

# Perception of cloth in assistive robotic manipulation tasks

Pablo Jiménez · Carme Torras

Received: date / Accepted: date

**Abstract** Assistive robots need to be able to perform a large number of tasks that imply some type of cloth manipulation. These tasks include domestic chores such as laundry handling or bed-making, among others, as well as dressing assistance to disabled users. Due to the deformable nature of fabrics, this manipulation requires a strong perceptual feedback. Common perceptual skills that enable robots to complete their cloth manipulation tasks are reviewed here, mainly relying on vision, but also resorting to touch and force. The use of such basic skills is then examined in the context of the different cloth manipulation tasks, be them garment-only applications in the line of performing domestic chores, or involving physical contact with a human as in dressing assistance.

**Keywords** Robotic assistance · Robotic cloth manipulation · Perception of cloth

## 1 Introduction

Robots perform quite competently nowadays in structured environments, tackling hard tasks under tough working conditions, and even handling incidences that could be anticipated. The requirements posed by assistive settings, however, point in a quite different direction. The top concerns are no longer precision and repeatability, but rather a high degree of adaptability to varying ambient conditions, ability to learn, multimodal human-robot interaction capabilities, and integrated safety. It is obvious that robots cannot replace humans entirely in assistive environments, probably they shouldn't either. Nonetheless, it is desirable that they be able to perform a variety of tasks within domestic and service environments such as hospitals or care homes. Such tasks fall under what is commonly known as *domestic chores*, whose fulfillment ensures not just tidy homes, but the very proper life conditions of physically disabled people, in acceptable standards of human dignity and coverage of needs. Human companionship cannot be obviated, but the burden of associated duties without added value in personal interchange can be certainly alleviated with a robotic helper.

Among these tasks, those that involve the manipulation of textile items, including of course all types of garments, but also other categories of fabric-made objects such as bed- and tablecloth, curtains, towels, kitchen rags and dishcloth, etc. have to be highlighted. The omnipresence of such items in human daily environments and the importance of handling them correctly are evident. The means of providing assistive robots with the necessary abilities to perform cloth manipulation are not so clear. Ideally one would like to replicate the human proficiency in manipulating clothes, but the robotic state-of-the-art is still far from achieving the needed perceptual skills and the required dexterity. Cloth perception and manipulation

---

This work was developed in the context of the project CLOTHILDE (“CLOTH manIPulation Learning from DEMonstrations”), which has received funding from the European Research Council (ERC) under the European Union’s Horizon2020 research and innovation programme (Advanced Grant agreement No 741930). This work is supported by the Spanish State Research Agency through the María de Maeztu Seal of Excellence to IRI (MDM-2016-0656)

---

P. Jiménez  
Institut de Robòtica i Informàtica Industrial, CSIC - UPC  
Llorens i Artigas 4-6, 08028 Barcelona, Spain.  
E-mail: pjimenez@iri.upc.edu

C. Torras  
E-mail: torras@iri.upc.edu

are extremely challenging, due to the deformable nature of fabrics. Textiles are highly flexible, and slight mechanical actions are amplified to large motion or draping variations Magnenat-Thalmann and Volino (2005). As a consequence, even tiny changes in the initial configuration or in the conditions in which the action takes place may lead the manipulated garment towards very different states. Deformations include wrinkles, folds, and inside-out reversals. As deformations may occur anywhere along the piece of fabrics, the range of possible shapes that the same garment displays (intra-garment variability) is practically infinite, which exacerbates the difficulties in recognizing the cloth item. Moreover, the bending energy stored by such deformations is really negligible, which means that cloth does not tend to recover the state previous to such deformations by its own means. This intricate behaviour of fabrics is due to the anisotropy and the nonlinear mechanical response of this material, as well as to the fact that it has two privileged dimensions, whereas thickness is orders of magnitude smaller. However, these challenges do not mean that it is pointless to aim at robots tackling cloth manipulation tasks, only that their versatility, ability to deal with uncertainties, learning capabilities and dexterity need to be fostered by research. Among all the skills robots have to be furnished with in order to manipulate textiles successfully, perception ranks possibly in the first place, due precisely to the difficulties in predicting the response of cloth while being manipulated. Predictive models can be certainly devised and used, but a constant monitoring of what is actually happening is needed. Perceptual feedback may eventually trigger correcting actions, and in any case it provides the required instantiation of the manipulation action parameters. Moreover, in the absence of predefined suitable models, such feedback provides the needed input for learning, be it by imitation -observing human teachers perform actions such as folding a shirt or helping a patient to dress up- or through reinforcement -stating the success or failure of the robot own actions.

This survey is about perception for robotic cloth manipulation in an assistive context. Robots to be deployed in domestic or other assistive environments will not gain their perceptual abilities from scratch, but will already be equipped with a set of common or generic perception skills. Such perceptual primitives, to be tuned through interaction with users, may include such different things as human gesture recognition or natural language understanding. Here we focus on the basic skills required for versatile cloth manipulation. Identifying textile items in a complex scenario, determining at which points a garment or a towel has to be grasped (depending on what the robot plans to do with it), finding out what kind of textile object it is, or which state along -for example- a folding process has been reached so far, all belong to this set of basic skills -and the state-of-the-art in research is systematically revised. Next, we examine which role these perception skills play in the context of the different assistance tasks. As we address assistance in an integral way, we consider all the tasks that are developed within a domestic environment to guarantee the functional autonomy of its inhabitants. Two types of tasks may be distinguished: a) tasks that involve manipulation of cloth alone or interacting with other objects as for example tables, beds or hangers; and b) manipulation of cloth in contact and interacting with a human user, that is, a person with disabilities. This distinction shapes Sections 3 and 4 of this survey, respectively. Some final considerations on the research done so far and what has still to be achieved are provided in Section 5.

## 2 Generic perceptual skills in cloth manipulation

Some basic perceptual skills such as determining the point where the cloth item has to be grasped or the current state of a garment are common to the different assistive applications to be described below. Of course, each such application provides its particular specifications to these tasks, but some shared generic mechanisms can be identified. For this reason we leave out too specific perceptual skills like seam tracking for sewing, which hardly will arise in an assistive context. Perception involves mainly computer vision, but there are complementary perception channels such as force sensing or touch that supply added value to visual perception. In what follows, these fundamental perceptual tasks are described, stressing their relevance in the context of assistive duties, as well as the challenges they present and how they are tackled. Figure 1 shows the perceptual skills addressed in this section and their possible outcomes.

### 2.1 Cloth isolation and wrinkle detection

Objects made of fabrics as for example garments are generally in expected locations including the drawers of a wardrobe, the hangers in a dressing room, the washing machine/drier or adjacent laundry baskets, hanging on a clothesline, etc., but they can as well be hanging from the back of a chair, scattered over a sofa, or thrown on the ground. Thus, the very first thing an assistive robot has to be able to do is to identify cloth items as such. Just storing all the garments and other fabric parts in memory is not enough, as -due to their

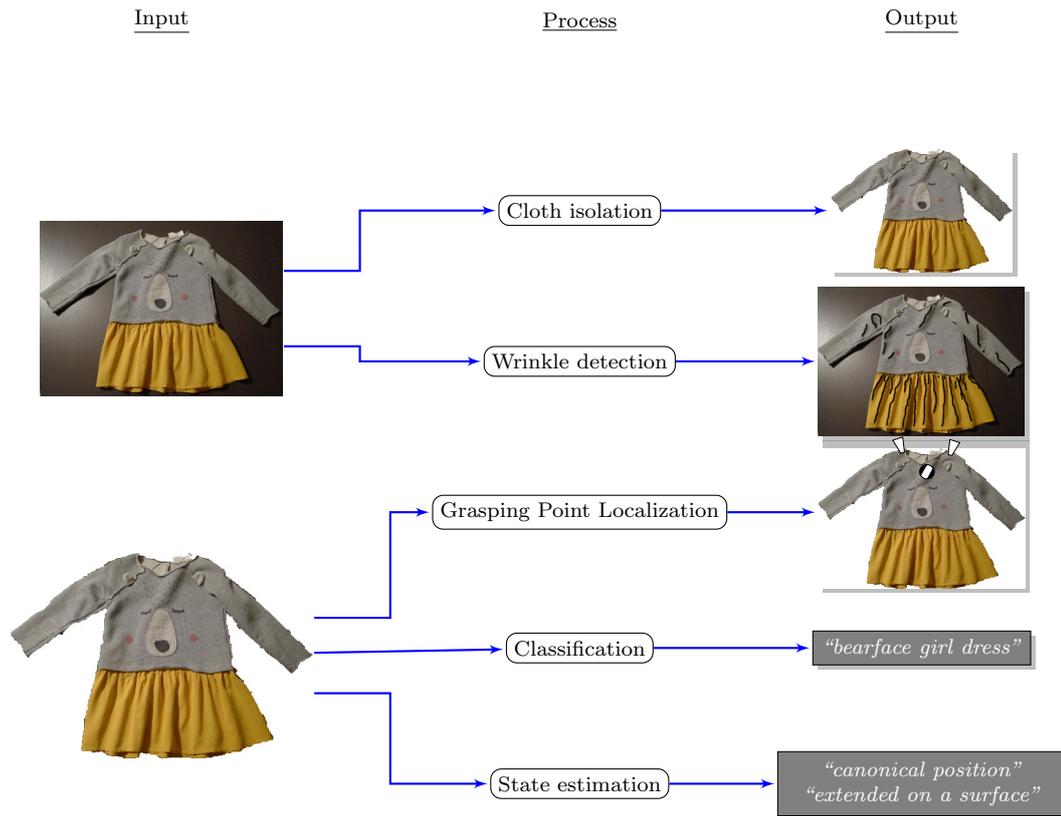


Fig. 1: Generic visual processes and the outputs they should provide. Here, **cloth isolation** is just background segmentation, as in many of the surveyed papers. One way to identify an item as textile in a scene is through **wrinkle detection**, but this basic procedure may also provide information to state estimation, for example in flattening processes. **Grasping point localization** refers either to find a generic graspable point on the cloth item, for instance a wrinkle (and additional information referring to the most suitable orientation of the gripper may also be given (circle)), or to determine the location of specific features, the shoulders in this case (triangles), which are more appropriate if the dress has to be folded. The output of **classification** is a label that designates a given cloth category, which can be as generic as “dress” or as specific as “bearface girl dress”. Last, **state estimation** provides a pose description, “extended on a surface”, which in turn may be used by the decision-making layer to compute the following manipulation action to perform, for example folding or flattening.

deformable nature- clothes may display infinite possible shapes, what we call *configurations*. Knowing the color, texture, pattern or the *canonical shape* (i.e., its configuration when lying flat and completely extended) of an object made of fabrics may be certainly helpful for identification, but there is one feature that reveals its textile nature without this previous knowledge, namely wrinkles. Excepting plastic bags (or crumpled paper, but this one shows sharper wrinkles), no other material in a house displays the characteristic fabric creases. For this reason, and also because sometimes such wrinkles have to be removed, be it while ironing laundry or while flattening a bedsheet, wrinkle detection and measuring is the first skill that the vision system of a domestic or assistive robot has to master. The most popular computer vision technique to detect wrinkles uses Gabor filters, which register abrupt intensity variations at different orientations. Wrinkles are discriminated from other features that respond to such filters by tuning their frequency (wrinkles are low-frequency features as opposed to edges or borders), as well as using the fact that the wrinkles usually appear in groups with similar orientations Yamazaki and Inaba (2009). Together with pure 2D image processing procedures such as the already mentioned Gabor filters, combining threshold function filtering with edge enhancement Paraschidis et al. (1995), or Wavelet Transform techniques Sun et al. (2011), the popularization of depth cameras has encouraged researchers to obtain directly volumetric information about the wrinkles, which is often much more informative. For example, Sun et al. (2015) use a high-resolution stereo-based sensor to capture a 2.5D depth map, from which the height and width of the wrinkles are computed by geometric means. In this work, wrinkle descriptions are used to guide a flattening-by-pulling process, in particular the main direction of the largest wrinkles. This work is continued in Sun et al. (2016a); Sun et al. (2018), where a hierarchical visual architecture is described: low-level features consisting in surface curvatures are computed from the b-spline surface fitted to the raw data, from these Shape Index features

are derived as mid-level features (they capture the local topology of the surface, i.e., whether it is a *cup*, a *trough*, a *saddle*, a *ridge*, a *dome*, etc. up to nine features), which finally allow to detect and quantify wrinkles as high-level features (fifth-order polynomials fitted to the ridges, while ruts and domes are used for splitting wrinkles).

## 2.2 Grasping point localization

The success or failure of any manipulative action on cloth depends largely on the adequacy of the grasps. In the well-studied case of rigid objects, the suitability of grasps is confirmed by form and force closure tests. Moreover, if different alternatives exist as for where to grasp, the choice responds to accessibility and stability criteria, not only while holding the part but also with regards to the final positioning of the part. When the objects to be manipulated are deformable, and in particular, if they are made of cloth, then the most suitable localization of grasping points depends largely on the manipulation goal, which generally involves some sort of deformation from the current state. For example, if we want to fold a shirt in the air, the most appropriate grasping points will be located at the shoulders, whereas if the very same shirt has just to be separated from a laundry heap, the localization of the grasping point will not be so relevant, as long as a stable grasp is guaranteed. Next, we explore these two alternatives (generic vs. specific grasping point) in more detail.

### 2.2.1 Generic grasping point selection (GGPS)

When the location of the grasping point is not relevant, the concern is just to ensure a stable grasp. This is the setting we may find for example in a laundry sorting scenario, where the clothes in a pile have just to be picked up. The exact location of the grasping point is unimportant here, what matters is that the garment be effectively grasped. Sometimes finding such a grasping point is just the first step within an active vision strategy where more specific grasping localizations are searched once the cloth item has been picked up. Such strategies often have an added benefit, namely solving automatically the problem of grasping more than one garment by accident: the regrasp by a second hand ensures that only one garment is held afterwards, and the other one, unless tightly entangled with the first, falls back to the pile Maitin-Shepard et al. (2010).

Earlier works required the use of at least a couple of cameras, either combining one zenithal (plan or top view) camera capturing the scene and another lateral one measuring the height of candidate points at wrinkles Hamajima and Kakikura (1996, 2000), or capturing the depth by stereovision (the two cameras are arranged side-by-side) Maitin-Shepard et al. (2010); Kita et al. (2011); Bersch et al. (2011); Willimon et al. (2011a). The first step consists in isolating within the image the part corresponding to the textiles, or, in other words, eliminating the background, which –depending on the particular layout– can be done by a simple method as for instance binarization Hamajima and Kakikura (1996, 2000). To this end, or to distinguish between different clothes in the pile, often some sort of image segmentation into regions is performed. Such segmentation may be a simple histogram-based procedure, or a more sophisticated method that takes the varying appearance of the parts of the same cloth piece into account Willimon et al. (2011a). The largest region is then chosen for grasping: for clothes of similar size, this can be an indication of being at higher positions on the pile. The selection of the grasping point within such region can proceed in different ways: just aim at the highest point Bersch et al. (2011), or take the centroid Hamajima and Kakikura (2000); Maitin-Shepard et al. (2010) (alternatively the farthest interior point to the border Willimon et al. (2011a)) of the region and measure its height before grasping. Figure 2 shows how the centroid is not always the best candidate for grasping, and for non-convex shapes some interior point has to be chosen instead.

Wrinkles are good candidates for grasping: the fingers of the gripper are positioned at both sides of the crease and a successful pinch grasp follows (see Figure 2). Thus, wrinkle-centered approaches are suitable alternatives to the region-oriented approaches above. This was already the perspective taken in Hamajima and Kakikura (1996), via 2D image analysis technique plus height measuring with lateral cameras. The advent of range cameras, as mentioned at the end of Section 2.1, paves the way to such approaches even more, by providing directly a full 3D description of the wrinkles. See for example Ramisa et al. (2011), which includes an experimental study of different grasping strategies. Also Sun et al. (2018) use their hierarchical vision architecture (recall Section 2.1) to compute, from the shape index and surface topologies, adequate grasping-triples (one ridge point and two wrinkle contour points on either side of the wrinkle).

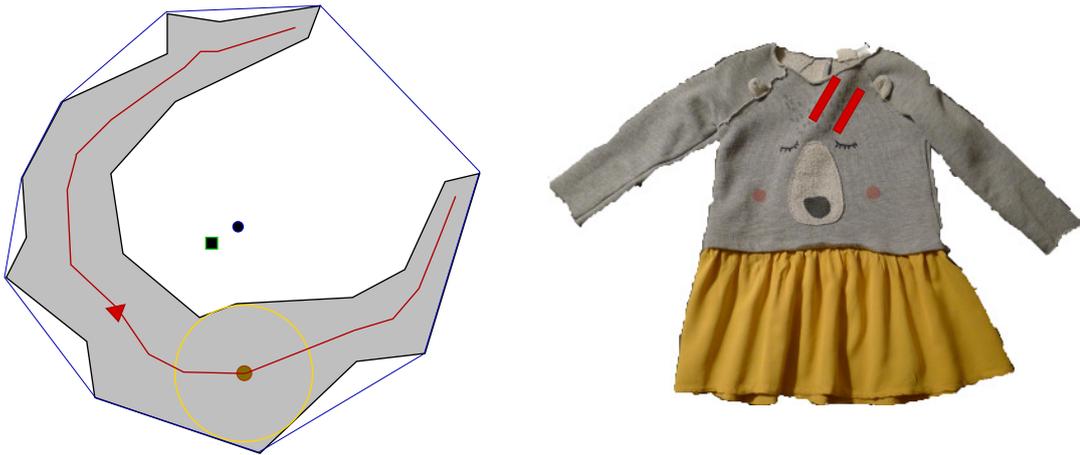


Fig. 2: For non-convex shapes of the cloth item, neither the centroid (small square) nor the center of the convex hull (blue dot) may provide suitable grasping points. Instead, the farthest interior point to the border (the center of the yellow circle), or some point -for example the middle point- along the shape’s skeleton (here we have depicted just the central “spine”, in red) are good candidates for grasping. On the right, an outstanding wrinkle provides a good potential grasping point. The orientation of the gripper is determined by the direction of the main axis (the ridge) of the wrinkle.

### 2.2.2 Specific grasping points selection (SGPS)

Successful manipulation often requires the garment to be grasped at specific points. Such is the case of folding a shirt in the air, to name an example, which has to be grasped at the shoulders before rotating it backwards and making it lay on the table. Usually such points are recognizable cloth features (in the sense that they can be identified by vision), which can be classified into three main groups:

- **Generic contour features**, that is, edges and corners.
- **Garment features**, i.e. collars, cuffs, pockets, and the like.
- **Node points of a cloth model**.

These three approaches are illustrated in Figure 3.

Regarding the first group, typical applications include detecting corners for unfolding cloth items laying on a surface, and active sensing strategies to bring the cloth item into a desired configuration while hanging in the air:

- **Unfolding on a surface**. Here we can cite the work Ono et al. (1998) where, after a foreground-background segmentation and encoding the outline information, the obtained sequence of edges and corners is matched against the stored contour of the cloth item in the canonical configuration. By comparison of the two contours, the potential location of folded corners can be estimated, thus guiding the robot towards the grasp location for unfolding. In Triantafyllou and Aspragathos (2011) the location of corners is computed via the Canny edge detector from local maxima of contour curvature. The edge opposite to the corner constitutes the corresponding “unfolding axis” around which the unfolding motion has to be performed. This work includes also a taxonomy of possible corner occurrences, with real corners as well as with pseudocorners arising from folds. Moreover, performance of their method is enhanced by exploiting varying illumination conditions. A third quite interesting work is presented in Willimon et al. (2011b), who proceed in two phases: a first maximal edge-detection phase on the binarized image of the garment along eight discrete directions, which are chosen for grasping and pulling (this already eliminates some folds), and a second corner-detection phase that combines depth information with the Harris corner detector to locate corners for unfolding.
- **Regrasping strategies of hanging garments**. The cloth item is assumed to be already grasped and hanging from the robot’s gripper. There is a high probability that the lowermost point corresponds to a corner of the garment, as observed in Osawa et al. (2007). This reference is a good example of active vision strategies, as it combines shaking for disentangling with rotation to obtain the view where the lowermost point displays maximum curvature. Repeated grasping of the lowermost point (alternating the two hands of a bimanual robot) progressively reduces uncertainty until the cloth item can be classified. The lowermost point is also considered in Hamajima and Kakikura (2000), combined with two hemline detection procedures. More recent research aims at corner detection by using 3D information. Two

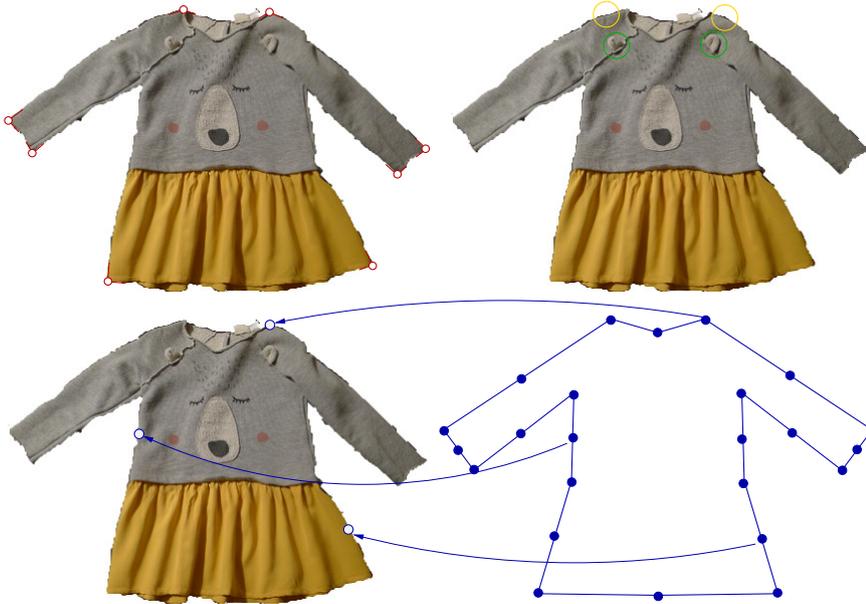


Fig. 3: Top-left: corners are easily detected generic features that provide good grasping points for applications similar to folding or unfolding on a surface. Only convex vertices with sufficiently long adjacent edges are considered. Top-right: User-defined grasping areas such as the shoulders (yellow circles) or garment features (green circles) belong to the second type of specific grasping points. Bottom: the correspondence between the model and the images allows the system to consider the nodes of the former as possible grasping candidates.

consecutive corners in the outline of a towel are searched in Maitin-Shepard et al. (2010) in order to perform a fold afterwards. Corners are detected combining a stereo multi-view border classification algorithm (which distinguishes actual towel borders from folds) with a RANSAC algorithm. At each iteration the towel is shaken and then rotated in front of the cameras at small discrete steps to find the most suitable grasp corner. Depth is also used in Triantafyllou et al. (2016) by computing edges from a combination of Canny’s edge detector, the Douglas-Peucker algorithm (separating edges of different orientations) and some geometric criteria. The outcome is a set of outline points, that once grasped brings the garment into a half-folded state, which is then matched against a polygon model of the folded garment. In this model, grasping point pairs are defined to be regrasped in order to complete the pre-established unfolding actions. Still within this category we could mention the seam-tracking method in Salleh et al. (2006, 2008). Although seam tracking and corner detection are based on a special gripper design with an integrated fototransistor-sensor pattern, an external 8-bit grayscale CCD camera confirms corner detection and provides 3D coordinates of the corner for grasping by a second hand (the latter by using motion stereo).

As for the second group, garment features cannot be identified by simple geometric cues (remember that they are also subject to deformations), and researchers have resorted to learning instead. Under this assumption, we may distinguish three different approaches:

- **Learned generic user-defined grasping features.** In Gibbons et al. (2009) no specific features are addressed, but the authors try to identify possible grasping areas from samples that human users considered graspable. These include but are not limited to edges, wrinkles and folds. A cross-correlation approach of appearance-based filters is followed.
- **Learned specific garment features.** In contrast, Ramisa et al. (2012) explicitly aims at polo shirt collars as suitable grasping candidates, which are identified by means of a Bag of Features (BoF, aka Bag of Visual Words, BoVW) detector combining appearance and 3D geometry. Logistic regression is used for a first rough classification of the images corresponding to a sliding window along the original image of the polo shirt, and then a more costly but also more reliable  $\chi^2$  Support Vector Machine (SVM) classifier is applied for further ranking of the candidate windows. Finally, among the best candidates, grasping points are selected by using a wrinkledness measure developed earlier Ramisa et al. (2011). The set of garment features to be detected (although not explicitly for grasping purposes) was later expanded to 11, including collars, sleeves, hemlines or hips of different garment types (jeans, polo shirts, T-shirts, shirts and sweaters) Ramisa et al. (2014). The approach is quite similar: a multitude of appearance- and

depth-based detectors are grouped together in a BoVW and again logistic regression is combined with local  $\chi^2$  SVM for classification. The difference with respect to earlier work lies in the particular detectors used: Scale-Invariant Feature Transform (SIFT) features for appearance, and Geodesic-Depth Histogram (GDH), Fast Point Feature Histogram (FPFH), Heat Kernel Signature (HKS) and Fast Integral Normal 3D (FINDDD) descriptors for depth.

- **Learned specific grasping features within an active vision strategy.** With an alternative setting (clothes hanging from the gripper instead of lying on a table) and also different descriptors, the approach in Doumanoglou et al. (2014a) looks for shoulders of shirts and T-shirts or the corners of the waist of trousers and shorts as possible grasping points. An active vision strategy with regrasping, guided by a Partially Observable Markov Decision Process (POMDP) that encodes the hanging states at the predefined set of grasping points and the transitions between states, is followed, and the used descriptors are Hough forests (similar to random trees but with additional information on the location of grasping points).

Finally, with respect to the third group, the nodes of an existing mesh model of the garment (or garment type), or a subset of such nodes, provide the grasping point candidates. Often it is one such pair of nodes (for bimanual grasps) that is predefined as the target locations for grasping. These nodes may not necessarily coincide with cloth features similar to the ones described above. In any case, an accurate correspondence between the 3D mesh model and the current image of the garment has to be established. As in the corners-edges case, contributions can be grouped into settings where the garment lies on a surface and others where it is hanging:

- **Garment on a surface.** The method presented in Miller et al. (2012) consists in fitting their *skeletal model* (a skeletal mesh with additional parameterized landmark nodes) to the contour of the garment in the image, relocating the landmark points of the model to the nearest neighbor on the contour. The procedure is applied to the different folded states that have been predetermined in a cloth folding application. Another approach Cusumano-Towner et al. (2011) begins with a *disambiguation* phase in a hanging state (an active sensing classification and state estimation method described below, with regrasping of the lowermost point), but in a second *reconfiguration* phase the cloth item is repeatedly laid on a table and regrasped. Contour matching via a dynamic time warp algorithm allows the robot to determine the location of nodes and to select a pair of points to grasp, lift, and lay again the garment following a path in the previously constructed *graspability graph* that corresponds to the desired sequence of actions.
- **Hanging garment.** In their first works, Kita et al. (2004) combined separated 2D visual information for matching right and left images to reconstruct via correspondence the 3D location of grasping point candidates. Later, Kita et al. (2009b, 2010, 2011) performed directly the matching of 3D observations with the 3D deformable triangular mesh models with their new trinocular stereo vision system, with an additional computation of the necessary orientation of the gripper for successful grasping Kita et al. (2009b). In Li et al. (2015a) a regrasping strategy is defined on the nodes of the garment’s model. At execution, the garment is picked up and its pose estimated, afterwards the regrasping point is found by optimizing an objective function measuring the closeness to the desired grasping points. The correspondences between the model mesh and the target mesh obtained from 3D scans of the grasped garment are computed via scaling, rigid transformation and non-rigid registration. 1-2 regrasps are generally enough for unfolding and holding the garment at the desired point. Finally, the work in Bersch et al. (2011) is worth mentioning despite their use of coded fiducial markers to short-cut the costly state recognition process, for their regrasping strategy. First, the held point at which the garment is hanging is quickly identified as the one minimizing the difference between the geodesic distance on the cloth surface and the Euclidean distance to the other points (corresponding to the markers). Then, the next grasping point is the one that brings the next grasp closer to the target configuration. The gripper orientation is determined through a collision detection procedure of the gripper’s fingers and the point cloud next to the candidate grasping point. Their approach requires both active sensing (including multiple rotations in front of the fixed camera and vertical motions of the camera mounted on the free arm of the bimanual robot) and a previous training of a score function that evaluates the probability of success of the grasp depending on the local point-cloud features.

### 2.3 Classification and state estimation

As we have just seen, some grasping point detection algorithms require some sort of state estimation (and thus, many of the works referenced above appear again in this section). But classification (that is,

identifying to which category a given garment belongs) and state (or pose) estimation are computer vision procedures that are required in any type of meaningful cloth handling processes. Most contributions perform classification and state estimation in a simultaneous fashion, but there are also some references that assume the cloth type already known beforehand Kita et al. (2004, 2009a,b, 2010, 2011), or that just perform classification as required by the application (like laundry sorting Willimon et al. (2011a); Yamazaki and Inaba (2013); Willimon et al. (2013)) or because the garment is assumed to be in a canonical state Miller et al. (2011). The following structure of this section is based on whether the whole shape of the garment or just specific features are considered in these processes.

### 2.3.1 Shape recognition methods

The shape of a cloth item in an image is compared either with images of the cloth model Kaneko and Kakikura (2001); Kita et al. (2004, 2009a,b, 2010, 2011); Cusumano-Towner et al. (2011) or with images of real clothes Osawa et al. (2007); Willimon et al. (2011a); Miller et al. (2011); Wang et al. (2011); Dumanoglou et al. (2014a); Stria et al. (2014a,b) in a database. We may categorize these contributions as follows:

- **Geometric features.** The point in these approaches is to compare geometric magnitudes that are easily obtained from the image. The convex closure of the binarized image of a garment held and stretched at two distant points allows Kaneko and Kakikura (2001) to define characteristic measures of the different regions appearing inside (both occupied and unoccupied by the cloth) that are used in a decision tree to determine both the cloth type and its hung-up state (only a finite set of discrete states are considered). Such characteristic measures have been previously obtained on cloth models. Instead, real stored images of clothes picked up at random points are used in Willimon et al. (2011a) for comparison with test images of garments also repeatedly grasped at random locations (two views rotated at  $\frac{\pi}{2}$  are taken). Matching is performed via a nearest-neighbor algorithm, with features such as the absolute difference in area of the two silhouettes, the absolute difference in eccentricity, the Hausdorff distance between their edges, and the Hausdorff distance between the Canny edges of the original grayscale images. Geometric features for garment classification are also applied for recognizing loosely extended garments on a surface in Hou et al. (2017); Hou and Sahari (2019). A set of measures obtained from the axis-aligned contour of the garment characterizes the garment type (not the absolute value of such measures but the relationships between them). These measures consist just in the contour width and height at regular intervals (e.g., a towel has always the same width and the same height).
- **Planar contour.** Different contour matching algorithms have been considered in literature:
  - *Region overlap*, i.e. the coincidence of the cloth region in the input image with the planar projection of the cloth model. This procedure is applied in Kita and Kita (2002); Kita et al. (2004), where one-hand held garment models are projected on a vertical plane, as corresponds to the real setting. These projected model images are transformed (vertical displacement and width normalization) to improve matching quality and the overlap region ratio to both the total cloth model region and real image region is considered. The cloth category is assumed to be already known and only state estimation is done (that is, determining at which node of the corresponding model the garment is hanging from). In some cases this simple criterion performs poorly. In the context of a folding application on a table, region overlap is also used in Colomé and Torras (2018) as a reward function during a reinforcement learning process. Folded shapes are assumed to be rectangular and thus the occupancy of the segmented cloth region wrt the enclosing rectangle is evaluated.
  - The average *Nearest neighbor distance* between the contour generated from the model parameters and the contour of the image is a measure of contour fit used within the energy function of an optimization algorithm in Miller et al. (2011). The real cloth image belongs to the category that displays the lowest energy (the state is known beforehand: garments are loosely spread out on a table).
  - A more accurate contour matching algorithm as for example *Dynamic time warping (DTW)* is justified in Cusumano-Towner et al. (2011) by the fact that they determine both the cloth type and its hung-up state (i.e., the pair of nodes of the corresponding mesh model at which it is grasped). The DTW algorithm provides a similarity measure for the two contours, considering pixel coordinates as well as first and second derivatives with respect to arc lengths (the derivatives correspond to corners and the like). A *simplified* polygonal contour of the test image and the polygonal models of the garments spread-out on the table is computed in Stria et al. (2014a,b) previously to the application of the DTW algorithm, which uses the vertices of the contours (their inner angles) as the chosen feature for similarity measure (in Stria et al. (2014b) also relative segment lengths are taken into account).

- **3D shape.** These methods rely on direct comparison of spatial information gathered through stereo vision or depth cameras with the 3D models of garments or volumetric information of real training images:
  - **3D models.** As in their previous work, Kita et al. (2009a,b, 2010, 2011) assume the cloth type is known beforehand. The 3D triangular meshes used as models are allowed to deform to some extent gradually from the fixed grasped patches tending towards matching the 12, under a force model that drives and bounds such deformations. As in their contour matching work, overlap criteria between projected regions of model and observed data are used to decide the current state. In contrast, Li et al. (2014a,b, 2015a) perform the two tasks of classification and state estimation, and they do it directly with the 3D models, via SIFT features and a SVM classifier Li et al. (2014a), or binary cell occupancy vectors with a nearest neighbor procedure based on Hamming distance Li et al. (2014b, 2015a). Test images are gathered with a Kinect sensor.
  - **Real training images.** Random Forests Doumanoglou et al. (2014a) and Active Random Forests Doumanoglou et al. (2014b) are the chosen classifiers that provide to an active sensing POMDP strategy the probabilities to belong to a given class and a given state, after the garment has been regrasped at the lowermost point. These classifiers have been previously trained on depth images of the set of garments and lowermost point regrasps. Instead, Sun et al. (2016b) implement classification as a Gaussian Process, on depth data obtained with the same cameras as for wrinkle computation Sun et al. (2015) (recall Section 2.1). The features used include the Shape Index histogram, Topology Spatial Distance and Multi-Scale Local Binary Patterns (LBP).
  - **Both synthetic and real training images.** It is a well-known fact that deep convolutional neural networks (CNN) require large amounts of training data. Thus, the two-layered CNN architecture developed in Mariolis et al. (2015) (the first one corresponding to classification, and the second, much larger, to state estimation, that is, determining the node from which the garment is hanging) has been first designed using synthetic data<sup>1</sup>. Once the most suitable layout has been found, the CNNs are retrained with real RGB-D data gathered with an Xtion camera. In a later work Kampouris et al. (2016), CNNs are combined with Random Forests on Histogram of Oriented Gradients (HOG), in a setting including not only vision but also photometric and tactile sensing, to obtain garment type, fabric pattern and material recognition. In Corona et al. (2018) the CNN (consisting of four non-square sized convolutional layers, plus four max-pooling layers between the convolutional ones, and two fully connected layers) is also first trained with the huge synthetic dataset and then refined with the smaller real depth images set. Only classification (no state estimation) is performed, and to this end the physical engine simulates the garment hanging at points all over its surface, and each view is captured by 36 virtual depth cameras. They also resort to deep learning to localize the predefined grasping points, now using only synthetic images. This avoids costly manual annotation: the predefined grasping points, with known location in the model, are projected on the image plane. After the first point is grasped, the set of possible cloth poses follows a normal distribution centered at the correct grasped pose. The networks also inform whether the grasping points are visible or not in the image. A similar approach is undertaken in Saxena and Shibata (2019) for garment classification and grasp point detection, as a first step in a dressing assistance context: the output of this step is the robot grasping the garment at two points of the hem- or waistline and presenting it to the patient (as the user is face-to-face to the robot, in the case of shirts the garment back is facing the robot, but not with trousers, otherwise the grasping points have to be swapped). Again the garment classes are learned via deep learning, on a real image dataset, and the trained convolutional networks used later for classification (three outputs in this case). Depth images are used so that the system is robust to colour or textile pattern variations. Also the grasping pairs of points are learned via convolution networks, now from two synthetic depth image datasets, the first one (SGD-I) following a similar strategy to Corona et al. (2018) (i.e., uniformly scattered grasping points on the garment), and the second one (SGD-II) consisting in points on distinctive features such as sleeve openings, hemlines, collars or waistlines (together with the two desired grasping points). Furthermore, whereas previous works required images of the whole garments, here most of the time only partially visible images are available, but despite this fact good classification results are obtained. The classification network (real image as input, one of the three considered garment types as output) as well as each one of the class-specific grasping point detection networks (synthetic image as input, spatial coordinates of the

<sup>1</sup> It should be noted that the use of deep neural networks not always entails the need of a costly off-line learning phase: in a cloth manipulation-related but different application, servocontrol learning of the position and deformation of soft material from the 3D point cloud, an online deep learning algorithm is presented in Hu et al. (2019). That is, the mapping between the manipulated and the feedback points of the deformable object is determined while manipulated.

two grasping points (SGD-I) or of all the feature and the two grasping points (SGD-II) as output) have similar architectures: four convolutional, four max-pooling and one hidden layer (SGD-II has two hidden layers), whereas the output layers are obviously different in each case.

Mentioning the problem stated at the beginning of this paragraph, but without adhering to a particular learning framework, Hou and Sahari (2019) propose obtaining a particle mesh model from the real image of a spread-out garment, whose hanging shape at each one of the nodes is then simulated at different orientations. In this way, a large dataset can be generated without having to resort to the real hanging images (nor constructing a model by hand).

### 2.3.2 Feature recognition methods

The aim of these approaches is to identify category-specific features in order to classify the cloth item at hand. Thus, despite the relative position of some features may provide clues for state estimation, these approaches are generally restricted to the classification problem. Class recognition is more reliable the more of such features are employed. Next, we review some work done in feature-based classification:

- **Fabric type.** The material and weaving (or knitting) pattern may already discriminate large groups of garment types: underwear is unlikely made of knitted wool, satin may help to distinguish a blouse from a rougher shirt, denim may be the material of which trousers, shirts or jackets are made but not pullovers, etc. The aspect of the wrinkles that appear when the garment is randomly thrown over the table helps Yamazaki and Inaba (2013) to identify the fabric type, using the Gabor filtering technique presented in previous work Yamazaki and Inaba (2009); Yamazaki et al. (2011) (see also Section 2.1). Not only wrinkles but also other features such as cloth overlaps, scale space extrema and contour (of the crumpled item, not to be confused with the contour methods of Section 2.3.1) and a SVM classifier help to determine the garment class. The same idea, material defining or at least constraining the set of garment types, inspires the work in Sun et al. (2017), who stress the robustness of this approach to occlusions and random configurations. As in their previous work, they resort to SI and LBP as descriptors for global features, together with a histogram of Topological Spatial Distances (TSD), which estimate the wrinkles' width and height. They also use local features, the Local B-Spline Patch (BSP) representation (restricted to the wrinkles, not the whole surface), encoded by LLC (Locality-constrained Linear Coding). Again, SVM is the chosen technique for classification, after fusing all descriptors.
- **Multi-level approach.** Combining low-level features (global ones including Color Histogram, Histogram of Line Lengths, Table Point Feature Histogram, and local such as SIFT and Fast Point Feature Histogram (FPFH)) allows Willimon et al. (2013) to determine via SVM some mid-level *characteristics* (buttons, pockets, hemlines...) which in turn are used in the high-level garment classification process.
- **Heels, toes and openings for state estimation.** The garment class (socks) is already known, but identifying these features from texture-based (MR8 filter bank and LBP) and shape-based (HOG) features allows Wang et al. (2011) to classify each sock within one of the canonical categories (sideways, heel up, heel down, and bunched states, all in the rightside-out or inside-out variants.)

Figure 4 summarizes the common computer vision skills seen so far.

## 2.4 Active vision strategies

Vision-guided manipulation does not only aim at fulfilling the nominal goals of the robot task, but also at enhancing the very same perception process, mainly by reducing uncertainty and ambiguity during grasp point localization, classification and state estimation. Most of the works that include such actions have been already referenced in the foregoing sections. Here we present a summary of the followed strategies:

- **Hanging clothes**
  - **Rotation:** The grasped and hanging cloth item is rotated before the camera(s), so that multiple views can be taken of the cloth in the same state, which increases the matching probability with stored garment types and/or states. This strategy has been followed in Osawa et al. (2007); Kita et al. (2009a, 2010); Maitin-Shepard et al. (2010); Bersch et al. (2011); Dumanoglou et al. (2014a); Mariolis et al. (2015); Kampouris et al. (2016); Saxena and Shibata (2019). In simulation, this method is equivalent to distributing several virtual cameras around the garment model Corona et al. (2018); Saxena and Shibata (2019).

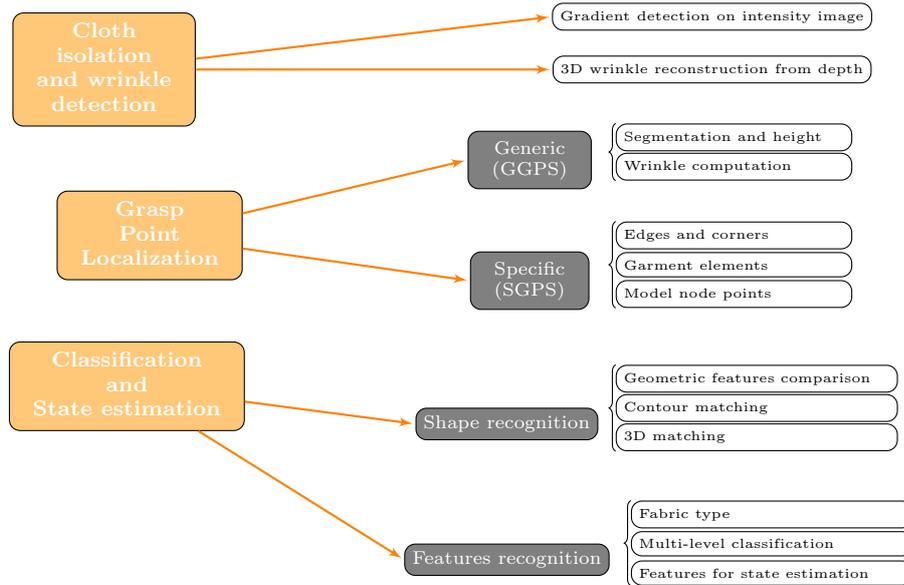


Fig. 4: Generic vision skills needed in assistive robot cloth manipulation.

- Shaking: Fast short motions of the grasped cloth item help to disentangle possible crumpled regions and to present the cloth in a more canonical way. See Osawa et al. (2007); Maitin-Shepard et al. (2010) for references using this strategy.
- Repeated lowermost point regrasping: This strategy aims at a progressive reduction and thus disambiguation of the possible states of the grasped item (up to symmetries). At each step the lowermost garment point is detected and grasped with a second robot, while the previously grasped point (by the first robot) is released. The final step generally involves no more release, but the two robots are now grasping the cloth item at different points. See Osawa et al. (2007); Cusumano-Towner et al. (2011); Saxena and Shibata (2019) for more details on specific implementations of this strategy.
- Spreading by pushing: This is also a bimanual active vision strategy, which consists in spreading the cloth part by pushing it with a second hand in front of the vision system Kita et al. (2010).
- **Clothes lying on a surface**  
Pulling at the corners and other flattening strategies help to attain a canonical configuration of a spread out garment which clearly enhances the classification chances of that particular item. A more specific method is described in Sun et al. (2016b), who define and apply the action *Grasp-Shake* (followed by dropping the garment back on the table, thus favoring the spreading-out of the cloth), and combine it with the action *Grasp-Flip*, which aims at displaying previously hidden parts of the garment. Applicability conditions for these two actions are defined, and if both are applicable they are randomly chosen.
- **Continuous perception**  
One step further consists in resorting to video sequences instead of fixed images: the vision process is now coupled with robot motion in a continuous fashion. For example, by recording the whole sequence of images while grasping and lifting garments in random configurations. This is the setting presented in Martínez et al. (2019) as a continuation of Sun et al. (2016b); Sun et al. (2017). They naturally resort to the same descriptors that had provided satisfactory results in their previous single-shot or static image works, plus other local and global descriptors, and the influence of each one on the overall performance has been thoroughly analyzed via an ablation study. Classification is now performed over sequences of frames, by using the Locality-Constrained Group Sparse Representation (LGSR) classification method.

## 2.5 Touch and Force

Robots that have to work in close proximity to people, even touching humans if required, must have some kind of sensory feedback both on contact and on the exerted force. Ideally, robots would be covered with an artificial sensitive skin or at least with a large number of touch sensors, and optimal performance would require humidity and temperature sensors as well. When it comes to manipulating cloth, touch and force may inform about issues including whether the cloth item has actually been grasped, whether one or more

garments have been picked up, the kind of textile material and quality of the fabric, or the exerted forces while dressing a patient. Next we review the achievements reached by research on these fields up to date.

### 2.5.1 Touch

The obvious way to provide some touch capabilities to the robot is to mount sensing devices on the gripper, more specifically on the fingertips. Such devices may be just simple binary presence detectors or sophisticated form-sensing matrices of sensors and force-measuring instruments. Tactile sensing technologies are systematized in Paul and Torgerson (1990) as for their different uses in cloth handling. Here we reproduce the resulting classification, together with some developments and devices:

- **Tension measuring** while stretching. In Gershon and Porat (1986); Gershon (1990) an implementation of a force-measuring device in a finger is mounted to this end, in the context of automated sewing, although this function is transferred to wrist-mounted F/T sensors in later works. In the context of assistive robotics, tensile force measuring is necessary in different applications such as bed-making or dressing.
- **Edge tracing**. In Salleh et al. (2006, 2008) special grippers and strategies have been developed for edge tracing while spreading a towel. The goal is to hold the towel at two consecutive corners. A pattern of IR LED – phototransistor couples is integrated in the fingertips of both the fixed gripper holding the towel at one corner and the gripper tracing the edge. These sensors report whether the gripper holds the towel or not, and in the first case if the grasp is at the interior, at the edge or at the corner of the cloth part (the latter being confirmed or not by an external CCD camera). Strain gauges are also included to keep the grasping force between the limits that ensure not losing contact while avoiding dragging the towel.
- **Texture measuring**. Recent projects like CLoPeMa (FP7-ICT-288553) have considered the need of distinguishing different fabric qualities by touch. By performing lateral finger movements, the tactile sensors mounted on the gripping surfaces described in Le et al. (2013) allow to gather data about textile roughness. The tactile experience can be enhanced if visual images are simultaneously considered (or vice-versa) in a multisensory integration approach, as done in Lee et al. (2019). The relationship between the two domains is learned via cGANs, a variant of Generative Adversarial Networks.
- **Fabric thickness**. A special finger design Ono et al. (1989, 1990) enables the measuring of the thickness of cloth between the fingers between 0.1 and 3.0 mm. The measuring device is a strain gauge mounted on the curved thin phosphor bronze finger, and the opposing finger is made from balsa wood. This device allows additionally to control the tension applied on the held fabric between 0.09 and 0.12 N. The authors report use of this gripper in pick and place Ono et al. (1990, 1991) and unfolding Ono et al. (1995, 1998) applications.

Some of these devices can accomplish also the basic tasks of cloth presence/absence detection (the optical sensors in Salleh et al. (2006, 2008) and the bronze finger in Ono et al. (1989, 1990)) or of contact detection with a surface, be it a table or a stack of fabric items (again Ono et al. (1989, 1990), although a simple touch sensor can do this as well Kemp et al. (1984)).

### 2.5.2 Force

We have just seen that there exist the possibility of measuring tensile forces with devices mounted on the gripper. Other possibilities include measuring the torques exerted at the actuators of the robot or to resort to a wrist-mounted force/torque (F/T) sensor.

- **Actuator torque measuring**. Actuator torques are not registered during the demonstration phase in the kinesthetic teaching setting for cloth handling in Lee et al. (2015), which compels the authors to perform this measuring while replaying the learned trajectory with high position and velocity gains. Alternatively, the external force applied to the robot can be inferred from its dynamic behavior, as done in Colomé et al. (2012, 2013, 2015). In their first works, just end-effector forces were computed, in particular to determine, by their weight, how many garments –if any– had been picked up by the robot from an unordered pile of clothes. This was achieved by learning the local Inverse Dynamic Model (IDM) of the employed Barrett’s WAM robot arm with a Locally Weighted Projection Regression algorithm, together with a robust disturbance state observer Colomé et al. (2012, 2013). Later the authors aimed at determining the forces exerted anywhere along the arm Colomé et al. (2015). Here the accent is put on obtaining a precise model of the friction term, as for low velocities and no external applied forces, friction is the second highest torque acting at each joint, after gravity. The parameters provided by the

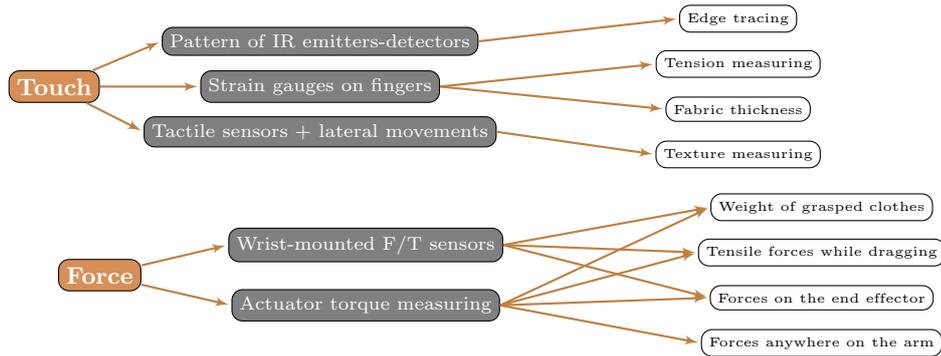


Fig. 5: Touch and force sensing in assistive robotic cloth manipulation.

robot’s manufacturer allow to obtain an explicit model of the robot’s IDM, and the inertia, Coriolis and gravity terms can also be computed. The integration of the complete IDM, together with a PD controller, in a Computed Torque Controller allows the robot to obtain a quite compliant response in the behavior of the robot. This compliance is a must in the scarf-wrapping application devised by the authors, where simple pushing of the arm enables the user to alter its trajectory online.

- **Wrist-mounted F/T sensors.** Besides being extensively used in sewing applications (see, for example, Schrimpf et al. (2014); Koustoumpardis and Aspragathos (2014) to cite some of the most recent ones), F/T sensors have been used to provide feedback on stretching out a held-up piece of cloth Consortium (2015). But the most interesting application, in the context of this survey, is the feedback they provide in dressing assistance tasks where vision alone proves to be insufficient, due to deformations and occlusions. In Kapusta et al. (2016) different force profiles are registered by a wrist-mounted F/T sensor in the case the insertion of the hand in the sleeve of a medical gown is missed, the case it is stuck during traversal, or the case the sleeve is successfully put along the patient’s arm. Discrimination between these three situations is provided by a HMM classifier trained with 12 volunteers. A higher classification accuracy is obtained later by the same authors by first training on real subjects, and then extending training by physical simulation with the parameters obtained in the first phase, thus avoiding long and tiresome testing on persons Yu et al. (2017). In the same context, such a sensor enables the system to discriminate between different base layers, i.e., the underlying fabric type, when putting a jacket on a human user Chance et al. (2017b). Almost the same accuracy can be achieved using the cheaper inertial measurement unit (IMU) together with the robot’s (a Baxter) internal sensors (whereas the accuracy drops drastically if they are used independently).

It should be noted that, in the same fashion as touch sensors, force sensing can also provide feedback on basic tasks such as contact assessment with the working surface. Figure 5 summarizes the reviewed touch and force skills.

Deep learning has also been resorted to as training paradigm of force-based interactions between a human and a robot in a dressing assistance scenario Erickson et al. (2018). A deep recurrent model trained with haptic and kinematic measurements in a physical simulation scenario can be used, as demonstrated by the authors, by the real robot system to predict the effects of the real robot’s actions and consequently lower the physical impact on the patient if needed. The task is pulling a sleeve along a person’s arm, and thanks to using Model Predictive Control (MPC) actions can be replanned online in real time.

### 3 Cloth manipulation in daily domestic chores

A robotic system aimed at providing autonomy to a disabled user should be able to perform (or at least to assist at performing) the different domestic chores needed for daily living. Among these, we consider all those that involve some kind of manipulation of fabrics. These tasks require no physical contact with the human, as opposed to dressing, which is tackled in Section 4. In what follows, we analyse these tasks, their variants, and the specific perception needs for successful completion. Recall Figure 1 and the previous sections to be aware of the role each perception skill has to play.

### 3.1 Cloth picking

In assistive environments, to pick up a cloth item is a task that arises in diverse situations, and thus requires different approaches, both on the perceptual and the manipulative sides. One such situation is to pick up garments thrown over a chair or on a bed, or even dropped on the floor. Depending on the garment type and its state, the appropriate action will be to put it on a hanger (e.g. a jacket) or to put it in the laundry bin (e.g., dirty socks). The involved perception assignments include:

- Cloth localization within a scene, in the fashion of Yamazaki and Inaba (2009), who survey a domestic scenario in search of the characteristic wrinkles that distinguish fabrics from other materials.
- Classification: by garment type, to decide a possible destination depending on the class.
- State estimation: rather than folded states or the mesh node it is hanging from, as seen in Section 2.3, what is required here are rough evaluations of the type dirty/clean, wet/dry, crumpled/flat, etc.
- GGPS, in the case the cloth item has to be just put away, in a laundry bin, or on a heap; otherwise:
- SGPS, e.g. the shoulders of a coat that has to be put on a hanger. This may require a regrasping strategy.

Another situation is to gather a folded item lying on a table to be put inside a drawer, or to take a garment from a drawer or hanger and present it to the user for dressing. Again, the casuistry is far from simple, as the duties of the perception systems will differ depending on the requirements of the user: e.g. whether any shirt or a specific one is wanted, in the latter case pattern and/or color matching will be necessary. The origin and destination of the cloth item (table, shelf, drawer, hanger) conditions the grasp type that best suits the involved manipulation (i.e., form preserving underneath grasps of folded items, or grasps at the shoulders for hanging, etc.) and thus the perceptual skill of grasping point localization.

A third situation is the laundry scenario. Dirty clothes have to be put inside the washing machine, washed clothes into the dryer or hung on a clothesline, clean and dry clothes may undergo ironing and/or folding processes. Some of the related perceptual skills are:

- GGPS: the aim is to grasp and pick up any item from a pile of washed clothes, what in Hamajima and Kakikura (1996, 2000) is called the *isolating task* (see also Monso et al. (2012) for a more recent approach). This task arises also in the previous step of “feeding” the washing machine from a heap of dirty clothes.
- Classification: previous to washing, some discriminating criteria may apply different from the garment type classification reviewed in Section 2.3, including colour/white, resistant/delicate, and the like.
- Classification: after washing and previous to drying, sorting by dryer-compatible or not (despite intrinsically being an issue related to fabrics material and quality, in general it can be shortcut by identifying the garment type).
- Classification: after drying, sorting by garment type for ironing and/or folding.

### 3.2 Folding

Folding is a common domestic chore, also performed in other assistive environments, through which cloth items are reshaped into manageable configurations suitable for storage. The task is accomplished by a sequence of individual bends affecting one (at the beginning) or more cloth layers. The most common folds involve the simultaneous grasping of two corners along an edge (thus assuming bimanual operation) and their transfer and release over the opposite edge or some intermediate location. Most foldings are done with the cloth item lying flat on a table Karakerezis et al. (1994); Paraschidis et al. (1995); Paul (2004); Osawa et al. (2006); Berg et al. (2011); Miller et al. (2012); Lakshmanan et al. (2012); Lee et al. (2015); Li et al. (2015b), although some researchers assume the cloth being already grasped and put the accent on the initial laying on the surface Zoumponos and Aspragathos (2010); Balaguer and Carpin (2011), and others consider foldings in the air Yamakawa et al. (2011); Colomé and Torras (2018).

The vision system has to provide feedback in the following tasks:

- SGPS, usually corners on the same edge. Often two or more cloth layers have to be grasped, which can be confirmed not only by visual means but also by sensors in the grippers.
- Classification allows the robot to trigger garment class specific folding procedures. That is, for each cloth type a predefined sequence of folds, starting at the extended canonical configuration, can be executed.
- State estimation is required at the beginning of a folding process, both for the hanging as well as for the lying situations. Although an initial canonical state may be assumed for simplicity, more versatile systems should be able to determine the state the cloth item is actually in, and execute the necessary regrasp or unfolding actions to bring the cloth item to the start configuration of the folding sequence.

- State estimation during a folding process evaluates the degree of attainment of the predicted folded states. Recall for example the rectangleness criterion of folded shapes explained in Section 2.3.1. By using a Kinect camera, Colomé and Torras (2018) not only resort to this geometric criterion, but also to a wrinkledness measure to determine the degree of attainment of a proper folded state.

### 3.3 Unfolding, spreading

The most common ways to unfold and spread cloth is either by performing the reverse fold operations on the item lying on a surface, together with pinching and dragging actions Ono et al. (1998); Willimon et al. (2011b), or via regrasps of the garment held in the air Hamajima and Kakikura (2000); Li et al. (2015a). More dynamic actions via swift manipulations and exploiting inertia for a first laying on a surface can also be considered (see Triantafyllou and Aspragathos (2011) for small rectangular parts, larger items including tablecloth still deserve research).

Perception duties include:

- SGPS, usually corners to unfold or to lift. While hanging, possibly the lowermost point is searched for regrasping.
- State estimation is required to determine whether the cloth item is completely spread out or still some folds remain.

### 3.4 Flattening

Maybe all the folds are removed, but still wrinkles prevent the cloth to be considered in its final state. Wrinkle removal can be done in various ways, depending on the previous state of the cloth item. Persistent, unaesthetic wrinkles (e.g. the ones on a shirt out of the dryer) are removed via ironing, whereas momentary wrinkles such as the ones appearing while spreading a tablecloth on a table are best removed by pulling and brushing. Ironing combines the proper iron drawing motions with repositioning, unfolding and folding actions. Iron drawing actions performed by human professionals have been analyzed in Dai (2012), to extract both basic *ironing profiles* and *ironing paths* (typical combinations of profiles) for an ironing robot. As for the second type of flattening actions, combinations of longitudinal brushing over wrinkles with pinching and pulling by a single-armed robot have been addressed in Sun et al. (2015, 2016a) and by bimanual manipulation Lee et al. (2015). More recently, Sun et al. (2018) do also resort to a dual-arm flattening setting, using the hierarchical visual architecture mentioned in Section 2.1.

In all cases, the main issue is to detect the wrinkles in order to proceed to their removal, but other sensing tasks are needed as well:

- Wrinkle detection. In the context of ironing, wrinkles are distributed more or less generally all over the cloth surface, and the motions of the iron are intended at covering the whole area. Some specific larger creases, appearing if the cloth has not been sufficiently extended, have to be monitored (or directly eliminated by pulling) as ironing over them may create sharp, costly-to-remove wrinkles. In the case of flattening by sweeping and dragging, the largest wrinkles are to be detected and measured. Pulling efforts are most effective orthogonally to the main direction of the wrinkle, whereas smaller wrinkles may be eliminated by gently sweeping over them.
- Edge and corner detection is necessary for grasping and pulling the cloth for wrinkle elimination. In this case, as mentioned before, the features to be grasped have to be searched in a direction perpendicular to the wrinkle.
- Force measuring and control is required in these operations, mainly while pulling the cloth: the dragging force has to be larger than the cloth-surface friction forces to allow the elimination of the wrinkle, but not excessively large to avoid pulling the whole cloth around and possibly creating new wrinkles.

### 3.5 Bed making

This task has been little explored in the Robotics literature despite being a fundamental domestic and hospital chore. It is a complex task that involves several simpler manipulations of diverse cloth elements, including bed sheets (both flat and fitted), blankets, duvets (and covers), comforters, quilts, and pillows (and pillowcases). Larger elements have to be extended, which means unfolding and possibly dynamic spreading

over the bed. Specific manipulations in bed making include extending the fits of a fitted bottom sheet over the mattress corners, tucking portions of bed sheets (flat bottom and top sheets) under the mattress, which in turn may require several manipulations as for example in the case of the so-called hospital corners, or fluffing the pillows before placing them at the head of the bed. Additional tasks may be quite complex as well, like putting the cover on the duvet. Other operations are more common to the cloth handling repertoire, such as flattening by pulling and sweeping, or removing items by grasping them at arbitrary locations. Bed making has been addressed in a quite recent work Seita et al. (2019), via detecting suitable pick points and pulling of the grasped bedcloth towards still uncovered parts of the bed. Details are provided below.

As for requirements on the perception side, the following tasks have to be addressed:

- GGPS, for removing pillows, duvets, top blankets, etc. before starting to make the bed. Larger elements will have to be grasped at two points, which should not be too close to each other.
- SGPS, basically the corners of the different elements. For the larger elements, the corners are too distant from one another, and convenient points on the same edge will have to be grasped instead for operations such as spreading (bimanual grasps) or pulling for flattening (single or bimanual). Suitable grasping points are determined in Seita et al. (2019) by deep transfer learning from depth images. The grasp detection network architecture includes a pre-trained YOLO single shot object detection network for feature extraction, plus two convolutional and two dense networks. Training was performed with human demonstrations. As for tucking portions of sheets under the mattress, this is not properly grasping but rather pushing (maybe while slightly lifting the mattress with the other hand), but this requires as well identifying the most suitable location on where to push, next to the portion that is already below the mattress.
- State estimation is about determining the current spatial bed/bedcloth relationship, in order to correct misalignments in orientation and wrong positionings (e.g., not being symmetric with respect to the bed’s main axis), and the degree of accomplishment along the bed making process. In Seita et al. (2019) the bed-making progress is measured through the resulting coverage of the last pulling action. Again, a deep neural network was trained to state, from the depth image, if the side of the bed where the robot stands is sufficiently covered.
- Wrinkle detection for flattening by pulling and sweeping.
- Force measuring and control is a must in tasks of the kind of fitting the corners of the bottom sheet or tucking sheets under the mattress.

### 3.6 Table setting

Like for the preceding task, very little has been done up to date for this application. Most notable is the work presented in Saxena et al. (2017), where a collaborative human-robot setting for laying tablecloth on a table is described. Both the human and the robot hold two corners of the tablecloth each one, and aim at positioning the cloth on the table by lowering the grasped corners. A head-mounted Kinect camera on the Baxter robot provides the necessary image feedback, which is submitted to the usual processes of background segmentation and contour detection. Gabor filtering is applied for wrinkle detection and also the corners are found by using the Shi-Tomasi detector. A structural similarity index is computed for the depth image and compared to the one corresponding to a successful placement of the tablecloth. Success is attained if the final image is a smaller rectangle than the original one (when the cloth was still over the table, not on it), there are four corners and no wrinkles. These are one-shot trials, no further correction is attempted for misplaced or wrinkled cloths. In general, laying a cloth on the table would mean essentially the same tasks as bed-making, of course without the tucking beneath the mattress and considering a slightly different behaviour in wrinkle removal due to different friction coefficients between the cloth and the table as compared to cloth-mattress. As for the latter, the fabric type of the tablecloth is surely more significant: as pointed out in Saxena et al. (2017) heavier and stiffer cloth is less prone to wrinkle formation than cotton or silk. Force measuring is not as critical as in bed making, unless correcting motions during laying and positioning within a collaborative setting are addressed.

## 4 Robot-assisted dressing

Helping a person to dress up is a strongly dynamical task, subject to a high variability in execution. Moreover, it is performed in close contact with a user who is frequently in a quite vulnerable situation. This makes it advisable to resort to sensory input, not only to certify that a given final state has been

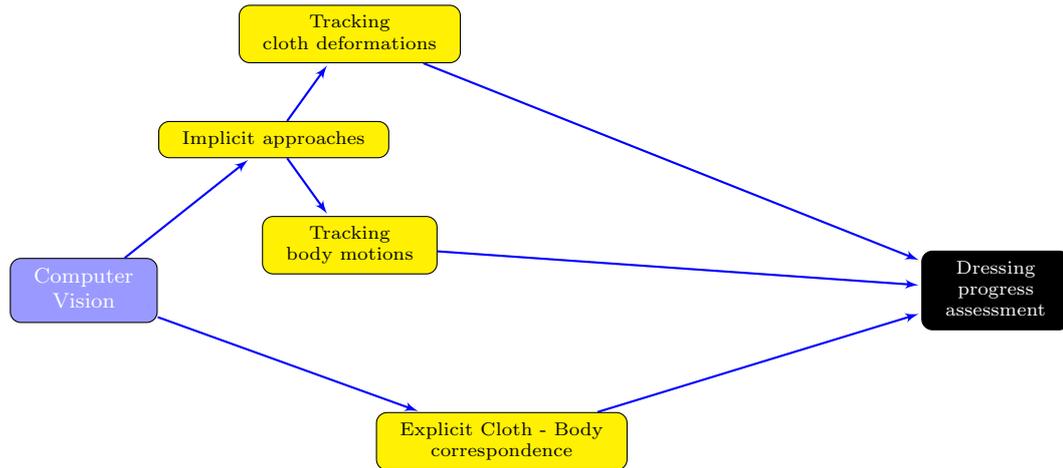


Fig. 6: Different approaches in processing visual information for robot-assisted dressing.

reached, but also to verify that a specific dressing action is proceeding correctly. E.g., if both feet have been introduced into the same trousers' leg, this should be corrected asap to avoid unnecessary discomfort or even pain. It should be noted that robotic assistance in dressing is a complex HRI subject which involves a great variety of aspects, including user's pose and gesture recognition Chance et al. (2017a) or resolving ambiguity of spoken deictic commands Jevtić et al. (2018). Although they are undoubtedly all relevant for the task, here we will concentrate on the manipulative aspects of helping a person to dress. Figure 6 displays the alternatives in visual tracking of robot-assisted dressing actions, as explained in the remainder of this section.

The progress of dressing actions can be evaluated either implicitly or by explicitly assessing the correspondence of human body parts with cloth features. Implicit approaches may in turn focus on cloth or on the human, that is, estimate the progress of dressing on the basis of processing just the cloth images, or assume that the robot grippers are already holding the garment in such a manner that just the correspondence between the end-effector and the body parts has to be tracked. All these alternatives are examined next:

- **Implicit cloth deformation tracking.** The optical flow-based method described by Yamazaki et al. (2013) performs a cloth-background segmentation and the analysis concentrates on the temporal evolution of cloth patches. Features such as the magnitude of flows of different cloth regions are computed, including the distance and relative angles of motion of the flows, as well as the local motion of the flows (i.e., wrt the center of the cloth region), and compared to the corresponding features of learned dressing sequences. Both correct and failed sequences have been considered while learning, which allows the system to identify the source of failure and eventually trigger the correcting action. The algorithm has been applied to the problem of dressing pajamas bottoms, and failures similar to the one pointed out above are implicitly identified by the resulting deformations on the pants.
- **Implicit body motion tracking.** It is assumed that the manipulator(s) is (are) holding the garment part at the convenient points (e.g. at the shoulders, two spaced points at the waistline, or the like) and by reaching a programmed sequence of poses, instantiated by the vision module, the dressing goal is achieved. In Klee et al. (2015) the emphasis is put on learning the user's mobility constraints, which (together with the sizes of the body parts of the particular user) instantiate the *templates* (sequences of desired poses of the robot) implementing the different dressing tasks. Robot motions are alternated with repositioning requests to the user, which, if not completely executed, allow to infer restrictions on the user's mobility. Here, the vision module has the mission of dimensioning the user and quantifying their degree of mobility, which is accomplished via depth image based skeleton tracking (OpenNI-tracker, on the depth images provided by a Kinect camera). In this way, the spatial coordinates of the user's joints (including the head position) can be tracked all the time. Experiments were conducted on putting a hat on users simulating to be constrained to sit on wheelchair and having limitations in repositioning their head. In a quite similar fashion, also the main emphasis is set on determining the actual range of human motions in Gao et al. (2015) and to adjust the robot motions accordingly. The top-view camera of the robotic system obtains zenithal images of the body and upper limbs of the user. In this work, the system is trained with RGB and depth images of a user that wears clothing with different colouring, facilitating the segmentation process into eight regions (L/R shoulders, upper arms, forearms and hands), in which

random decision forests intervene for robust classification of the pixels. The midpoint of each resulting region is computed and represents the position of the joints of the different body parts. The distribution of the positions of these joints is learned along a number of trials, as well as a classification of regions according to their accessibility by a user with specific mobility constraints. Such positions constitute the successive goals for a motion planner for each of the two arms of a Baxter robot, whose grippers hold the shoulders of the garment, a sleeveless jacket in this case. It should be stressed that for this kind of approaches, where the vision system focuses exclusively on the human pose, a large number of algorithms and even commercial software -as we have just seen- can be adapted, as the subject of human pose estimation is quite popular in a huge variety of applications.

- **Explicit cloth-body correspondence.** Vision now does not only concentrate on the user, but follows also the displacements and deformations of cloth features, and they are put in correspondence. For example, the collar of a T-shirt has to be aligned first with the top of the head and afterwards with the base of the neck. This is what Tamei et al. (2011); Koganti et al. (2013, 2014, 2015); Koganti et al. (2017); Koganti et al. (2019) aim at by using topological coordinates matching. The neckline (or the sleeve contour) is approximated by a polyline, and also the head-neck-body (or the arm) is represented as a sequence of segments (see Figure 7). The topological coordinates used by these researchers include the *writhe* (for the collar-body relationship), which quantifies the relative twisting of the curves around each other, and the center of location of these curves (for the collar-head and the sleeve-arm curve pairs). Tamei et al. (2011) employ markers and a motion capturing system to locate the body and the cloth features, whereas in Koganti et al. (2013) a commercial color and depth sensor captures the 3D shape of distinctively coloured collar and sleeves, and both motion capture plus depth sensor are combined in Koganti et al. (2014). The use of markers is restricted in Koganti et al. (2015) to a previous offline training phase, which allows later to rely exclusively on depth by resorting to the information stored in a shared latent space, using a dimensionality reduction technique called shared Gaussian Process Latent Variable Model (GP-LVM). This idea is extended further in Koganti et al. (2017) via a Bayesian Gaussian Process Latent Variable Model (BGPLVM), where approximate Bayesian inference together with the Automatic Relevance Detection (ARD) kernel avoids overfitting and enables automatic dimensionality reduction (the technique is known as Manifold Relevance Determination (MRD)). Finally, the authors demonstrate in Koganti et al. (2019) that policy search reinforcement learning in this task-specific latent space is much more efficient than learning in the high-dimensional joint configuration space of the robot. The GPLVM formalism is used by the same researchers in a more complete dressing assistance learning from demonstration setting that includes a cloth reaching phase (via point-to-point motion planning), an arm dressing phase (with Dynamic Movement Primitives, DMPs) and the body clothing phase (with GPLVM), as described in Joshi et al. (2019). Imitation learning from human demonstration is viewed as a form of prior knowledge that makes the learning process much more efficient.

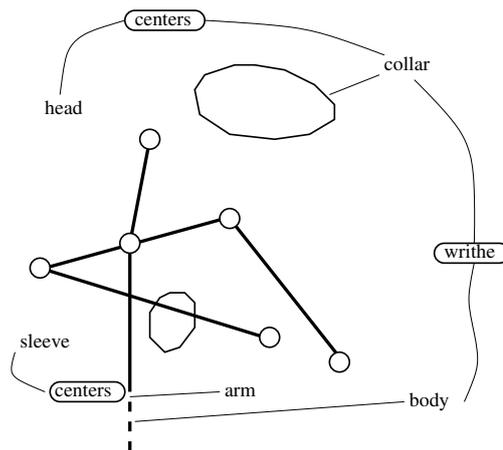


Fig. 7: The human body and the significant features of the clothing part (neckline or collar, and sleeve) represented as connected segment sets. The topological coordinates are exploited as for their expressive power regarding the relationships between human and garment features.

An issue arising in the context of vision-based user-centered tracking approaches is related to occlusions of the user’s body parts, including their limbs. Three types of occlusions may appear, namely garment-

occlusions (i.e., the very same garment hides a body part while dressing, as for example the hand that disappears inside the sleeve), robot-occlusions (due to the robot arm(s) interrupting the line of sight), or self-occlusions (the user’s own body interferes with the visibility of other parts of their body). While the two last types of occlusions may appear in different contexts, the first one is quite specific to the dressing assistance scenario. A possible way out is to infer the location of occluded body parts from the pose of other parts, as done in Chance et al. (2018). More specifically, the position of the elbow which has disappeared within the jacket is predicted from regression trees for different body features, which point at the shoulder and the hips as the most significant predictors (together with the same arm’s hand, but the latter is also occluded by the sleeve during the process of dressing). To this end, recurrent neural networks were trained on unoccluded motion tracking data, collected from a set of human-human assisted dressing interactions, and applied to the Kinect data obtained from a robot-assisted dressing task. An obvious but still quite unexplored alternative (or rather complementary procedure), would be to resort to cloth deformations, similarly as in the research cited above Yamazaki et al. (2013), but establishing an explicit correspondence between them and the location of body parts.

## 5 Conclusions

Cloth handling is tricky and involved, a challenge for robotic settings. In the case of rigid bodies, manipulation means generally just repositioning the object, but handling clothes goes far beyond simple translation or reorientation. Furthermore, it differs also from the manipulation of other types of deformable material such as paper or metal sheets, in that its deformations cannot be considered elastic (no original shape is recovered under cease of applied forces) nor plastic in a strict sense Jiménez (2012) (see also Nadon et al. (2018); Sanchez et al. (2018) for other recent surveys on robot manipulation of general deformable objects). In other words, cloth is very easy to deform, and these deformations are quite unstable under further manipulation. The mechanical behavior of textiles is hard to predict, and thus accurate and purposeful cloth manipulation is highly dependent on sensory feedback. This is even more marked in unstructured environments including assistive settings, with a tight and versatile interaction with humans.

In this survey, we have approached the topic of perception in robotic manipulation of cloth in assistive environments both from the side of describing the tools, i.e., the basic required perceptual skills, and the side of their use, that is, how such skills are applied in the execution of assistive tasks. The set of such basic skills include cloth isolation and wrinkle detection, grasping point(s) selection (both generic and specific ones), classification and state estimation. The solutions developed so far by research concerning these skills have been briefly and systematically described (some more detail can be found in Jiménez Schlegl (2017), although this survey does obviously not include the most recent contributions listed here). As for the applications, we have reviewed the most common domestic chores involving cloth manipulation and described the necessary role played by the different skills in the execution of these tasks. These are the tasks that do not imply physical contact with the human. But assistive tasks include also personal care in the case of disabled people, which regarding cloth manipulation means basically dressing assistance. In this application we have obviated the common perceptual skills (we have assumed the garments are already correctly identified and grasped at the most convenient locations) and have focused on the specific skills that are required in this context, namely tracking the development of the dressing task. We have seen that this can be done either implicitly by tracking cloth deformations or body motions, or explicitly by monitoring the cloth-body correspondence. This can be seen as a kind of state estimation, but in a continuous fashion. The methods and algorithms that have been presented along this survey are hard to compare with each other. Different environments and experimental settings, including non-identical garments and other cloth items, diverse simplifications and assumptions, among others, make it really difficult to establish a fair comparison as for performances. Establishing accurate benchmarks is a pending issue, and should include the very same experimental garments, if the physical aspects of manipulation are to be taken into account, for example in active sensing procedures.

Regarding future scientific research and technological development on the manipulation side, there is still a lot of work to do to come up with robotic hands that can equal human dexterity in grasping and manipulating textiles, not to say in fine manipulation tasks such as buttoning. As for perception, which is the topic of this survey, the achievements are already striking, but the to-do list is also still far from completed. Most experiments have been executed in controlled laboratory conditions and focus on one or two specific skills. This is far from the human ability to approach a bunch of clothes and rapidly decide the type and state of each individual item, to grasp them where needed, reverse an inside-out sleeve, match the pair of socks, fold what needs not to be ironed, or decide that that shirt needs to be washed again. High-resolution depth cameras may pave the way towards more accurate wrinkle measuring and state estimation, but processing

times are still too long for real-time applications. Perceptual skills have to gain in speed and accuracy, and have to be tightly coupled with manipulation in order to allow active vision strategies to resolve uncertainties in an agile way. Integration with other perception sources, mainly force and touch, is quite promising and should be further explored. In this sense, recent surveys such as Park et al. (2018) for tactile sensing may be really helpful. The crucial tasks of state estimation and tracking require further advances in versatility and uncertainty handling to come closer to effectively mimicking human comprehension of cloth states and our intuitive discretization of what is actually a continuum of deformed states. Some initiatives including the EU-project CLOTHILDE (ERC-2016-ADG-741930)<sup>2</sup> point in this direction.

## References

- Balaguer B, Carpin S (2011) Combining imitation and reinforcement learning to fold deformable planar objects. In: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, pp 1405–1412, DOI 10.1109/IROS.2011.6094992
- Berg JVD, Miller S, Goldberg K, Abbeel P (2011) Gravity-based robotic cloth folding. In: Hsu D, Isler V, Latombe JC, Lin M (eds) Algorithmic Foundations of Robotics IX, Springer Tracts in Advanced Robotics, vol 68, Springer Berlin Heidelberg, pp 409–424, DOI 10.1007/978-3-642-17452-0\_24
- Bersch C, Pitzer B, Kammel S (2011) Bimanual robotic cloth manipulation for laundry folding. In: Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, pp 1413–1419, DOI 10.1109/IROS.2011.6095109
- Chance G, Caleb-Solly P, Jevti A, Dogramadzi S (2017a) What’s up? resolving interaction ambiguity through non-visual cues for a robotic dressing assistant. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp 284–291, DOI 10.1109/ROMAN.2017.8172315
- Chance G, Jevti A, Caleb-Solly P, Dogramadzi S (2017b) A quantitative analysis of dressing dynamics for robotic dressing assistance. *Frontiers in Robotics and AI* 4:13, DOI 10.3389/frobt.2017.00013, URL <https://www.frontiersin.org/article/10.3389/frobt.2017.00013>
- Chance G, Jevti A, Caleb-Solly P, Aleny G, Torras C, Dogramadzi S (2018) elbows outpredictive tracking of partially occluded pose for robot-assisted dressing. *IEEE Robotics and Automation Letters* 3(4):3598–3605, DOI 10.1109/LRA.2018.2854926
- Colomé A, Torras C (2018) Dimensionality reduction for dynamic movement primitives and application to bimanual manipulation of clothes. *IEEE Transactions on Robotics* 34(3):602–615, DOI 10.1109/TRO.2018.2808924
- Colomé A, Pardo D, Alenyà G, Torras C (2012) External force estimation for textile grasp detection. In: Intelligent Robots and Systems (IROS), Workshop: Beyond Robot Grasping-Modern Approaches for Learning Dynamic Manipulation, 2012 IEEE International Conference on, URL <http://www.robot-learning.de/Research/IROS2012>
- Colomé A, Pardo D, Alenyà G, Torras C (2013) External force estimation during compliant robot manipulation. In: Robotics and Automation (ICRA), 2013 IEEE International Conference on, pp 3535–3540, DOI 10.1109/ICRA.2013.6631072
- Colomé A, Planells A, Torras C (2015) A friction-model-based framework for reinforcement learning of robotic tasks in non-rigid environments. In: Robotics and Automation (ICRA), 2015 IEEE International Conference on
- Consortium C (2015) The robot - clopema - clothes perception and manipulation. URL <http://clopemaweb.felk.cvut.cz/the-robot/>
- Corona E, Aleny G, Gabas A, Torras C (2018) Active garment recognition and target grasping point detection using deep learning. *Pattern Recognition* 74:629 – 641, DOI <https://doi.org/10.1016/j.patcog.2017.09.042>, URL <http://www.sciencedirect.com/science/article/pii/S0031320317303941>
- Cusumano-Towner M, Singh A, Miller S, O’Brien JF, Abbeel P (2011) Bringing clothing into desired configurations with limited perception. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA) 2011, pp 1–8, URL <http://graphics.berkeley.edu/papers/CusumanoTowner-BCD-2011-05/>
- Dai J (2012) Task analysis and motion generation for service robots: With reference to region segregation and path generation for robotic ironing. chap 3, pp 30–50, DOI 10.4018/978-1-4666-0291-5.ch003

<sup>2</sup> <https://www.iri.upc.edu/project/show/187>

- Doumanoglou A, Kargakos A, Kim TK, Malassiotis S (2014a) Autonomous active recognition and unfolding of clothes using random decision forests and probabilistic planning. In: Robotics and Automation (ICRA), 2014 IEEE International Conference on, pp 987–993, DOI 10.1109/ICRA.2014.6906974
- Doumanoglou A, Kim TK, Zhao X, Malassiotis S (2014b) Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V, Springer International Publishing, Cham, chap Active Random Forests: An Application to Autonomous Unfolding of Clothes, pp 644–658. DOI 10.1007/978-3-319-10602-1\_42, URL [http://dx.doi.org/10.1007/978-3-319-10602-1\\_42](http://dx.doi.org/10.1007/978-3-319-10602-1_42)
- Erickson Z, Clever HM, Turk G, Liu CK, Kemp CC (2018) Deep haptic model predictive control for robot-assisted dressing. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), pp 4437–4444, DOI 10.1109/ICRA.2018.8460656
- Gao Y, Chang HJ, Demiris Y (2015) User modelling for personalised dressing assistance by humanoid robots. In: Proceedings of the 27th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 1840–1845
- Gershon D (1990) Parallel process decomposition of a dynamic manipulation task: robotic sewing. Robotics and Automation, IEEE Transactions on 6(3):357–367, DOI 10.1109/70.56654
- Gershon D, Porat I (1986) Robotic sewing using multi-sensory feedback. In: Proc. 16th ISIR, Brussels, pp 823–834
- Gibbons P, Culverhouse P, Bugmann G (2009) Visual identification of grasp locations on clothing for a personal robot. In: Proceedings of Taros’09, Londonderry, pp 78–81, URL <http://104.131.144.114/download~54395a94335b7bbe298b4576>
- Hamajima K, Kakikura M (1996) Planning strategy for unfolding task of clothes – isolation of clothes from washed mass. In: SICE’96, vol 2, pp 1237–1242
- Hamajima K, Kakikura M (2000) Planning strategy for task of unfolding clothes. Robotics and Autonomous Systems 32(2–3):145–152, DOI 10.1016/S0921-8890(99)00115-3
- Hou YC, Sahari KSM (2019) Self-generated dataset for category and pose estimation of deformable object. Journal of Robotics, Networking and Artificial Life 5:217–222, DOI <https://doi.org/10.2991/jrnal.k.190220.001>, URL <https://doi.org/10.2991/jrnal.k.190220.001>
- Hou YC, Sahari KSM, Weng LY, How DNT, Seki H (2017) Particle-based perception of garment folding for robotic manipulation purposes. International Journal of Advanced Robotic Systems 14(6):1729881417738727, DOI 10.1177/1729881417738727, URL <https://doi.org/10.1177/1729881417738727>
- Hu Z, Han T, Sun P, Pan J, Manocha D (2019) 3-d deformable object manipulation using deep neural networks. IEEE Robotics and Automation Letters 4(4):4255–4261, DOI 10.1109/LRA.2019.2930476
- Jevtić A, Valle AF, Alenyà G, Chance G, Caleb-Solly P, Dogramadzi S, Torras C (2018) Personalized robot assistant for support in dressing. IEEE Transactions on Cognitive and Developmental Systems pp 1–1, DOI 10.1109/TCDS.2018.2817283
- Jiménez P (2012) Survey on model-based manipulation planning of deformable objects. Robotics and Computer-Integrated Manufacturing 28(2):154 – 163, DOI <https://doi.org/10.1016/j.rcim.2011.08.002>, URL <http://www.sciencedirect.com/science/article/pii/S0736584511000986>
- Jiménez Schlegel P (2017) Robotics and Autonomous Systems URL <http://digital.csic.es/handle/10261/166458>
- Joshi RP, Koganti N, Shibata T (2019) A framework for robotic clothing assistance by imitation learning. Advanced Robotics 0(0):1–19, DOI 10.1080/01691864.2019.1636715, URL <https://doi.org/10.1080/01691864.2019.1636715>
- Kampouris C, Mariolis I, Peleka G, Skartados E, Kargakos A, Triantafyllou D, Malassiotis S (2016) Multi-sensorial and explorative recognition of garments and their material properties in unconstrained environment. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp 1656–1663
- Kaneko M, Kakikura M (2001) Planning strategy for putting away laundry -isolating and unfolding task. In: Proc. of the 4th IEEE Int. Symposium on Assembly and Task Planning, pp 429–434
- Kapusta A, Yu W, Bhattacharjee T, Liu CK, Turk G, Kemp CC (2016) Data-driven haptic perception for robot-assisted dressing. In: 2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp 451–458, DOI 10.1109/ROMAN.2016.7745158
- Karakerezis A, Ippolito M, Doulgeri Z, Rizzi C, Cugini C, Petridis V (1994) Robotic handling for flat non-rigid materials. In: Systems, Man, and Cybernetics, 1994. Humans, Information and Technology., 1994 IEEE International Conference on, vol 1, pp 937–946, DOI 10.1109/ICSMC.1994.399957
- Kemp D, Taylor G, Taylor PM (1984) An adaptive, sensory gripper for fabric handling. In: Proc. of the 4th IASTED Int. Symp. on Robotics and Automation, Amsterdam

- Kita Y, Kita N (2002) A model-driven method of estimating the state of clothes for manipulating it. In: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision, IEEE Computer Society, Washington, DC, USA, WACV '02, pp 63–69, URL <http://dl.acm.org/citation.cfm?id=832302.836877>
- Kita Y, Saito F, Kita N (2004) A deformable model driven visual method for handling clothes. In: 2004 IEEE International Conference on Robotics and Automation, ICRA 2004, New Orleans (LA), USA, April, 2004, pp 3889–3895, DOI <http://dx.doi.org/10.1109/ROBOT.2004.1308874>
- Kita Y, Ueshiba T, Neo ES, Kita N (2009a) Clothes state recognition using 3d observed data. In: 2009 IEEE International Conference on Robotics and Automation, ICRA 2009, Kobe, Japan, May 12-17, 2009, pp 1220–1225, DOI <http://dx.doi.org/10.1109/ROBOT.2009.5152741>
- Kita Y, Ueshiba T, Neo ES, Kita N (2009b) A method for handling a specific part of clothing by dual arms. In: 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 11-15, 2009, St. Louis, MO, USA, pp 4180–4185
- Kita Y, Neo ES, Ueshiba T, Kita N (2010) Clothes handling using visual recognition in cooperation with actions. In: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems, October 18-22, 2010, Taipei, Taiwan, pp 2710–2715
- Kita Y, Kanehiro F, Ueshiba T, Kita N (2011) Clothes handling based on recognition by strategic observation. In: 11th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2011), Bled, Slovenia, October 26-28, 2011, pp 53–58
- Klee SD, Ferreira BQ, Silva R, Costeira JP, Melo FS, Veloso MM (2015) Personalized assistance for dressing users. In: Social Robotics - 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings, pp 359–369, DOI [10.1007/978-3-319-25554-5\\_36](https://doi.org/10.1007/978-3-319-25554-5_36), URL [https://doi.org/10.1007/978-3-319-25554-5\\_36](https://doi.org/10.1007/978-3-319-25554-5_36)
- Koganti N, Tamei T, Matsubara T, Shibata T (2013) Estimation of human cloth topological relationship using depth sensor for robotic clothing assistance. In: Proceedings of Conference on Advances In Robotics, ACM, New York, NY, USA, AIR '13, pp 36:1–36:6, DOI [10.1145/2506095.2506146](https://doi.org/10.1145/2506095.2506146), URL <http://doi.acm.org/10.1145/2506095.2506146>
- Koganti N, Tamei T, Matsubara T, Shibata T (2014) Real-time estimation of human-cloth topological relationship using depth sensor for robotic clothing assistance. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp 124–129, DOI [10.1109/ROMAN.2014.6926241](https://doi.org/10.1109/ROMAN.2014.6926241)
- Koganti N, Ngeo JG, Tamei T, Ikeda K, Shibata T (2015) Cloth dynamics modeling in latent spaces and its application to robotic clothing assistance. In: Proceedings of the 27th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 3464–3469
- Koganti N, Tamei T, Ikeda K, Shibata T (2017) Bayesian nonparametric learning of cloth models for real-time state estimation. *IEEE Transactions on Robotics* 33(4):916–931, DOI [10.1109/TRO.2017.2691721](https://doi.org/10.1109/TRO.2017.2691721)
- Koganti N, Shibata T, Tamei T, Ikeda K (2019) Data-efficient learning of robotic clothing assistance using bayesian gaussian process latent variable model. *Advanced Robotics* 33(15-16):800–814, DOI [10.1080/01691864.2019.1610061](https://doi.org/10.1080/01691864.2019.1610061), URL <https://doi.org/10.1080/01691864.2019.1610061>, <https://doi.org/10.1080/01691864.2019.1610061>
- Koustoumpardis PN, Aspragathos NA (2014) Intelligent hierarchical robot control for sewing fabrics. *Robotics and Computer-Integrated Manufacturing* 30(1):34–46
- Lakshmanan K, Sachdev A, Xie Z, Berenson D, Goldberg K, Abbeel P (2012) A constraint-aware motion planning algorithm for robotic folding of clothes. In: Desai JP, Dudek G, Khatib O, Kumar V (eds) *Experimental Robotics*, STAR 88, Springer, Springer Tracts in Advanced Robotics, vol 88, pp 547–562, URL <http://dblp.uni-trier.de/db/conf/iser/iser2012.html#LakshmananSXBG12>
- Le THL, Jilich M, Landini A, Zoppi M, Zlatanov D, Molfino R (2013) On the development of a specialized flexible gripper for garment handling. *Journal of Automation and Control Engineering* 1(3):255–259
- Lee AX, Lu H, Gupta A, Levine S, Abbeel P (2015) Learning force-based manipulation of deformable objects from multiple demonstrations. In: IEEE International Conference on Robotics and Automation, ICRA 2015, Seattle, WA, USA, 26-30 May, 2015, pp 177–184, DOI [10.1109/ICRA.2015.7138997](https://doi.org/10.1109/ICRA.2015.7138997), URL <http://dx.doi.org/10.1109/ICRA.2015.7138997>
- Lee J, Bollegala D, Luo S (2019) "touching to see" and "seeing to feel": Robotic cross-modal sensory data generation for visual-tactile perception. In: International Conference on Robotics and Automation, ICRA 2019, Montreal, QC, Canada, May 20-24, 2019, pp 4276–4282, DOI [10.1109/ICRA.2019.8793763](https://doi.org/10.1109/ICRA.2019.8793763), URL <https://doi.org/10.1109/ICRA.2019.8793763>
- Li Y, Chen CF, Allen PK (2014a) Recognition of deformable object category and pose. In: *Robotics and Automation (ICRA)*, 2014 IEEE International Conference on, pp 5558–5564, DOI [10.1109/ICRA.2014.6907676](https://doi.org/10.1109/ICRA.2014.6907676)

- Li Y, Wang Y, Case M, Chang SF, Allen PK (2014b) Real-time pose estimation of deformable objects using a volumetric approach. In: Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on, pp 1046–1052, DOI 10.1109/IROS.2014.6942687
- Li Y, Xu D, Yue Y, Wang Y, Chang S, Grinspun E, Allen PK (2015a) Regrasping and unfolding of garments using predictive thin shell modeling. In: IEEE International Conference on Robotics and Automation, ICRA 2015, Seattle, WA, USA, 26-30 May, 2015, pp 1382–1388, DOI 10.1109/ICRA.2015.7139370, URL <http://dx.doi.org/10.1109/ICRA.2015.7139370>
- Li Y, Yue Y, Xu D, Grinspun E, Allen PK (2015b) Folding deformable objects using predictive simulation and trajectory optimization. In: Proceedings of the 27th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 6000–6006
- Magenat-Thalmann N, Volino P (2005) From early draping to haute couture models: 20 years of research. *The Visual Computer* 21(8-10):506–519
- 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: International Conference on Robotics and Automation (ICRA), pp 2308–2315
- Mariolis I, Peleka G, Kargakos A, Malassiotis S (2015) Pose and category recognition of highly deformable objects using deep learning. In: 2015 International Conference on Advanced Robotics (ICAR), pp 655–662, DOI 10.1109/ICAR.2015.7251526
- Martínez L, del Solar JR, Sun L, Siebert JP, Aragon-Camarasa G (2019) Continuous perception for deformable objects understanding. *Robotics and Autonomous Systems* 118:220 – 230, DOI <https://doi.org/10.1016/j.robot.2019.05.010>, URL <http://www.sciencedirect.com/science/article/pii/S0921889019300417>
- Miller S, Fritz M, Darrell T, Abbeel P (2011) Parametrized shape models for clothing. In: International Conference on Robotics and Automation (ICRA), pp 4861–4868, DOI 10.1109/ICRA.2011.5980453
- Miller S, van den Berg J, Fritz M, Darrell T, Goldberg KY, Abbeel P (2012) A geometric approach to robotic laundry folding. *International Journal of Robotic Research* 31(2):249–267
- Monso P, Alenyà G, Torras C (2012) Pomdp approach to robotized clothes separation. In: 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2012, Vilamoura, Algarve, Portugal, October 7-12, 2012, pp 1324–1329
- Nadon F, Valencia A, Payeur P (2018) Multi-modal sensing and robotic manipulation of non-rigid objects: A survey. *Robotics* 7(4):74, DOI 10.3390/robotics7040074
- Ono E, Okabe H, Akami H, Aisaka N (1989) Robot hand with a sensor for cloth handling. *Journal of the Textile Machinery Society of Japan* 37(1):14–24, URL [https://www.jstage.jst.go.jp/article/jte1955/37/1/37\\_1\\_14/\\_pdf](https://www.jstage.jst.go.jp/article/jte1955/37/1/37_1_14/_pdf)
- Ono E, Okabe H, Ichijo H, Aisaka N, Akami H (1990) Robot hand with sensor for handling cloth. In: Intelligent Robots and Systems '90. 'Towards a New Frontier of Applications', Proceedings. IROS '90. IEEE International Workshop on, pp 995–1000 vol.2, DOI 10.1109/IROS.1990.262526
- Ono E, Ichijo H, Aisaka N (1991) Robot hand for handling cloth. In: Advanced Robotics, 1991. 'Robots in Unstructured Environments', 91 ICAR., Fifth International Conference on, pp 769–774 vol.1, DOI 10.1109/ICAR.1991.240582
- Ono E, Kita N, Sakane S (1995) Strategy for unfolding a fabric piece by cooperative sensing of touch and vision. In: Proc. Intelligent Robots and Systems 95. 'Human Robot Interaction and Cooperative Robots', vol 3, pp 441–445, DOI 10.1109/IROS.1995.525922
- Ono E, Sakane N, Sakane S (1998) Unfolding folded fabric using outline information with vision and touch sensors. *Journal of Robotics and Mechatronics* 10(3):235–243
- Osawa F, Seki H, Kamiya Y (2006) Clothes folding task by tool-using robot. *Journal of Robotics and Mechatronics* 18(5):618–625
- Osawa F, Seki H, Kamiya Y (2007) Unfolding of massive laundry and classification types by dual manipulator. *Journal of Advanced Computational Intelligence and Intelligent Informatics* pp 457–463
- Paraschidis K, Fahantidis N, Petridis V, Doulgeri Z, Petrou L, Hasapis G (1995) A robotic system for handling textile and non rigid flat materials. *Computers in Industry* 26(3):303–313, DOI 10.1016/0166-3615(95)00022-V, URL [http://dx.doi.org/10.1016/0166-3615\(95\)00022-V](http://dx.doi.org/10.1016/0166-3615(95)00022-V)
- Park M, Bok B, Ahn JH, Kim M (2018) Recent advances in tactile sensing technology. *Micromachines* 9(7), DOI 10.3390/mi9070321
- Paul FW (2004) Acquisition, placement, and folding of fabric materials. *International Journal of Clothing Science and Technology* 16(1/2):227–237, DOI 10.1108/09556220410520504
- Paul FW, Torgerson E (1990) Tactile sensors: Application assessment for robotic handling of limp materials. In: Taylor P (ed) *Sensory Robotics for the Handling of Limp Materials*, NATO ASI Series, vol 64, Springer

- Berlin Heidelberg, pp 227–237, DOI 10.1007/978-3-642-75533-0\_15, URL [http://dx.doi.org/10.1007/978-3-642-75533-0\\_15](http://dx.doi.org/10.1007/978-3-642-75533-0_15)
- Ramisa A, Alenyà G, Moreno-Noguer F, Torras C (2011) Determining where to grasp cloth using depth information. In: Artificial Intelligence Research and Development - Proceedings of the 14th International Conference of the Catalan Association for Artificial Intelligence, Lleida, Catalonia, Spain, October 26–28, 2011, pp 199–207, DOI 10.3233/978-1-60750-842-7-199
- Ramisa A, Alenyà G, Moreno-Noguer F, Torras C (2012) Using depth and appearance features for informed robot grasping of highly wrinkled clothes. In: 2012 IEEE International Conference on Robotics and Automation (ICRA), pp 1703–1708, DOI 10.1109/ICRA.2012.6225045
- Ramisa A, Alenyà G, Moreno-Noguer F, Torras C (2014) Learning rgb-d descriptors of garment parts for informed robot grasping. *Engineering Applications of Artificial Intelligence* 35:246 – 258, DOI <http://dx.doi.org/10.1016/j.engappai.2014.06.025>
- Salleh K, Seki H, Kamiya Y (2006) Tracing manipulation in clothes spreading by robot arms. *J Rob Mechatron* 18(5):564–571
- Salleh K, Seki H, Kamiya Y, Hikizu M (2008) Inchworm robot grippers for clothes manipulation. *Artificial Life and Robotics* 12(1–2):142–147, DOI 10.1007/s10015-007-0456-6
- Sanchez J, Corrales JA, Bouzgarrou BC, Mezouar Y (2018) Robotic manipulation and sensing of deformable objects in domestic and industrial applications: a survey. *The International Journal of Robotics Research* 37(7):688–716, DOI 10.1177/0278364918779698, URL <https://doi.org/10.1177/0278364918779698>, <https://doi.org/10.1177/0278364918779698>
- Saxena K, Shibata T (2019) Garment recognition and grasping point detection for clothing assistance task using deep learning. In: 2019 IEEE/SICE International Symposium on System Integration (SII), pp 632–637, DOI 10.1109/SII.2019.8700343
- Saxena K, Labuguen R, Joshi RP, Koganti N, Shibata T (2017) A study on human-robot collaboration for table-setting task. In: 2017 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp 183–188, DOI 10.1109/ROBIO.2017.8324415
- Schrimpf J, Bjerkgeng M, Mathisen G (2014) Velocity coordination and corner matching in a multi-robot sewing cell. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, September 14–18, 2014, pp 4476–4481, DOI 10.1109/IROS.2014.6943196, URL <http://dx.doi.org/10.1109/IROS.2014.6943196>
- Seita D, Jamali N, Laskey M, Berenstein R, Tanwani AK, Baskaran P, Iba S, Canny J, Goldberg K (2019) Deep Transfer Learning of Pick Points on Fabric for Robot Bed-Making. In: International Symposium on Robotics Research (ISRR)
- Stria J, Průša D, Hlaváč V (2014a) Polygonal models for clothing. In: Mistry M, Leonardis A, Witkowski M, Melhuish C (eds) *Advances in Autonomous Robotics Systems*, Lecture Notes in Computer Science, vol 8717, Springer International Publishing, pp 173–184, DOI 10.1007/978-3-319-10401-0\_16, URL [http://dx.doi.org/10.1007/978-3-319-10401-0\\_16](http://dx.doi.org/10.1007/978-3-319-10401-0_16)
- Stria J, Průša D, Hlaváč V, Wagner L, Petřík V, Krsek P, Smutný V (2014b) Garment perception and its folding using a dual-arm robot. In: Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on, pp 61–67, DOI 10.1109/IROS.2014.6942541
- Sun J, Yao M, Xu B, Bel P (2011) Fabric wrinkle characterization and classification using modified wavelet coefficients and support-vector-machine classifiers. *Textile Research Journal* 81(9):902–913, DOI 10.1177/0040517510391702
- Sun L, Aragon-Camarasa G, Rogers S, Siebert J (2015) Accurate garment surface analysis using an active stereo robot head with application to dual-arm flattening. In: 2015 IEEE International Conference on Robotics and Automation (ICRA), pp 185–192, DOI 10.1109/ICRA.2015.7138998
- Sun L, Camarasa GA, Khan A, Rogers S, Siebert P (2016a) A precise method for cloth configuration parsing applied to single-arm flattening. *International Journal of Advanced Robotic Systems* 13(2):70, DOI 10.5772/62513, URL <https://doi.org/10.5772/62513>, <https://doi.org/10.5772/62513>
- Sun L, Rogers S, Aragon-Camarasa G, Siebert J (2016b) Recognising the clothing categories from free-configuration using gaussian-process-based interactive perception. In: 2016 IEEE International Conference on Robotics and Automation (ICRA), pp 2464–2470
- Sun L, Aragon-Camarasa G, Rogers S, Stolkin R, Siebert JP (2017) 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), pp 6699–6706, DOI 10.1109/IROS.2017.8206586
- Sun L, Aragon-Camarasa G, Rogers S, Siebert JP (2018) Autonomous clothes manipulation using a hierarchical vision architecture. *IEEE Access* 6:76646–76662, DOI 10.1109/ACCESS.2018.2883072

- Tamei T, Matsubara T, Rai A, Shibata T (2011) Reinforcement learning of clothing assistance with a dual-arm robot. In: *Humanoid Robots (Humanoids)*, 2011 11th IEEE-RAS International Conference on, pp 733–738, DOI 10.1109/Humanoids.2011.6100915
- Triantafyllou D, Aspragathos NA (2011) A vision system for the unfolding of highly non-rigid objects on a table by one manipulator. In: *Intelligent Robotics and Applications - 4th International Conference, ICIRA 2011, Aachen, Germany, December 6-8, 2011, Proceedings, Part I*, pp 509–519, DOI 10.1007/978-3-642-25486-4\_51, URL [http://dx.doi.org/10.1007/978-3-642-25486-4\\_51](http://dx.doi.org/10.1007/978-3-642-25486-4_51)
- Triantafyllou D, Mariolis I, Kargakos A, Malassiotis S, Aspragathos N (2016) A geometric approach to robotic unfolding of garments. *Robot Auton Syst* 75(PB):233–243, DOI 10.1016/j.robot.2015.09.025, URL <https://doi.org/10.1016/j.robot.2015.09.025>
- Wang PC, Miller S, Fritz M, Darrell T, Abbeel P (2011) Perception for the manipulation of socks. In: *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2011, San Francisco, CA, USA, September 25-30, 2011*, pp 4877–4884, DOI 10.1109/IROS.2011.6095013, URL <http://dx.doi.org/10.1109/IROS.2011.6095013>
- Willimon B, Birchfield S, Walker I (2011a) Classification of clothing using interactive perception. In: *Robotics and Automation (ICRA)*, 2011 IEEE International Conference on, pp 1862–1868, DOI 10.1109/ICRA.2011.5980336
- Willimon B, Birchfield S, Walker I (2011b) Model for unfolding laundry using interactive perception. In: *Intelligent Robots and Systems (IROS)*, 2011 IEEE/RSJ International Conference on, pp 4871–4876, DOI 10.1109/IROS.2011.6095066
- Willimon B, Walker I, Birchfield S (2013) Classification of clothing using midlevel layers. *ISRN Robotics* DOI <http://doi.acm.org/10.5402/2013/630579>
- Yamakawa Y, Namiki A, Ishikawa M (2011) Motion planning for dynamic folding of a cloth with two high-speed robot hands and two high-speed sliders. In: *Robotics and Automation (ICRA)*, 2011 IEEE International Conference on, pp 5486–5491, DOI 10.1109/ICRA.2011.5979606
- Yamazaki K, Inaba M (2009) A cloth detection method based on image wrinkle feature for daily assistive robots. In: *Proceedings of the IAPR Conference on Machine Vision Applications (IAPR MVA 2009)*, Keio University, Yokohama, Japan, May 20-22, 2009, pp 366–369
- Yamazaki K, Inaba M (2013) Clothing classification using image features derived from clothing fabrics, wrinkles and cloth overlaps. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Tokyo, Japan, November 3-7, 2013, pp 2710–2717
- Yamazaki K, Nagahama K, Inaba M (2011) Daily clothes observation from visible surfaces based on wrinkle and cloth-overlap detection. In: *Proceedings of the IAPR Conference on Machine Vision Applications (IAPR MVA 2011)*, Nara Centennial Hall, Nara, Japan, June 13-15, 2011, pp 275–278
- Yamazaki K, Oya R, Nagahama K, Inaba M (2013) A method of state recognition of dressing clothes based on dynamic state matching. In: *System Integration (SII)*, 2013 IEEE/SICE International Symposium on, pp 406–411, DOI 10.1109/SII.2013.6776728
- Yu W, Kapusta A, Tan J, Kemp CC, Turk G, Liu CK (2017) Haptic simulation for robot-assisted dressing. In: *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pp 6044–6051, DOI 10.1109/ICRA.2017.7989716
- Zoumponos G, Aspragathos N (2010) A fuzzy strategy for the robotic folding of fabrics with machine vision feedback. *Industrial Robot* 37(3):302–308, URL <https://www.scopus.com/inward/record.url?eid=2-s2.0-77951898664&partnerID=40&md5=45350a8053770007f33ecba076502569>