the junction types that lie to its ends. The axes are highlighted with red colour (colour figure online)

the rest of the junction types are also depicted (nevertheless, “I” junction has only one option, since it is a matter of orientation to alter the sides of the depths).

Folding Axis In order to define the possible types of axis that may occur once, the garment is led in a half folded configuration, the junctions’ “dictionary” is utilized. In particular, four types of axis are observed based on the junction-types that lie to the axis’ two ends. To facilitate the definition of the axis-types, we consider that the axis is located horizontally and that the fold is downwards as in Fig. 5. The possible junction combinations, from the left to the right end of the axis, are:

1. An “arrow-1” and an “arrow-2” junction (Fig. 5a).
2. An “arrow-1” and an “L-2” junction (Fig. 5b).
3. An “L-2” and an “arrow-2” junction (Fig. 5c).
4. Two “L-2” junctions (Fig. 5d).

Although we cannot exclude the existence of other combinations of junctions that form an axis, these four categories were the result of the manipulations stage (as it is described in [1]) in 100 tests.

Upper layer The upper layer is defined by a sequence of edges that along with the folding axis form a non-intersecting polygon. The edges of the layer are recognized based on the detection and classification of the junctions formed on the garment, the configuration they imply and their interrelations. A more detailed description of the proposed upper layer extraction approach is provided in Sect. 5.2.

Lower layer The edges located onto the area of garment belong either to the upper layer, the lower layer or refer to noise (wrinkles, edges occurring from decorative features of the garment such as pockets). The edges located to the outline of the half-folded garment that do not belong to the upper layer are considered edges of the lower level.

4 Low-level features extraction

This section presents the extraction and classification of the low-level features. It begins with the preprocessing of the depth image and the extraction of the garment’s edges, which is a process firstly introduced in [2]. Furthermore, it introduces the detection of junctions based on the detected edges and presents an extended version of the junctions’ dictionary in comparison to [19]. Finally, the implementation of the junctions’ “dictionary” utilizing the method of Random Forests is presented.

4.1 Preprocessing and extraction of oriented edges

Once the garment is located half-folded on a working table, a preprocessing procedure, which was presented in [2], extracts oriented edges that will be used for the detection of its low- and high-level features (Fig. 6). The preprocessing starts with the acquisition of a depth image using a range sensor. The image is filtered using bilateral filtering to diminish noise and enhance edge pixels that are extracted by Canny detector [20]. The detected pixels are separated into clusters by the DBSCAN algorithm and, through line simplification [21], result in straight line segments. Finally, to diminish noise, all edges with length smaller than a predefined threshold are rejected. The edges extracted from the surface of a garment laid on a table can be seen in Fig. 6b.

The extracted edges are attributed with an orientation so as to facilitate further analysis of the garment’s configuration and the extraction of its upper level. This orientation depends on: (1) the depth differences between the two sides of the edge (the depth difference calculation of an edge is described in detail in Sect. 4.4); (2) the direction used (clockwise or not) in order to reveal the sequence of edges that comprise the outline of the upper layer (Fig. 7). In particular, the orientation of an edge with two endpoints p and q is defined by the formula:



Fig. 6 Extraction of edges on the surface of a garment: **a** the colour image of the garment, **b** the extracted edges using depth information (colour figure online)

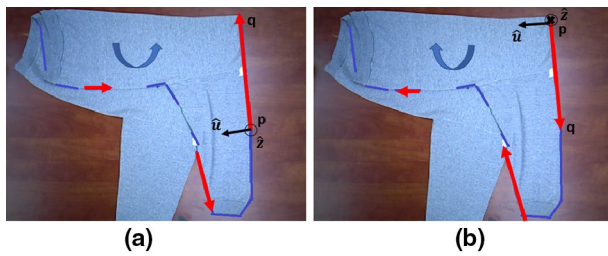


Fig. 7 The edges’ orientation based on the direction of the upper layer’s sequence exploration and the depth differences: **a** not-clockwise direction of the layer’s exploration, **b** clockwise direction of the layer’s exploration. The blue edges correspond to examples of edges extracted through preprocessing while on three of them the oriented edge is depicted with red arrows (colour figure online)

$$\frac{\overrightarrow{pq}}{|\overrightarrow{pq}|} = \hat{u} \times \hat{z} \tag{1}$$

where \hat{u} is the unit vector that is perpendicular to the edge and its direction occurs from the depth difference between the two sides of the edge (from the most distant side of the edge comparing to the camera to the closest one) and \hat{z} is the unit vector of the perpendicular to the table rotation axis, determining the upper layer exploration detection. Two examples on the edges’ orientation extraction for clockwise and anticlockwise exploration of the upper layer can be seen in Fig. 7.

4.2 Junctions’ extraction

The results of the preprocessing step show that the information obtained from the depth sensor differentiates from the perfect scenario, presenting gaps on the edges and extra noise (Fig. 8). The small depth differences of the garment’s layers can be sometimes in the range of the sensor’s error tolerance [22]; hence, parts of the folded garment’s layer might not be detected. Moreover, wrinkles of the fabric or decorative features of a garment, such as pockets, large seams, collars etc., could create extra edges that complicate the task of layer detection (from now on, these edges will be referred as noise edges).

Taking into account the aforementioned difficulties, a method is developed aiming to group edges of close distance into junctions, classify them and, based on the configuration they imply, connect them to reveal the garment’s layers. Therefore, the first important step is to detect and classify the edge junctions correctly and unaffected from the noise edges. Since the junctions of interest include two (“I” and “L”) or three (“arrow” and “T”) edges, only combinations of edge pairs or triads are taken into account.

To avoid investigating all possible edge combinations, a distance metric between the oriented edges is taken into account. In particular, in a pair of edges, an edge is considered to

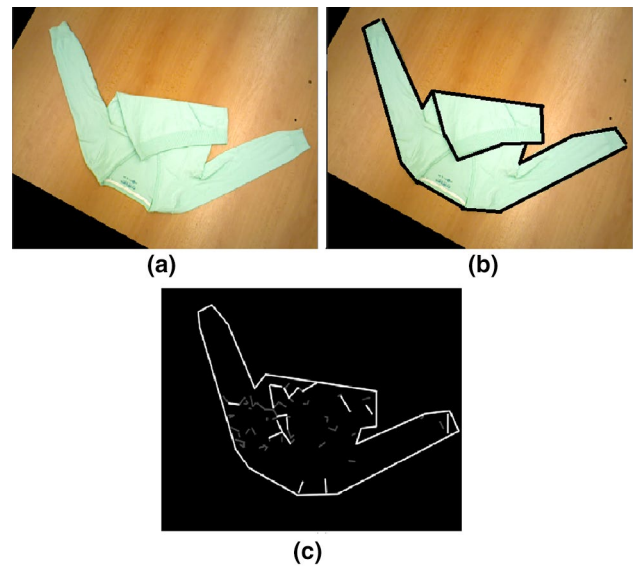


Fig. 8 The differences between the ideal edge extraction and the true edges that are extracted: **a** a colour image of the folded garment, **b** an example of an ideal situation where all the edges of the layers are visible and all the noise edges of the wrinkles or the V-neck are not visible, **c** the true edges as they are extracted through the preprocessing procedure (colour figure online)

participate into a junction with another edge when the Euclidean distance between the end of the first one and the beginning of the second one is below a threshold that depends on the garment’s dimensions. Therefore, as it is depicted in Fig. 9, an edge \overline{ab} is connected with another edge \overline{cd} to form a junction when:

$$\text{dist}(b, c) < \lambda \cdot \max_{dx} \tag{2}$$

where $\text{dist}(b, c)$ is the Euclidean distance between the points b and c , λ is a constant value that represents a percentage of

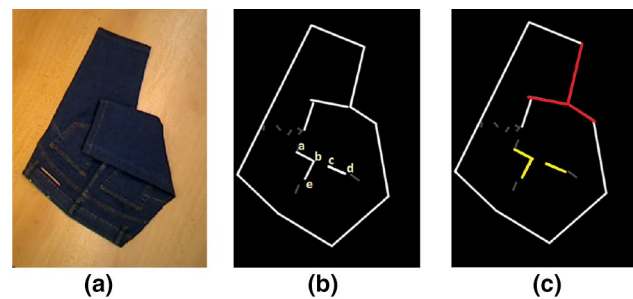


Fig. 9 A true “T”-junction and a combination of an “I”-junction and an edge occurring from a decorative seam: **a** the colour image of the folded garment, **b** the edges extracted through the preprocessing procedure-edges \overline{ab} , \overline{be} and \overline{cd} form a junction, **c** the “T”-junction and the combination of “I”-junction and an edge occurring from a seam are highlighted in red and yellow respectively (colour figure online)

\max_{dx} and \max_{dy} is the maximum distance of the points of the garment’s outline on the horizontal axis of the sensor’s image. The \max_{dx} acts as a rough estimator of the garment’s dimensions. Finally, if an edge satisfies Eq. (2) for more than one edge, triads of junction candidates are formed.

We should point out, that since this criterion is applied on oriented edges, it ensures that the edges are connected in such way that they both have their highest parts on the same side. The use of relative values (higher/lower) is preferred in relation to the actual depth values since different parts of the same layer might have different values due to wrinkles, decorative features of the garment that have extra volume, e.g. pockets, or even due to the error tolerance of the depth sensor.

Moreover, another criterion for the formulation of a junction is applied in cases of edges that belong to the outline of the folded garment. In these cases, the inclination of the internal edges regarding the corners of the outline of the garment is considered as a strong indicator of a non-accidental relation that could refer to an “arrow” or a “T” junction providing useful information on the garment’s configuration. This inclination is measured by calculating the distance of the corners’ top from the lines defining the internal edges (Fig. 10). An edge with short distance from the top of a corner is considered as a good candidate to form a junction. Therefore, an internal edge \overline{ab} and a corner formed from outline edges with top point c are considered to form a junction when:

$$d_{\text{per}}(a, b, c) < \kappa \cdot \max_{dx} \tag{3}$$

where $d_{\text{per}}(a, b, c)$ is the perpendicular distance of point c from the edge \overline{ab} , κ is a constant value that represents a percentage of \max_{dx} and \max_{dy} is defined in Eq. (2).

It should be commented that for the junction formulation, edges of the garment are merged; nevertheless, the case of splitting larger edges to smaller ones in order to investigate new junctions is not explored. The main reason

is that, during preprocessing (Sect. 4.1), line simplification is applied on large contours resulting in smaller segments. The line simplification parameters were adjusted through trial-and-error experiments so that the occurring edges represent adequately the curvature and the different parts of the layers’ outline. Rare cases of segments that would provide useful information and new junctions if they were further split might occur, nevertheless creating and testing more edges would lead to extra computational cost.

4.3 Junctions’ “dictionary” elements

The junction types studied in this paper are composed from two or three edges. In the case of two edges, the classification is easy and can be based on the angle and the depth difference of the two areas formed between them. A similar approach could be applied on junctions with three edges; nevertheless, during the recognition of the junctions formed on the cloth, there might be edge combinations that include noise edges. These cases make it difficult to distinguish them from a junction corresponding to the categories “T” and “arrow”. For example, a combination of an “I” junction with an edge occurring from a decorative seam would be easily confused with a “T” junction (Fig. 9). Therefore, to deal with this problem extra elements are added for recognition to the junctions’ dictionary. These new elements comprise combinations of “L” or “I” junctions with an edge which can be located to either side, lower or higher, of the junction in a random position. In Fig. 11, a summary of illustrations of the additional dictionary elements is provided. Due to the complexity of the information, the junction recognition is not relying on simple calculations and the Random Forest classifier [23] is utilized so as to be trained to more complicated examples.

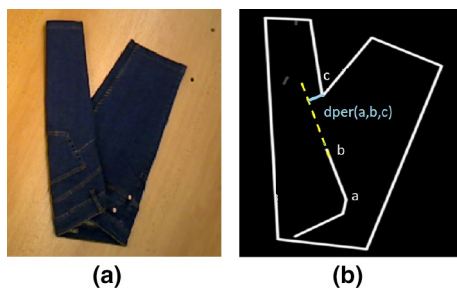


Fig. 10 Correlation of an edge in the interior of the garment with a corner located on the outline: **a** the colour image of the garment, **b** the extracted edges are depicted while the perpendicular of the point c to the line of the edge ab is marked with blue (colour figure online)

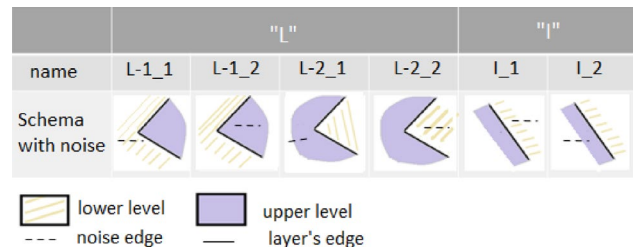


Fig. 11 Extra junctions explored by the junctions dictionary to deal with combinations with noise edges. The noise edge may be either to the upper or the lower level side without specific position or orientation

4.4 Junctions’ “dictionary” utilizing random forests methodology

The classification of the elements of the “dictionary” is a complicated problem that cannot be resolved through simple calculations. The Random Forest classifier (Fig. 12) provides the capability to learn by examples in a fast, parallelized manner while it produces probabilistic classification results that permit, in cases of multiple choices, to select the one with the highest probability. Moreover, according to [24], it is better suited for multi-class problems than classifiers such as SVM while it requires less training data and computational power than deep neural network approaches.

The goal of the proposed classifier is to classify three-edge junctions to elements of the “dictionary”. Therefore, its inputs refer to features regarding the schema and the depth differences of the junction under classification while the assigned classes are the junction types. In order to extract the inputs of a junction, one of its edges is considered as initial and the other two are taken into account based on the order of appearance in a clockwise manner (Fig. 13). Based on this order, the Euclidean distances and angles of interest between the first and the other two edges are extracted while the depths on both sides of each one of the edges are calculated. In addition, the information that an edge belongs to the outline of the folded garment and its side that belongs to the working table are declared.

For better understanding of these features, a schema of a junction that includes three edges ($\vec{e}_1 = \vec{ab}$, $\vec{e}_2 = \vec{cd}$ and $\vec{e}_3 = \vec{fg}$) is presented in Fig. 14 while equations for calculating the depths alongside an edge, the distance and angle of interest between two edges are provided in order to facilitate the definition of the classifier inputs.

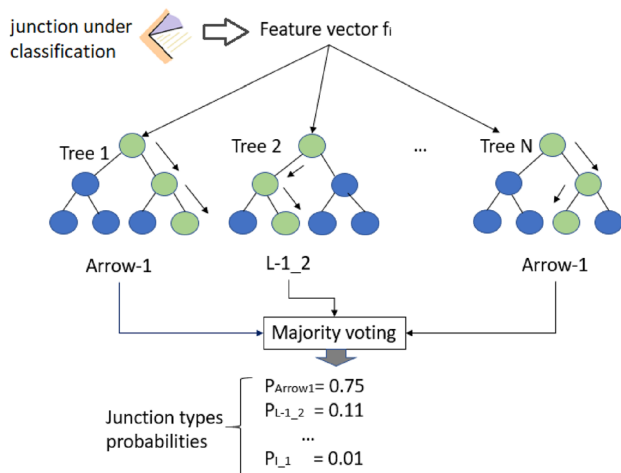


Fig. 12 Random Forest diagram. The feature vector f_i of a junction is explored by the trees of the Random Forest and, through majority voting, the probabilities of the junction for each class are calculated

Therefore, the distance between two oriented edges is calculated as in Eq. (2), meaning it is the Euclidean distance between the end of the first edge and the start of the second one. For example the distance between \vec{e}_1 and \vec{e}_2 is:

$$l_{12} = \text{dist}(b, c) \tag{4}$$

where $\text{dist}(b, c)$ the Euclidean distance between points b and c.

The angle of interest between two junction edges $\vec{e}_1 = \vec{ab}$ and $\vec{e}_2 = \vec{cd}$ is defined by the following equation:

$$a_{12} = \begin{cases} \widehat{abc}, & \text{if } l_{12} > l_{\min} \\ \widehat{abd}, & \text{otherwise} \end{cases} \tag{5}$$

where l_{\min} is a small predefined length that considers that the two edges are very close to each other.

For the extraction of depth differences between the two sides of an edge, two rectangular areas on the two sides defined by the edge are extracted (Fig. 14c). The rectangular areas length is the edge’s length l while their width w is predefined (10 pixels). The mean depth of each area is extracted while their difference defines the width of the garment along the edge.

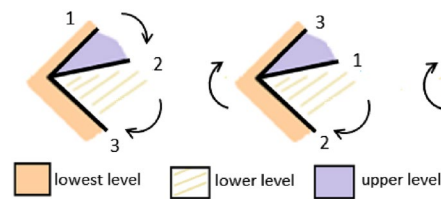


Fig. 13 The three different orientations of an “arrow-1” junction regarding the edges ordering

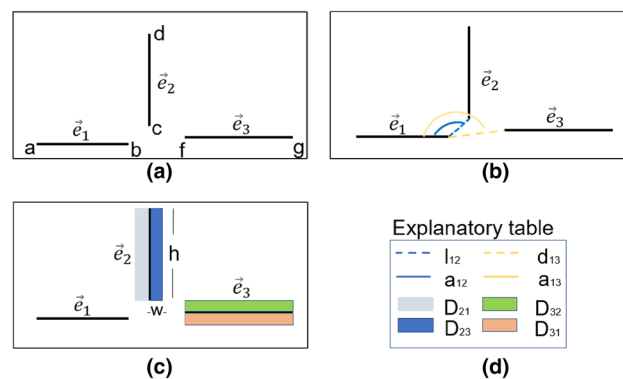


Fig. 14 Width extraction along a garment’s edge: **a** the edges of a junction, **b** the distances and angles between the edges of a junction, **c** the rectangular areas explored for the width extraction of an edge are highlighted with blue and yellow colours, **d** explanatory table (colour figure online)

Based on the aforementioned definitions, the inputs of the Random Forests classifier utilized for training and evaluation are the following:

1. The distance l_{12} between the initial and the second edge. Larger distances tend to have higher probabilities to refer to connections with noise-edges.
2. The distance l_{13} between the initial edge and the third edge.
3. The difference $dl_{23} = l_{12} - l_{13}$ between the two aforementioned distances. Values near zero indicate equally strong connections between the two candidate edges and the initial edge while large positive or negative values show the advantage of one edge over the other, i.e. higher probabilities to belong to the same layer with the initial edge.
4. The angle a_{12} between the initial and the second edge. Considering that two edges of a junction could belong to the outline of the same layer, their angle would signify the layer's change of curvature providing information on the junction's schema.
5. The angle a_{13} between the initial and the third edge.
6. The difference $da_{23} = a_{12} - a_{13}$ between the calculated angles.
7. The $dD_2 = D_{21} - D_{23}$ difference between the depths of the two sides of the second edge where D_{21} and D_{23} are considered in a clockwise manner. The depth differences provide valuable information that indicates the sides of the edges that most probably belong to the upper or the lower layer. Cases with small differences have higher changes to belong to noise edges.
8. The $dD_3 = D_{31} - D_{32}$ difference between the depths of the two sides of the third edge.
9. The difference $dD'_2 = D'_{21} - D'_{23}$ between depths D'_{21} , D'_{23} that are the same with the depths of the second edge unless for depths that do not belong to the garment, i.e. they belong to the table, and their depth is considered to be zero. This information shows that these edges are located at the outline of the half-folded garment and cannot be wrinkles.
10. The difference $dD'_3 = D'_{31} - D'_{32}$ of depths D'_{31} , D'_{32} , that refer to the depths of the third edge but similarly to (9) take into account areas that do not belong to the garment by setting them to zero.

Therefore, according to these features, for each junction j under classification, a feature vector f_j is created.

The outputs of the classifier are the probabilities of the input junction to belong to each one of the defined junction types of the dictionary. The class with the highest probability is considered to be the junction's type. Actually, the classes of interest are all the three-edge junction types as they are depicted in Figs. 4 and 11. Moreover, these classes

are divided into three subclasses according to the order of the edges at the feature vector f_j provided as input. In this way, apart from the class type, each edge of the junction is mapped on the edges of its type schema.

5 High-level feature extraction

In this section the extraction of high-level features of a garment's configuration are explored. The low-level features are combined with each other based on the semantics of the "dictionary" so that the axes and the layers of the garment can be detected through their components.

5.1 Axis recognition

In this section, the detection of the unfolding axis is presented. At first, only one axis is considered while the type of the garment or any information on the axis location is not available (e.g. we do not take into account that the fold occurred from predefined robotic manipulations). To facilitate its detection, an axis' definition based on Sect. 3.1, is utilized. Therefore, an axis can be defined as a sequence of junctions that fulfill the following rules:

1. The axis starts with a junction that is classified as an "arrow-1" or an "L" junction.
2. The axis ends with a junction that is classified as an "arrow-2" or an "L" junction.
3. In-between the predefined junctions, zero to multiple "T" junctions can occur.

The proposed axis detection methodology explores all the outline edges of the garment (Fig. 15b), combines internal edges with the corners of the outline to create junction candidates (Fig. 15c), classifies them (Fig. 15d) and, once they

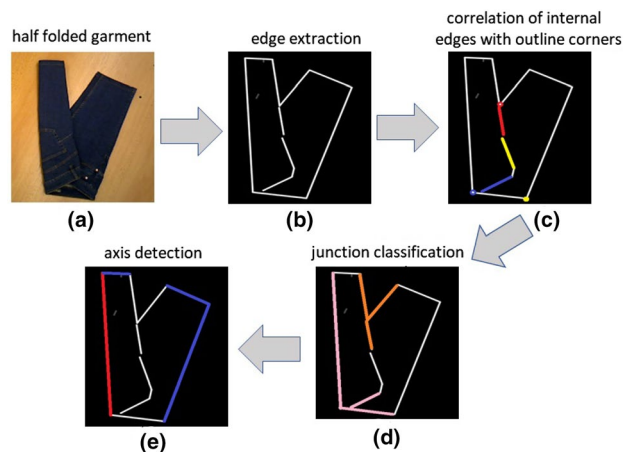


Fig. 15 The axis detection procedure

satisfy the aforementioned criteria, ends up through an SVM classifier to the final axis (Fig. 15e).

At the first step of the axis detection process, all the edges of the folded garment's outline are combined into junctions. The junctions might include either only two consecutive edges that form a corner or additionally include an internal edge that could refer to a part of an "arrow" or a "T" junction providing useful information on the garment's configuration. Nevertheless, it could be a noise edge that was combined with a simple "L" or "I" junction. Therefore, each outline corner might end up with zero, one or multiple internal edges. At this point, the "junction's dictionary" is applied and the candidate junctions are examined. In the cases that no internal edge is assigned to a corner then the junction is classified as an "L" or "I" while in cases of multiple edges, only the junction that has the highest recognition probability from the Random Forest classifier is considered as valid.

Once all the corners of the garment's outline are classified, axes that fulfil the aforementioned criteria are detected. Although these measures limit the possible cases, more than one axis candidates might occur. To recognize the real axis, an SVM with polynomial Kernel of 5th degree was utilized (Fig. 16). The SVM classifier is selected since it is suitable for binary problems without large amount of training data. The inputs to this SVM are enumerated in the following lines:

1. The number of "arrow" junctions participating to the axis. According to the rules mentioned above an axis can have N_{ar} "arrow" junctions, where $N_{ar} = \{0, 1, 2\}$. Since the "arrow" junctions are strong indicators of folds their existence supports the hypothesis that an edge with "arrow" junctions is an axis.
2. The relative width W_r of the garment at the axis candidate area. For its extraction, the width of each candidate edge is calculated from the depth difference between the side of edge where there is the garment and the side of the table (for the depth difference calculation see Sect. 4.4 and Fig. 14c). Then, all the widths are normalized based on the larger one calculating in this

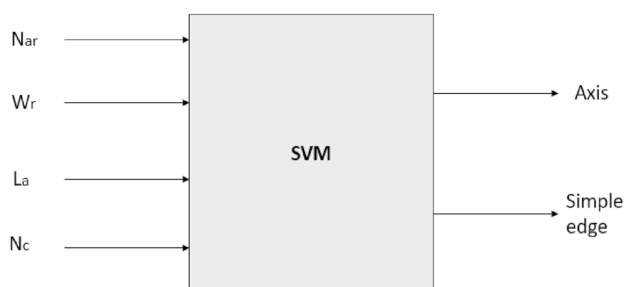


Fig. 16 The SVM classifier has four inputs while its output is binary with two classes: (1) axis, (2) simple edge

way the relative widths. Larger widths can be justified by a garment's fold; nevertheless, as it will be proved in the experiments, it is not always the case. There are cases where wrinkles or decorative features (e.g. pockets) amplify the width of other parts of the garments. Furthermore, in thin fabrics, the depth differences are close to the camera's tolerance as a result the calculated widths might have deviations that affect their ranking.

3. The length L_a of the axis candidate, which is defined by the Euclidean distance between its two ends. This parameter discourages the classifier from the selection of small edges, (e.g. the ones corresponding at the opening-end of sleeves).
4. The number of arrow junctions that belong to other candidate axes (N_c). Although an "arrow" junction indicates a fold, the case of false detections cannot be ignored. Therefore, the existence of more than one edges with "arrow" junctions could indicate more folds but also it could be the result of a misclassification.

Although the SVM does not use as direct input the actual pixels and their depth values around the candidate axis, the three-dimensional information of the area is taken into account in the form of the number of arrow junctions N_{ar} and the relative width W_r . In particular, if $N_{ar} > 0$ the formulation of two layers is implied. Moreover, the junction classification process that decides on the existence of "arrow" junctions uses the depth differences along the edges of the junctions as input. Finally, the relative width W_r , that comes from the calculation of the mean depth along each candidate, provides the ranked value in an effort to extract the thicker edge. This piece of information is more valuable than the absolute width of the edges since different garments can have different widths, i.e. different thicknesses.

For the evaluation of the axis detection methodology, two scenarios were explored: (1) the number of axes (N_{axis}) is considered a-priori known; (2) the number of axes is not known. In the first scenario, the results of the classification are ranked so that only the first N_{axis} candidates with the highest probabilities are considered as axes, while, in the second scenario, each candidate is evaluated separately. Since the results of a simple SVM are binary, a modified version is used in the first case in order to extract probabilities [25].

5.2 Upper layer extraction

Once the axis of the half-folded garment is known, the search of the upper layer can begin. The goal of this task is to find the sequence of edges that connects the two ends of the axis and forms the outline of the garment's upper layer. To determine this sequence of edges, a search algorithm is applied. The algorithm is defined by the tuple $\langle E, A, T \rangle$

that includes: (1) a state space E , that includes all detected edges apart from the axis; (2) an action space A that includes the possible interconnections between the states; (3) a state transition function T , that determines the next state of the sequence based on the states suggested by the action space A and (4) a search procedure, that deals with dead end sequences, meaning with sequences that do not end up connecting the two axis ends.

As it is already mentioned in Sect. 3, this algorithm introduces new features to the methodology of [2], that also uses a state space, an action space and a transition function. In particular the same state space is used in both cases, nevertheless this time a different action space and transition function are proposed. Therefore, the criteria of proximity, collinearity and continuity suggested for the calculation of the action space in [2] are substituted by the junctions' formulation criteria while the junctions' "dictionary" takes the place of a neural network transition function. More details on the steps of the proposed algorithm are provided in the following subsections.

5.2.1 State space

The state space E is defined as the set $E = \{s_1, s_2, \dots, s_N\}$, where N is the number of detected edges excluding the axis (to avoid taking this shortcut). The initial state s_1 and goal state s_G are defined by the junctions located at the ends of the axis. In the cases of "L" junctions, they are the edges that do not belong to the axis whereas in "arrow" junctions, the edges that signify the existence of the upper layer. In combinations that include one "arrow" and one "L" junction the exploration of the upper layer's edge sequence starts from the "arrow" junction while the goal state is the edge of the "L" junction.

5.2.2 Action space

The action space $A = \{a_1, a_2, \dots, a_n\}$, defines the n possible interconnections between the states as they are defined by Eqs. (2) and (3) that determine the criteria for the formulation of a junction. In this way, only certain actions are available at each state of the path. In Fig. 17, an example of the possible actions suggested by the action space is presented. Three different actions are allowed, i.e. three edges fulfil the criteria to form a junction with the edge corresponding to the state s_i .

5.2.3 State transition function

The state transition function $T(s_i, (a_{i1}, a_{i2}, \dots, a_{in})) = s_{(i+1)}$ is based on the junctions' dictionary. The function combines the current state edge s_i with the edges occurring from the action space A and, according to the formed junction, the

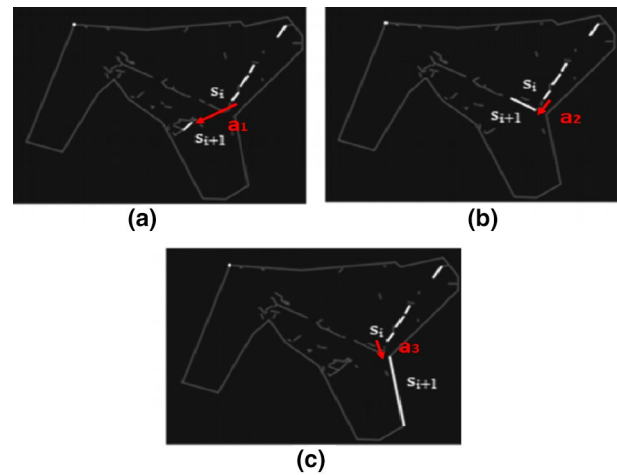


Fig. 17 Three different actions a_1, a_2, a_3 suggested by the action space in order to proceed from state s_i to the next candidate state s_{i+1} are depicted in the corresponding images

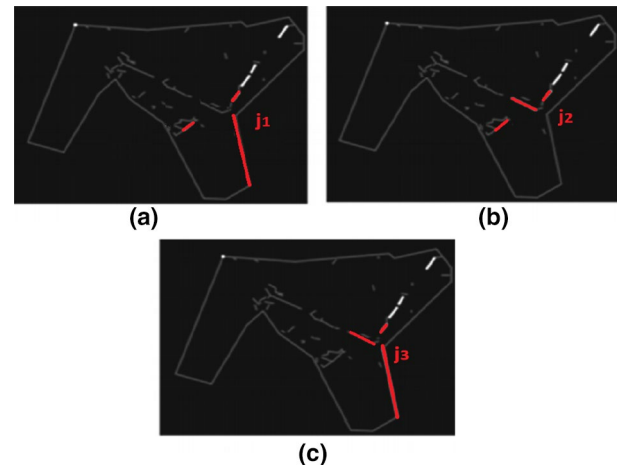


Fig. 18 Three different junctions j_1, j_2, j_3 based on the suggestions of the action space A and considered by the transition function T

edge $s_{(i+1)}$ that suggests the continuity of the upper layer is selected. In cases where the action space A suggests more than two edges (Fig. 17), all the possible junctions (Fig. 18) are taken into account and the one with the highest probability defines through its configuration the next state.

5.2.4 Search procedure

There are cases that, due to a false classification of the junctions' dictionary at the state transition function, the selected sequence of edges might not end up to the goal state or might lead to a loop intersecting itself, creating in this way a dead-end situation. To deal with these circumstances, a depth first algorithm is utilized. It can be observed that each time the

action space provides to the transition function more than one state, then a branch is created. The sequence of the upper layer edges incorporates the edge-state suggested by the transition function; nevertheless, the rest of the choices are kept as branches. Therefore, once a dead-end situation occurs, the algorithm returns back to the last node where there was a branch and re-evaluates the possible junctions. This time, only junctions that do not suggest the previous edge state as a part of the upper layer are taken into account. This procedure is iterated until a path of edges that leads to the goal state is found or until there are no other options.

Once the path of edges that signifies the upper layer is extracted, the edges are connected so that the complete outline of the upper layer is formed (Fig. 19).

5.3 Lower layer extraction

The outline edges of a garment constitute edges of its upper and lower layer. Once the upper layer is found the remaining outline edges can be mapped to the lower layer. In Fig. 20 the upper and the lower layer of a folded garment are depicted.

6 Experimental results

The experiments are divided in three main parts. In the first part, the performance of the junction dictionary is evaluated and, in particular, the performance of the Random Forest classifier. In the second part, experiments regarding the axis detection methodology are described. Finally, the layer extraction is evaluated in the third part.

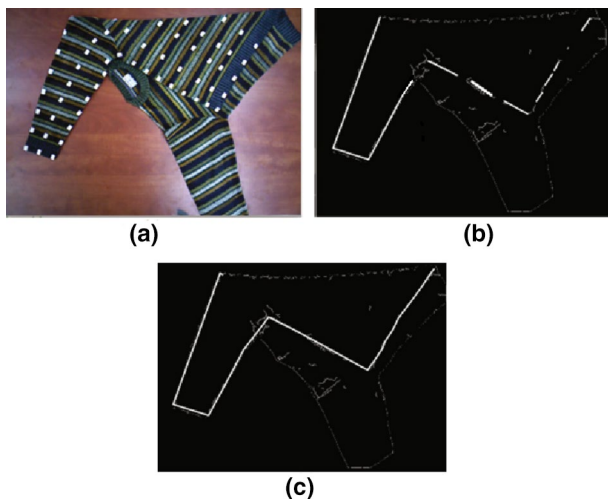


Fig. 19 Extraction of the whole outline that defines the upper layer: **a** colour image of folded garment, **b** the edges that define the outline of the upper layer, **c** the complete outline of the upper layer as it is defined by the union of its edges (colour figure online)

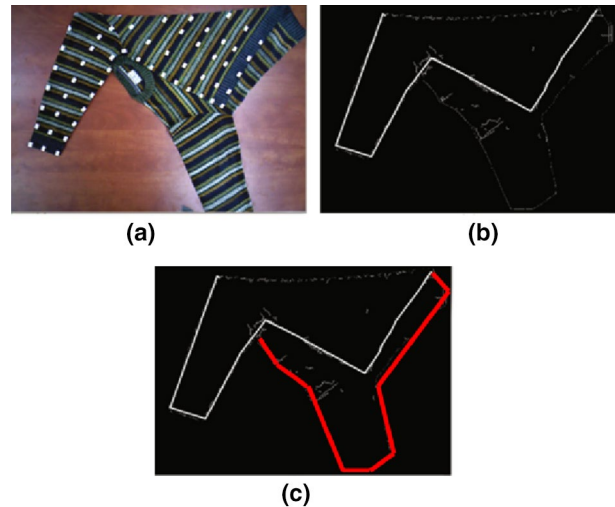


Fig. 20 Extraction of the garment's lower layer: **a** colour image of folded garment, **b** the extracted upper layer, **c** the lower layer showed in red (colour figure online)

For the evaluation of the proposed methodologies on one-fold examples, an online dataset [17] and a dataset that was collected specifically for this purpose, with garments occurring from the robotic manipulations as they are described in [1], are utilized. The datasets include 30 different garments (13 in [17] and 17 in our own dataset) which belong to various types, such as shirts, T-shirts, shorts, skirts and towels, while they are made of a variety of fabrics such as cotton, wool, polyester, denim or leather. Moreover, for the experiments on garments with two folds, another dataset including 30 configurations of garments of various types (skirt, shorts, trousers, towel, T-shirt) was collected while such examples did not exist in an online dataset.

For the dataset collection, the depth sensor is located so that it has a top view of the working table in a distance of 0.7–1.0 meters. During the garment selection there was not any restriction to the garments' width, nevertheless the usage of very dark coloured cloths was avoided since the depth sensor cannot operate correctly and provide data in these cases. In particular, the thinner width of the garments was less than 2 mm (3.5 mm at the seam lines) and it referred to a T-shirt while the thicker was a long sleeved shirt whose width was approximately 1.5 cm (these measurements refer to our dataset and were made by a meter to avoid errors by the depth sensor).

6.1 Junctions' dictionary evaluation

During the training of Random Forests 200 examples were used for each junction type that were located in 400 different garment configurations of 15 different garments (from our dataset with one fold configurations) of various types such as

skirts, T-shirts, long-sleeve shirts, shorts, towels etc. These garments have various widths and a variety of decorative features such as pockets, collars or thick seams. The target of Random Forest is to classify the junctions and their orientation (as it is defined by the clockwise order of appearance).

For the training of the Random Forests 160 trees were utilized with max depth equal to 20. The classifier was trained on 80% of the samples and tested on the remaining 20%. It managed to recognize correctly the 71.5% of the cases. Nevertheless, the biggest part of the misclassifications did not fail to recognize the edges of the junctions that belong to the same layer and the errors mainly referred to misinterpretations of the relative inclination between the edges (for example, an “L” junction combined with a noise edge could be mistaken for an “I” junction with a noise edge). In this scope, the classifier’s output provided correct information on the edges that belong to the same layer in 91.5% of the test cases. This feature is very useful since, in many cases, this piece of information can lead to correct layer recognition despite the misclassification of junctions participating to it. Moreover, it was observed that thicker garments provided better results while the performance was independent from the garment type.

During the development of the axis detection methodology, it was observed that the possible junction types that can be detected on the outline of the garment comprise a subset of the total set of junctions. In particular, in these cases, the lower layer of the junctions, i.e. the working table, is a priori known since they are located on the outline, hence, the possible orientations are limited. Taking advantage of this fact, the Random Forest classifier is trained on fewer classes. This time, the performance reaches up to 84.6%, which leads to more correct classifications than in the case that the whole set of junctions is taken into account. Furthermore, in 97% of the test cases, the classifier recognized correctly the edges that belong to the same layer. The results of the Random Forest classifier are summarized in Table 1 for both cases of junctions’ sets, the whole set and the subset, explored in this paper.

Table 1 Performance of the Random Forest classifier regarding the output class and the correct correspondence of edges to the upper or lower layer

	All junctions (%)	Outline junctions (%)
Classification performance	71.5	84.6
Layer correspondence performance	91.5	97

The whole set of junctions and the subset of the junctions at the outline of the garment are evaluated

6.2 Axis detection evaluation

For the evaluation of the axis detection methodology, 375 different configurations of the garments that belong to the aforementioned datasets (with garments with one fold) were formed. For the SVM training 60% of the configurations was used while the rest 40% was left for testing. Moreover, during the training, k -fold cross-validation (where $k = 5$) was used to achieve generalization and avoid over-fitting.

Various scenarios regarding the number of axes and whether this number was a-priori known were evaluated. A common step to all the scenarios was the application of the rules mentioned in Sect. 5.1 so that the number of candidate axes is diminished. In 95% of the tested configurations, the true axis was included in the candidates suggested by the aforementioned rules. In the rest 5%, the correct axis was excluded due to misclassifications of the “arrow” and “L” junctions that lay to its ends.

The first scenario that was tested considers only one axis. In this case, in the 91% of the cases the axis was successfully detected. In the second scenario, the number of axes is not known, hence each edge is evaluated independently of the others. This time, in 80% of the tested configurations all the edges were classified correctly to axes and simple edges. Regarding only the results of the SVM classifier, meaning the evaluation of the results per edge and not per configuration, the precision was 86.21%, the recall was 92.59% and F1-score 89.3% (Table 2). In all the scenarios, the existence of “arrow” junctions at the ends of the axes proved to be a valuable factor. Actually, cases with two “arrow” junctions were always recognized correctly in all the scenarios, while cases with two “L” junctions were proved to be the most prone to errors. Examples of successful axis detection are depicted in Fig. 21, while cases of failure are illustrated in Fig. 22.

The results presented in the previous paragraph referred to examples with one axis. To further evaluate the methodology, tests were also made to configurations with two folds. The number of the folds is considered known a priori. The methodology detected successfully at least one of the axes in 97% of the cases, while it detected successfully both axes in 87% of the cases. Examples of successful detection of both axes are presented in Fig. 23.

In relation to similar work [17], our proposed method presented better results regarding the correct axis detection per configuration for garments with one fold, while other scenarios were not explored. Actually, in their approach the axis is extracted correctly in 87% of the cases while, for the same scenario, our method reached the performance of 91%. A summary of the experimental results referring to all the axis detection scenarios that are explored in this paper and the results of [17] is presented in Table 3.



Fig. 21 Results of successful axis detection method in various garments. The colour image with the axis marked in red and the extracted edges from the depth image are provided for each example (colour figure online)

6.2.1 Upper layer extraction evaluation

For the evaluation of the upper layer detection methodology, 112 configurations of garments of various types (skirts, shorts, shirts and T-shirts) were utilized while the performance of the upper layer method reached the value of 83.04%. The reasons the proposed method failed in some

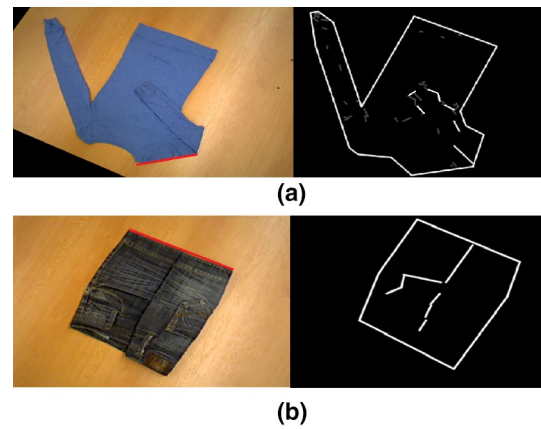


Fig. 22 Examples of incorrect axis detection: **a** error due to misclassification of the “arrow” junction formed by the fold, **b** error occurring from the SVM classifier. The axis suggested by the classifier is marked in red (colour figure online)

Table 2 The performance of the SVM classifier for the evaluation of the garment’s outline edges as axes or simple edges

	Precision (%)	Recall (%)	F1-score (%)
Successful axis detection per edge	86.21	92.59	89.3

Table 3 Evaluation of the performance of axis detection scenarios per configuration

	One axis (%)	Two axes/ one detec- tion (%)	Two axes/ two detec- tions (%)	Unknown (%)
Proposed approach	91	97	87	80
Stria et al. [17]	87	–	–	–

cases to extract the upper layer are either due to misclassifications of the junctions’ “dictionary” or due to missing edges that hampered the layer extraction procedure. In case of missing edges, that are more common in thin garments, the action space was not able to include the correct next state to the suggested actions, leading to failure. Moreover, highly wrinkled garments were more prone to errors since they usually lead to the examination of multiple junctions. Examples of the detected upper layers on half-folded garments are presented in Fig. 24 (online dataset) and Fig. 25 (our dataset). To accomplish better understanding, for each tested configuration three images are provided: (1) an image depicting the extracted edges and (2) a colour image of the garment, where the edges that comprise the upper layer are drawn with green colour, (3) the outline of the upper layer



Fig. 23 Results of successful axis detection in cases with two folds. The colour image with the axis marked in red and the extracted edges from the depth image are provided for each example (colour figure online)

as it is dictated by the extracted edges. Furthermore, cases of unsuccessful layer detection are depicted in Fig. 26.

For further evaluation of the upper layer detection algorithm, some extra experiments are performed utilizing garments with two folds. These tests were limited to cases that the whole areas of the layers were visible to an observer, i.e. the layers did not intersect. Based on this restriction 30 different configurations were evaluated, each one including two folds. Various garment types were utilized during the

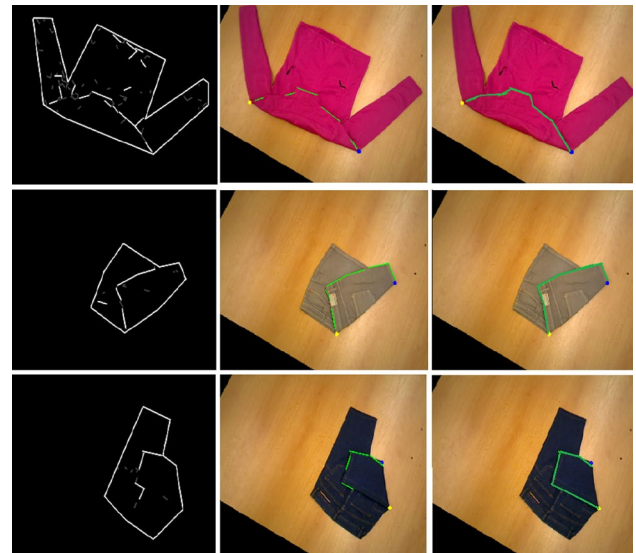


Fig. 24 Examples of successful upper layer detections of garments in a half-folded configuration from online dataset. For each garment three images are depicted. In the first one the detected edges are depicted, in the second one the edges that comprise the upper layer are presented in green while in the third one the outline of the upper layer is highlighted (colour figure online)

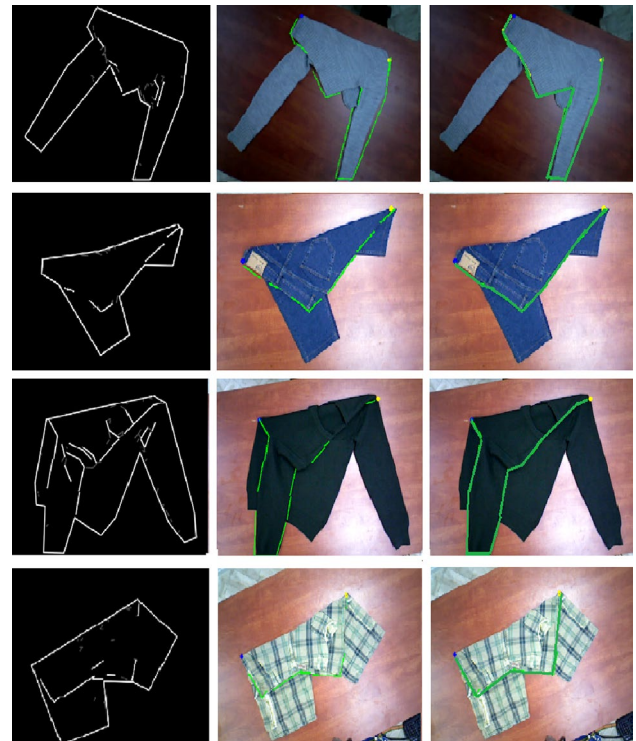


Fig. 25 Examples of successful upper layer detections of garments in a half-folded configuration from our dataset. For each garment three images are depicted. In the first one the detected edges are depicted, in the second one the edges that comprise the upper layer are presented in green while in the third one the outline of the upper layer is highlighted (colour figure online)

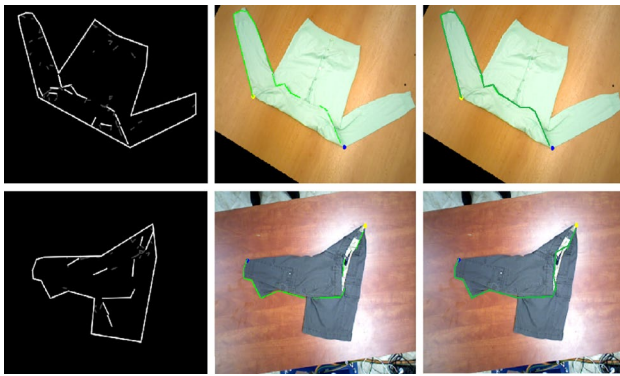


Fig. 26 Examples of unsuccessful layer detection. For each garment three images are depicted. In the first one the detected edges are depicted, in the second one the edges that comprise the upper layer are presented in green while in the third one the outline of the upper layer is highlighted (colour figure online)

tests such as skirts, shorts and T-shirts. In 90% of the cases at least one of the layers was detected correctly while in 80% of the cases both layers were recognized. Examples of successful detections of the layers in configurations with twofold are presented in Fig. 27.

Comparing to two related works [2, 17], our approach, that was successful in 83.04% of the cases, has better results than [2], whose performance is 79.4%, while it does not

have as good performance as [17] since their results reach the value of 89%. Nevertheless, the other two approaches do not extend to other scenarios with more than one folds. A summary of the performance of the upper layer detection method for all the explored cases is provided in Table 4.

Apart from being capable to generalize to more than one folds, the layer extraction approach is able to function in set ups that are not identical while it can handle garments of different type or fabric of the ones used for training. This mostly relies on the performance of the Random Forest classifier. Despite the fact that the classifier is trained on data from our own dataset for one fold examples, the method generalized to the other two datasets, meaning the online dataset and our dataset with configurations with two folds. A factor that facilitated it is the similar set ups between the cameras and the working tables, although they are not identical. Nevertheless, the main reason is that the classifier depends on depth differences and not on depth values and, therefore, is more adjustable to minor set up changes. Moreover, during the collection of the training dataset minor changes on the set up occurred making the system more flexible to changes. Furthermore, a major accomplishment of the layer extraction method is that it handled new types of fabrics, such as leather that is included in the online dataset, and new types of garments such as a jacket of the online dataset and a pair of trousers of our own two-fold configuration dataset.

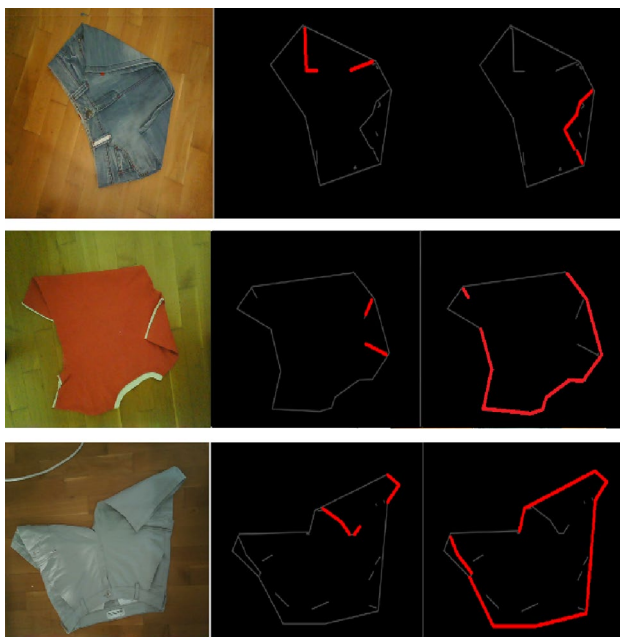


Fig. 27 Examples of successful detection of both layers in cases of configurations with two folds. For each case three images are depicted: (1) the colour image of the folded garment, (2) the image depicting the detected edges with the edges of the first detected layer marked in red, (3) the image depicting the detected edges with the edges of the second detected layer marked in red (colour figure online)

7 Conclusions

In this paper, a method for the extraction of the upper layer of a folded garment lied on a table by analysing and perceiving its configuration is proposed. The presented method is placed in a pipeline for the robotic unfolding of garments and constitutes a crucial part for its completion. In this scope, the garment is analysed into its conceptual parts, starting from primitive features, like junctions, and building up to more complex features, like the folding axis and the garment’s layers. In this scope, a “dictionary” translating junctions into indications of localized configurations is introduced while new methodologies for axis and layer detection are proposed. The method integrates human knowledge on the

Table 4 Evaluation of the performance of layer detection scenarios per configuration

	One layer (%)	Two layers/ one detection (%)	Two layers/two detections (%)
Proposed method	83.04	90	80
Stria et al. [17]	89	–	–
Triantafyllou et al. [2]	79.4	–	–

semantics of visual features that indicate the garment's configuration and uses machine learning approaches to classify them and combine them.

The proposed methodologies are based on a generic folded configuration of a garment; therefore, they are independent of its type. In this way, a variety of garments' shapes can be handled which, especially in women's wardrobe, can be large according to fashion. Moreover, tests showed that the proposed methodologies have the potential to generalize in cases with more than one folded layer.

Experiments proved the effectiveness of the method providing very good performance on different test sets and in a variety of garments, fabrics and configurations while tests showed that it can handle successfully new fabrics or garments' types. Thicker garments and garments with less wrinkles provided, as it was expected, better results.

In this direction, a camera with better error tolerance and a strategy that straightens the garment from wrinkles could be tested in an effort to enhance the performance. Future work dictates the completion of the unfolding task through robotic manipulations. The robot's aim will be to unfold the upper layer of the garment utilizing a strategy that takes into account the shape of the upper layer, the limitations of the robotic workspace and the working table while it will try to avoid the formulation of extra folds that could occur from the manipulations.

References

- Triantafyllou D, Mariolis I, Kargakos A, Malassiotis S, Aspragathos N (2016) A geometric approach to robotic unfolding of garments. *J Robotics Auton Syst* 75:233–243
- Triantafyllou D, Aspragathos N (2019) Upper layer extraction of a folded garment towards unfolding by a robot. In: RAAD pp 597–604
- Tu Z, Chen X (2014) Junction detection based on line segments. In: 9th IEEE conference on industrial electronics and applications, Hangzhou, pp 1231–1234
- Doumanoglou A, Kim T K, Zhao X, Malassiotis S (2014) Active random forests: an application to autonomous clothes unfolding. In: ECCV 2014: computer vision—ECCV, pp 644–658
- Doumanoglou A, Kargakos A, Kim TK, Malassiotis S (2014) Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning. In: IEEE international conference on robotics and automation pp 987–993
- Corona E, Alenya G, Gabas A, Torras C (2018) Active recognition and target grasping point detection using deep learning. *Pattern Recogn* 74:629–641
- Saxenaand K, Shibata T (2019) Garment recognition and grasping point detection for clothing assistance task using deep learning. In: IEEE/SICE international symposium on system integration, pp 14–16
- Mariolis I, Peleka G, Kargakos A, Malassiotis S (2015) Pose and category recognition of highly deformable objects using deep learning. In: International conference on advanced robotics, pp 655–662
- Maitin-Shepard J, Cusumano-Towner M, Lei J, Abbeel P (2010) Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In: IEEE international conference on robotics and automation, pp 2308–2315
- Yamazaki K (2015) A method of grasp point selection from an item of clothing using hem element relations. *Adv Robot* 29
- Kita Y, Sian Neo E, Ueshiba T, Kita N (2010) Clothes handling using visual recognition in cooperation with actions. In: IEEE/RSJ international conference on intelligent robots and systems, pp 2710–2715
- Li Y, Wang Y, Case M, Chang SF, Allen KP (2014) Real-time estimation of deformable objects using a volumetric approach. In: IROS, pp 987–993
- Osawa F, Seki H, Kamiya Y (2007) Unfolding of massive laundry and classification types by dual manipulator. *J Adv Comput Intell Intell Inform* 11:5
- Cusumano-Towner M, Singh A, Miller S, O'Brien JF, Abbeel P (2011) Bringing clothing into desired configurations with limited perception. In: 2011 IEEE international conference on robotics and automation, pp 3893–3900
- Hamajima K, Kakikura M (2000) Planning strategy for task of unfolding clothes (classification of cloths). *J Robot Mechatron* 12:577–584
- Manabu K, Massayoshi K (2003) Study on handling clothes-task planning of deformation for unfolding laundry. *J Robots Mechatron* 15:429–434
- Stria J, Petrik V, Hlavac V (2017) Model free approach to garments unfolding based on detection of folded layers. In: 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), pp 3274–3280
- Estevez D, Victores JD, Fernandez-Fernandez R et al (2020) Enabling garment-agnostic laundry tasks for a robot household companion. *Robotics Auton Syst* 123:134–139
- Triantafyllou D, Aspragathos N (2014) Definition and classification of primitives for the robotic unfolding of a piece of clothing. In: KEOD, vol 417, p 422
- Zhou P, Ye W, Wang Q (2011) An improved Canny algorithm for edge detection. *J Comput Inform Syst* 7(5):1516–1523
- Hershberger J, Snoeyink J (1994) An $O(n \log n)$ implementation of the Douglas–Peucker algorithm for line simplification. In: Proceedings of the 10th annual symposium on computational geometry, pp 383–384
- Khoshelham K (2012) Accuracy analysis of Kinect depth data. In: ISPRS—international archives of the photogrammetry, remote sensing and spatial information sciences
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Geron A (2019) Hands on machine learning. O'Reilly Media, Inc., Newton
- Platt J (1999) Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Adv Large Margin Classif* 10(3):61–74

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.