

Modelling folded garments by fitting foldable templates

Ioannis Mariolis · Sotiris Malassiotis

Received: 13 June 2014 / Revised: 18 December 2014 / Accepted: 22 February 2015 / Published online: 11 March 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract This work presents a novel method performing shape matching between folded garments and corresponding unfolded templates. It incorporates both partial and global shape analysis techniques, estimating point correspondences between contours of folded garments and generic templates of unfolded garments. The established correspondences are used in the estimation of polygonal models of the folded garments. The initial matching results are also used for modifying the original templates to better fit the folded garments. The method is applied in an iterative fashion, using the fitted templates as a new input, resulting to more accurate matching as indicated by the experimental results. The produced polygon models can be used for planning the unfolding strategy during autonomous robotic manipulation.

Keywords Fold detection · Partial matching · Shape analysis · Unfolding

1 Introduction

Autonomous unfolding of real garments is still an extremely challenging task for robotic manipulators, mainly due to the required perceptual abilities. The problem of perception is particularly challenging because garments are deformable objects, lying in high-dimensional configuration spaces. One of the critical challenges in automating garment unfolding is the ability to recognize the configuration of a not fully unfolded article of clothing. A limited number of existing studies address this problem, employing different assumptions and approaches.

A template matching approach has been adopted in Osawa et al. [1], which focuses on the unfolding of massive laundry. The degree of similarity between templates of folded garments is calculated by evaluating covariance between them and images of the manipulated garments. To limit the set of employed templates, a lowest grasping point approach has been adopted during manipulation. In Hamajima et al. [2], unfolding is facilitated by the detection of hemlines, which was based on the appearance of shadows and the outline of the clothes in the hung-up state. However, in case no hemline is detected a lowest grasping point approach is also adopted. Maitin et al. [3] address robotic handling of towels using stereo cameras for searching geometric cues. Based on the classification of depth-discontinuity edges, the towel's corner grasp points were detected and used for its reconfiguration. A system handling different types of garments by comparing observed contours to simulated ones is presented in Cusumano et al. [4]. The system is using a strain-limiting finite element model for cloth simulation, whereas the tracking of the garment's configuration is based on a Hidden Markov Model that uses simple perceptual cues consisting of the height of the garment when held by a single gripper and its silhouette when held by two grippers. The comparison between the observed and the simulated contours is performed by means of the dynamic time warping algorithm. A system that is capable of unfolding a crumpled T-shirt equipped with fiducial markers is presented in Bersch et al. [5]. The system is based on machine learning techniques applied to a dataset of random grasps. Employing the same techniques the system can also bring the T-shirt into a folded configuration. The presented studies are aiming to unfold hung garments by locating suitable re-grasping points.

In a series of works, Willimon et al. [6–9], address the problems of interactive perception and surface estimation of non-rigid deformable objects, such as real garments. In [6],

I. Mariolis (✉) · S. Malassiotis
Centre for Research and Technology-Hellas (CERTH), Thessaloniki,
Greece
e-mail: ymariolis@iti.gr

features such as corners, peak region, and continuity of the cloth are used to determine a location and orientation to interact with the cloth and unfold or flatten it, whereas in [7] colour, shape, and flexibility features are employed for interacting and isolating a large variety of both rigid and non-rigid objects. In [8], an energy minimization approach and its extension have been proposed for estimating the current configuration of a non-rigid object using RGBD images, whereas the method is extended in [9], using graph cuts and removing the need for feature correspondence or texture information. In another recent study [10], Wilimmon et al. propose the use of a multilayer approach to classify garments within a laundry heap into seven categories. The problem is compartmentalized into low, mid-level, and high layers, employing color, texture, shape, and edge information using both local and global perspective. The proposed classification approach has been compared to a baseline system using support vector machines, and a superior performance has been reported.

In a different direction, in Triantafyllou et al. [11], edges and corners of a piece of fabric lying on a table are extracted and classified according to an estimated topology. The classified corners are used to detect folds and facilitate unfolding by a single manipulator. However, that study addresses folds on polygonal pieces of fabric instead of real garments. Unfolding of planar polygonal pieces of fabric has been also addressed in Ono et al. [12] using visual and tactile sensors. In that study, a template of the unfolded fabric is acquired and partially matched to its folded configuration. Using the matching results, the position of a corner point is predicted on the folded fabric as candidate for grasping. If after a trial to unfold the fabric the point is considered erroneous, estimation is improved by repeating the process with remaining candidates or performing partial pattern-matching again. The employed partial matching approach is matching the side lengths and corners of the folded fabric to those of an unfolded template. Although, the employed template is an image of the fabric before folding, a significant percentage of erroneous matches is reported.

In Miller et al. [13], the inverse problem of folding autonomously a spread-out garment lying on a table is addressed. In that study, fitting a parameterized shape model to the unfolded garment is proposed. However, to provide feedback for the folding result, the proposed model can be augmented to include the desired folding line, as well. Using a coarse-to-fine energy optimization approach, the augmented model is fitted to the resulted configuration after folding, closing the perception-action loop. In the experimental evaluation it is reported that a static approach has been employed, where the optimization is performed over the folding axis, leaving the unfolded portion of the model untouched. Although, adequate experimental results are provided for evaluating the accuracy of the model fitting in the spread-out case, no quantitative results are reported on the fitting performance in case

of the augmented model. The proposed approach is optimal for the task at hand, i.e. re-examining the garment and providing feedback on the resulted fold. However, when considered for fitting a model to a folded garment in a more general setup, it presents certain limitations. Similarly to [12], the unfolded configuration of the garment must be known when addressing the folded configuration. Moreover, a parameterized model for the unfolded configuration must be already extracted, whereas an approximate position of the fold must be also known.

In this work, modelling the folded garment by fitting an unfolded template to its contour, is proposed, to identify suitable re-grasping points for unfolding. The proposed approach considers approximately planar configurations of the folded garments, presenting a zero-curvature fold. The approach is based on the critical, yet easily met, assumption that the folding axis is part of the outer contour of the folded garment. This assumption is always valid when only a single zero-curvature fold exists. Although, the presented method mainly addresses single folds, it allows recursive application in case of multiple folds, simply by using the virtually folded template of the current iteration as the unfolded template of the next iteration.

The main application assumption is that using a bimanual robot, a crumpled garment can be brought into an approximately planar configuration by picking it from two points near its outline. A robotic algorithm has been developed for performing such task autonomously, but it will not be described in this study due to space limitations. When the manipulated garment is subsequently laid on a table it will present one or more folds. At that point, the garment's type and unfolded configuration are considered unknown. A set of generic unfolded templates is employed and it is known that the folded garment belongs to the same garment type as one of the templates. Also unknown is considered the location of the folding axis on the unfolded template. The proposed approach matches the folded contour to an unfolded template using a partial matching approach and predicts the location of the folding axis by testing hypotheses on its location. Matching results are, then, used for updating the template to better fit the folded contour. Using the matching correspondences between the folded contour and the folded garment a polygon model is estimated for the latter. The model can be used for identifying landmark points on the folded contour suitable for unfolding. Similarly to [10], the proposed method employs both local and global perspective, with partial matching operating on a local scale and hypothesis testing and modelling operating on a global scale. However, as opposed to [10], which employs a supervised learning approach, no training is needed, depleting the need of acquiring large annotated training datasets.

This paper significantly extends previous work on folded garment modelling [14]. The main contribution of the cur-

rent study is that a novel method for modifying the unfolded template to better fit the outline of the examined garment is proposed. The use of a fitted template is aiming to tackle the challenges introduced by the garments' large variability in style and shape and their deformable properties. Moreover, new measures for evaluating the method's performance are introduced and more extensive experiments on real garments are presented. Finally, a polygon model is also estimated using the matching results and suitable re-grasping points are identified on the garment's contour to bring it to an unfolded configuration.

The proposed method can be applied to any kind of folded shapes as long as corresponding unfolded templates are employed. However, in case the shapes correspond to the contours of folded garments, additional challenges are introduced due to non-rigid deformations and articulations. To the authors' knowledge, apart from the work in [14] and its extension proposed here, [12] and [13] are the first studies addressing the problem of matching deformable folded shapes to unfolded templates. However, both in [12], and [13], the unfolded template corresponds to the folded garment's spread-out configuration. In [14], even though the method is not assuming that the unfolded configuration is known beforehand, a significant reduction in performance is reported when a different garment is used for defining the unfolded template. Thus, the extended approach proposed here can be considered the first one fully addressing the problem of fitting generic unfolded templates to folded contours, whose exact unfolded configuration is unknown. Moreover, in [12], only pieces of fabric are considered, whereas in [13], where real garments are used, the location of the fold on the template is approximately known (since a deterministic folding procedure is applied) and a model of the unfolded portion of the garment is available.

This paper is organized as follows. In Sect. 2, the proposed method performing matching of folded shapes and fitting the unfolded templates is presented. In Sect. 3, the adopted modelling approach is described and a re-grasping strategy is proposed for unfolding. In Sect. 4, the experimental results are demonstrated, whereas in Sect. 5 a short discussion on the method and the produced results is presented. The paper concludes in Sect. 6.

2 Matching folded garments to unfolded templates

The proposed approach operates in two subsequent stages. The first one is called the matching stage, where the contour of the folded garment is matched to a reference template. The matching method used in this stage is adopted by the work in [14]. For completeness it is also presented here in more detail, including useful notations for linking it to the second stage which is called the fitting stage. In this stage, the

output of the matching stage is used for updating the reference template to better fit the folded contour. The proposed method can be applied recursively, using the updated template as a new input for the matching stage.

2.1 Matching stage

At this stage, an RGB image of a folded garment lying on a table is acquired. Color edge detection is applied on the image and the resulted edge image is thresholded. A binary mask corresponding to the largest object is selected, after filling existing holes. The mask's contour is extracted and approximated by a polygon using Douglas–Peucker algorithm for line simplification [15]. Then, each side of the simplified polygon is examined as potential folding axis. After removing the examined side from the folded contour the method proposed in Riemenschneider et al. [16] is used to detect partial matches between the resulting open contour and the unfolded template. Each partial match is used for estimating an affine transformation, which is applied to the examined side generating a hypothesis about the folding axis location on the template. To test the generated hypothesis, the template is virtually folded over the transformed axis and the resulting polygon is matched to the folded garment using Inner Distance Shape Contexts (IDSC) algorithm [17]. Based on the hypothesis producing the best match, point correspondences between the matched polygons are established, whereas the location of the folding axis on the template is estimated. An overview of the aforementioned procedures is presented in the block diagram of Fig. 1.

2.1.1 Partial matching

Matching the folded garment's contour $FC_{\text{garm}} \in \mathbb{R}^{2 \times M}$ to the template's unfolded contour $UC_{\text{temp}} \in \mathbb{R}^{2 \times N}$ is an extremely challenging task, especially if generic templates are employed. After folding, the original contour can be fragmented over several pieces, whereas some fragments can be missing due to overlaps. Moreover, fragments belonging to different parts of the original contour can be wrongly connected. To tackle the above challenges, a partial matching method has been adopted. An additional advantage of using an approach based on partial matches is that it makes the method more robust with respect to local deformations of the garments. In this study, the partial matching method proposed in Riemenschneider et al. [16] has been employed. In [16], object category localization is performed by partially matching edge contours to a single shape prototype of the category. The aforementioned method has been selected, since object category localization presents similar challenges to those of folded contour matching.

The adopted matching technique is translation and rotation invariant [16]. In this work, a novel strategy for extracting

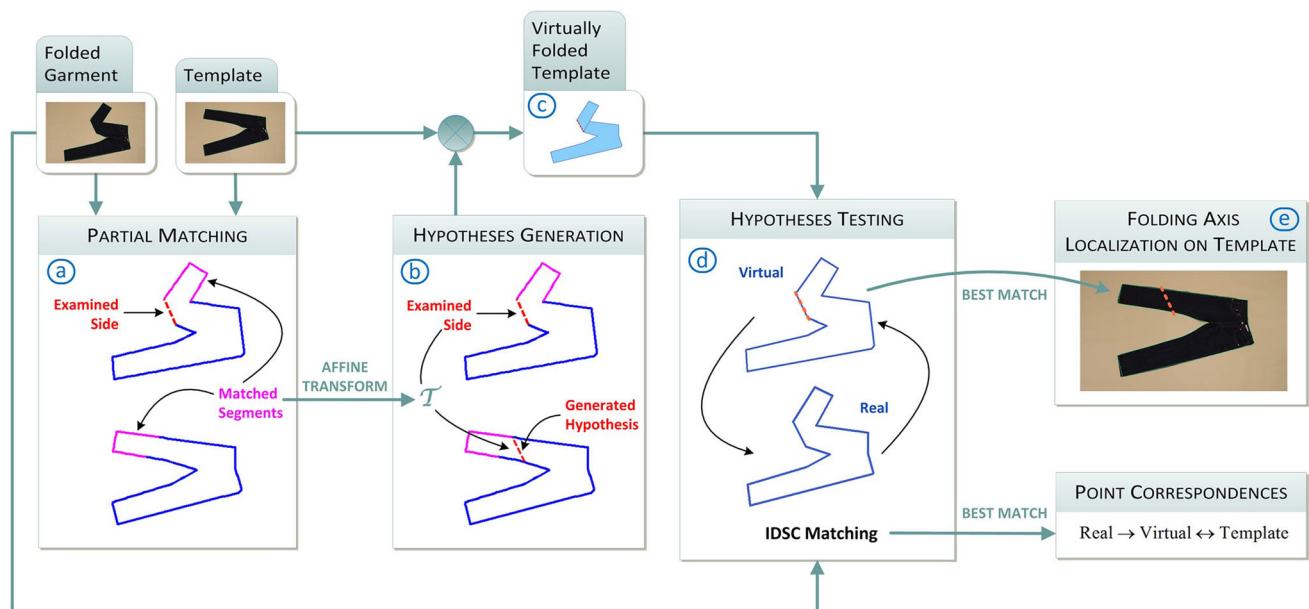


Fig. 1 Block diagram of the proposed method performing shape matching of folded garments and unfolded templates

the descriptors is proposed achieving reflection invariance, as well. This is a very useful property in case of folding, since some segments of the folded contour may represent reflections of segments belonging to the unfolded contour. According to the proposed strategy, an initial contour orientation is selected for both UC_{temp} and FC_{garm} and the descriptors are extracted, whereas a second set of descriptors is extracted by FC_{garm} after inverting its orientation. The proposed strategy guarantees that the same sets of descriptors will be extracted even if a reflected version of the contour is examined (e.g. in case folding is performed in the opposite direction). Using the extracted descriptors, partial matching is performed twice. The first time the initial set of descriptors is employed, matching segments of the contours. The second time the other set of descriptors is employed, matching segments of UC_{temp} to reflected segments of FC_{garm} . Notice, however, that due to the unknown front–back configuration and possible symmetries of the templates the method is in fact agnostic to whether the reflected segments belong to folded or unfolded parts of the garment and vice-versa. Each partial match is assigned with an affinity score As , which is efficiently calculated using integral images optimization [18, 19], and after thresholding only strong matches remain. Then, similar overlapping contours are merged and the remaining matches are aggregated using the technique described in [16].

The described matching procedure is repeated for every side of FC_{garm} . In each repetition, UC_{temp} is matched to the open contour that results after removing from the FC_{garm} the examined side. In Fig. 1a, the simplified contours of a folded pair of pants and an unfolded template are illustrated. The examined side is denoted by a red dashed line, whereas

matched segments are denoted by magenta solid lines. The proposed method examines all sides and produces a large number of matches, which are not always correct. Each match is used to generate a hypothesis about the location of the folding axis on the template and the validity of the hypothesis is tested.

2.1.2 Hypothesis generation and testing

Each partial match is used to estimate a local affine transformation, which is then applied to the examined side. The new location of the transformed side generates a hypothesis about the position and orientation of the folding axis, with respect to the employed template. In Fig. 1b, an example of a generated hypothesis is illustrated. The red dashed line on the template corresponds to the generated hypothesis about the folding axis location. In this example, the presented hypothesis is valid.

Each generated hypothesis is assigned a confidence score Cs , which is computed as a product of the affinity score As of the match used for generating the hypothesis and Rs score, which is given by:

$$Rs = \begin{cases} e^{-\frac{L_{out}}{L_{in}}}, & \text{if } L_{in} \neq 0 \\ 0, & \text{if } L_{in} = 0 \end{cases} \quad (1)$$

where L_{out} denotes the axis length that is predicted to be outside the template and L_{in} denotes the axis length that is predicted to be inside the template. Both As and Rs take values in the $[0, 1)$ interval. The higher their values, the greater

is the confidence on the validity of a certain hypothesis. In case $R_s = 0$, the hypothesis is automatically rejected.

A significant advantage of using partial matches to generate the hypotheses is that insensitivity to local deformations and segmentation noise is introduced, provided that erroneous hypotheses are discarded during testing. Namely, in case of local deformations, the position and orientation of the predicted folding axis will be absurd, resulting to erroneous hypotheses. This is compensated by the fact that non-deformed parts of the contour can produce valid hypotheses that will be selected over the erroneous ones. However, different partial matches can result in very similar hypotheses, in case the matches are correct. Therefore, an approach similar to the one used for merging the matched segments is proposed for aggregating the surviving hypotheses. Aggregation is based on a binary image, where each pixel represents a hypothesis. The horizontal image axis corresponds to angles, the vertical axis corresponds to distances and each pixel represents the folding axis of each hypothesis in polar form. Similar hypotheses are expected to create connected regions in this binary image, which can be extracted by connected component analysis. After extracting a component, the associated confidence scores are summed and the result is thresholded. The hypotheses of each component that survives thresholding are merged by calculating a weighted average of the corresponding folding axes. The assigned C_s scores are used as weights for the calculation. By thresholding and aggregating, only a small number of hypotheses proceed to the final testing stage, which is the most computationally intensive. Each hypothesis determines a folding axis on UC_{temp} . The predicted axis is used to virtually fold the template and a predicted folded contour $VFC_{temp} \in \mathbb{R}^{2 \times M}$ is generated (Fig. 1c). VFC_{temp} is matched to FC_{garm} using IDSC and a matching cost is estimated (Fig. 1d). IDSC, which is a global shape matching technique, has been selected due to its insensitivity to articulations [17]. The hypothesis resulting to the minimum matching cost is selected as the most probable one. The selected hypothesis is accepted only if the associated matching cost is lower than the one produced when the unfolded template is employed. The IDSC correspondences $PC \in \mathbb{N}^{2 \times M}$ between VFC_{temp} and FC_{garm} contour of the selected hypothesis are used for matching the folded garment to the virtually folded template. The selected hypothesis provides also $F_{axis} \in \mathbb{R}^{2 \times 2}$, which indicates the location of the folding axis on the template.

In case multiple templates belonging to different garment types are employed, the selected hypothesis can be also used to predict the actual type of the folded garment. Thus, the garments can be recognized while they are folded and the estimated configuration can be exploited for planning the unfolding strategy.

Classification and matching results depend on the similarity of the reference template to the garment before folding.

Therefore, to improve performance, a novel method for fitting the generic template to better match the folded garment is proposed.

2.2 Fitting stage

The reference template is a generic contour representing an entire class of unfolded garments. Typical classes, that are also considered in this work, are shirts, T-shirts, pants, shorts, skirts, and towels. Due to the large intra-class variation of real garments' size, shape, and style, a large number of reference templates are needed for each class, unless the original template is modified to better fit the specific folded garment. A better fit of the original template to the folded garment can be estimated by employing the point correspondences between the virtual and actual folded contours. At first, the virtually folded contour is registered to the actual contour using a collection of similarity transformations [20] estimated in a piecewise fashion. Then, the folding axis estimation is used to virtually unfold the registered contour. The original template is registered to the virtually unfolded contour, using weighted local similarity transformation. Straight line segments between points of the simplified contour are estimated using the RANdom SAMple Consensus (RANSAC) method [21], whereas their intersections are used to define the contour points of the fitted template. The updated template is used as new input for the matching stage. An overview of the fitting method is depicted in the block diagram of Fig. 2.

2.2.1 Registering folded contours

At first, a better fit of the template to the visible contour parts of the garment is estimated, by registering the template's virtually folded contour to the actual contour of the folded garment. To perform the registration, the point correspondences PC that were established in the matching stage between VFC_{temp} and FC_{garm} are used to estimate for each contour point i a non-reflective similarity transformation $T_s^i \in \mathbb{R}^{3 \times 3}$, which is then applied to $VFC_{temp}(i)$. Namely, for each point $VFC_{temp}(i)$, a small neighbourhood of Np^s adjacent points is selected in both VFC_{temp} and FC_{garm} and a non-reflective similarity transformation T_s^i is estimated. The neighbourhood selection is such that the contour point of interest (point i) lies always in its centre. The estimation of T_s^i is based on Least Squares minimization of the Euclidean distance between the points of FC_{garm} and the transformed points of VFC_{temp} that belong to the selected neighbourhood. VFC_{temp} is registered to FC_{garm} by applying the corresponding local transformation to each point, resulting to a registered contour $RVFC_{temp} \in \mathbb{R}^{2 \times M}$.

Non-reflective similarity is a transformation that preserves shapes and angles. It may include a rotation, a translation, and a scaling and it has four degrees of freedom. The mathemat-

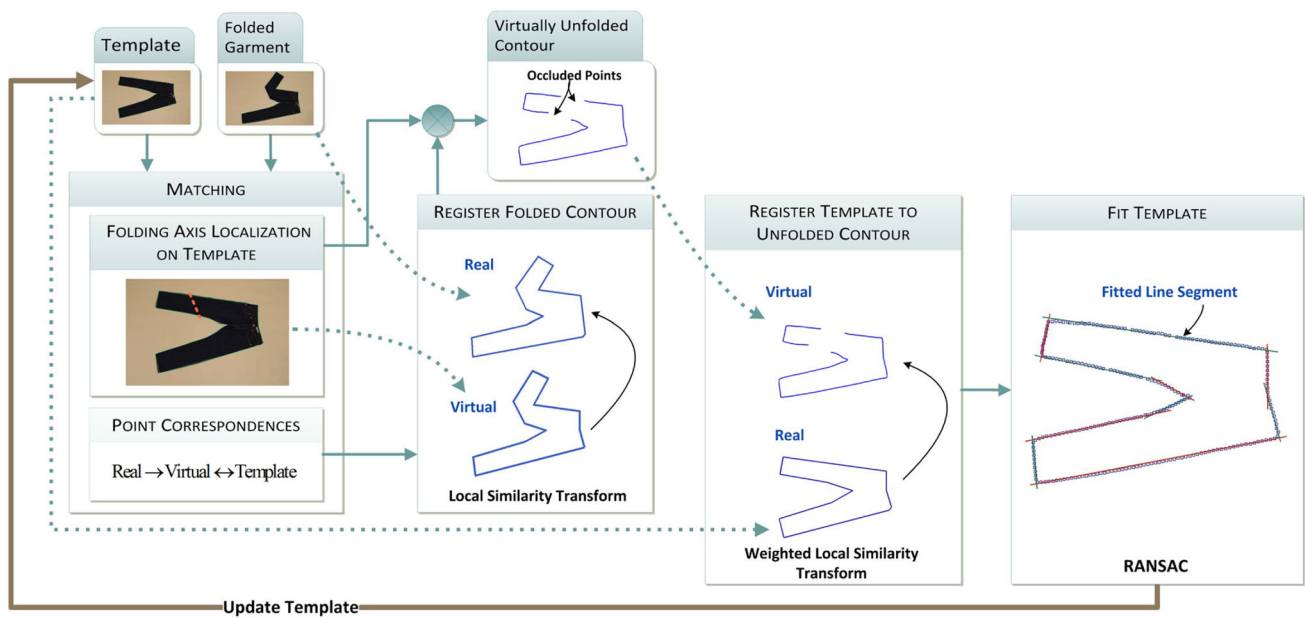


Fig. 2 Block diagram of the proposed method performing updating of the unfolded template

ical model of the transformation is given in (2), where s , ϕ , and (t_x, t_y) are scaling, rotational and translational differences between the points, whereas (x, y) and (u, v) are the original and transformed coordinates, respectively.

$$\begin{aligned} u &= s(x \cos \phi - y \sin \phi) + t_x \\ v &= s(x \sin \phi + y \cos \phi) + t_y. \end{aligned} \quad (2)$$

With the exception of the scaling factor, similarity transformation is basically a rigid transformation. However, using adequately small neighbourhoods, deformations due to the elastic properties of the garments can be also fitted. Moreover, the adopted piecewise approach allows the use of different scaling factors for different parts of the contour. This facilitates the correct alignment of the contours when the shape proportions of the generic template significantly differ from those of the actual garment.

Since the folding axis is part of the virtually folded contour, the estimated similarity transformations are also applied to the axis itself. Thus, a new folding axis $\text{RFAxis} \in \mathbb{R}^{2 \times 2}$ is computed by applying the corresponding local transformations to FAxis . Then, by reflecting its points over RFAxis , $\text{RVFC}_{\text{temp}}$ is virtually unfolded producing $\text{VUC}_{\text{temp}} \in \mathbb{R}^{2 \times N}$. However, due to self-occlusion, the unfolded contour is not complete, presenting multiple gaps. In VUC_{temp} , these gaps are padded with NaNs. To fill these gaps, UC_{temp} is registered to VUC_{temp} .

2.2.2 Registering the template to the virtually unfolded contour

Each point j of the unfolded template UC_{temp} is registered to the corresponding point of the virtually unfolded contour

VUC_{temp} , using a weighted local similarity transformation $T_{\text{ws}}^j \in \mathbb{R}^{3 \times 3}$. T_{ws}^j is estimated in a similar manner to that of T_s^i , with the main difference that the closer the points to the neighbourhood centre the higher their contribution to the estimation. The reason for using different weights for each point is that, as shown in Fig. 2, there are some missing points in the virtually unfolded contour, which correspond to self-occluded parts of the folded garment. Therefore, larger neighbourhoods of contour points are employed, which sometimes can contain a lot of NaN values. A Gaussian kernel with experimentally defined parameters is used for generating the corresponding weights for each transformation estimation. The neighbourhood's centre is used for centring the employed kernel, as well.

After applying T_{ws}^j to all points of UC_{temp} , the registered contour of the unfolded template $\text{RUC}_{\text{temp}} \in \mathbb{R}^{2 \times N}$ is produced, which better fits the garment's shape. However, due to the missing contour points of VUC_{temp} , some outlier points may be present in RUC_{temp} . To discard these outliers, a smoothing technique based on RANSAC is applied.

2.2.3 Updating the template using RANSAC

The goal is to fit each side of RUC_{temp} by straight line segments using a suitable linear regression model. However, due to the missing contour information in VUC_{temp} , some outliers may be present in RUC_{temp} . Therefore, the RANSAC method is used to perform the fitting without employing those outliers. Namely, using the vertices of the simplified contour of RUC_{temp} , the contour points can be grouped to C different sets, corresponding to the sides of the simplified contour.

Then, a straight line segment LS^i is fitted to each set S^i , after removing the set's outliers using RANSAC. The C intersections between adjacent line segments (LS^i, LS^{i+1}) are used for defining the contour points of the updated template $UC_{fit} \in \mathbb{R}^{2 \times C}$.

UC_{fit} can be re-sampled and used as a new input for the matching stage. Thus, the proposed matching-fitting procedure can be iterated n times (5 in our experiments) and the template resulting to the best match can be finally selected.

3 Modelling folded garment to facilitate unfolding

Once matching and fitting are performed, a polygon model of the folded garment is estimated using the derived correspondences. The produced model can be used for identifying point pairs on the folded contour that are accessible for grasping and suitable for bringing the garment to an unfolded configuration.

3.1 The simplified polygon model

Using the unfolded template's simplified contour, a polygon model can be derived, denoted henceforth as the simplified polygon model (SPM). SPM is based on contour points and presents no interior structure. Similarly to the polygon model introduced in [13], SPM is parameterized about a set L of landmark points l that correspond to the m corners of the original template's simplified contour, as described by (3).

$$L = [l_1(x) \ l_1(y) \ \dots \ l_m(x) \ l_m(y)] \quad (3)$$

SPM differs from the polygon model introduced in [13], only to the selection of each garment's landmarks points. The selected landmark points in SPM case are depicted as red dots on the unfolded templates of Fig. 3. In case of pants

and towel, SMP and the polygon model in [13] coincide. Using the point correspondences between UC_{temp} and UC_{fit} , an SPM can be fitted to UC_{fit} . Let the fitted model be denoted by L_0 . Following a similar notation to the one used in [13] for the parameterized shape models, L_0 can be augmented to model the virtually folded template VFC_{fit} , using the estimated folding axis location F_{axis} on VFC_{fit} . The augmented SPM is described by (4).

$$L_{folded} = [F_{axis} \mid L_0] \quad (4)$$

where F_{axis} specifies the line about which the polygon defined by L_0 is to be folded.

3.2 Locating re-grasping points for unfolding

Once a model of the folded garment is estimated, suitable re-grasping points for unfolding can be identified. These points correspond to landmark point pairs that have been manually defined as candidate re-grasping points for unfolding on the unfolded templates. The selected pairs are sorted according to preference order, preferring those that result to the most natural unfolded configuration of the garment (e.g. for pants they may be the two points on the waist). Using the augmented SPM model of the folded garment the suitable landmark point pairs are identified on the garment's contour, and their accessibility is examined. Namely, the overlap between the model's folded layers is computed and only landmark points with no overlap within a predefined radius around them are labelled as accessible. The list of the sorted landmark pairs is updated by dismissing pairs containing inaccessible points. Using the correspondences between the model and the actual contour, the location of the top pair is identified on the folded contour, and it can be used for planning the re-grasping motion for unfolding.



Fig. 3 Top synthetic templates of unfolded garments belonging to six different garment types. Landmark points used for producing the SMPs are denoted by red dots on the templates' contours. Bottom images of

real garments used in the experiments. An unfolded configuration of the garments is presented here for comparison with the employed templates



Fig. 4 Example images of the folded garments in the real dataset

4 Experimental evaluation

In previous work [14], the effectiveness of the matching stage has been evaluated, using a synthetic dataset consisting of virtually folded templates. The results reported adequate accuracy of folding axis localization achieving on average a correct localization rate of 97 % for five different garment types. Similar results have been produced in case of real garments, achieving on average a correct localization rate of 96.7 % when each garment was also used as a template. However, as expected, when different garments were employed as templates, the correct localization rate dropped to 88.5 %.

4.1 Real dataset and synthetic templates

In this study, the image database of folded garments used in [14] has also been employed. The database¹ consists of 54

images of folded garments belonging to six different types, i.e. pants, shorts, shirt, T-shirt, skirt and towel. To test the effectiveness of the proposed approach in case previously unseen templates are used, six synthetic templates have been manually constructed. As illustrated in Fig. 3, the templates, depicted in the top row of the image, correspond to the six different garment types used in the database. In the other two rows of the image the garments used in the database are illustrated before folding. In Fig. 4, example images of the employed database are also presented, depicting the selected garments in different folded configurations. A qualitative comparison of the similarity between the employed generic templates and the garments used for testing can be performed by visual inspection of Fig. 3. In addition, to allow for a quantitative comparison, the mean Euclidean distance in centimeters between the contour points of the test gar-

¹ The employed image database, including manual ground truth annotation, is publicly available for testing and benchmarking. For more

Footnote 1 continued
information on how to acquire the database, please contact the corresponding author via email.

Table 1 Folding axis localization results for 12 garments, using 54 folding scenarios

Type	EucDist (cm)	CLRm (%)	CLRf (%)	dA (%)	dθ (°)	NMM	MCD (%)
Pants	2.6/4.7	100	100	1.5 ± 1.1	2.8 ± 2	0	11
Shorts	21.3/23.1	77.8	77.8	2.9 ± 2.3	1.6 ± 1.4	2	23.2
Shirt	8.5/7.4	70	90	3.9 ± 2.5	4.2 ± 3.8	0	20.9
T-shirt	9.4/20.2	66.7	77.8	4.8 ± 3	2.8 ± 3	2	29.3
Skirt	8.4/4.6	88.9	100	4.8 ± 3.2	3.6 ± 1.9	0	10.6
Towel	2.5/2.3	42.9	100	3.7 ± 3.7	3.1 ± 2.6	0	30.6
Overall	9.58	75.9	90.7	3.6 ± 2.6	3 ± 2.5	4	20.4

ments and the generic templates (aligned using IDSC matching correspondences and similarity registration) before folding is also provided in the second column (EucDist in cm) of Table 1. Since two test garments have been used for each type, both distances are provided separated by “/”. On the left, the distances for the garments illustrated in the second row of Fig. 3 are presented, whereas on the right the distances for the garments presented in the third row of Fig. 3 are presented. The towel template is the most similar to the test garments, whereas the shorts template is the most dissimilar one. Mean EucDist averaged over all templates and test garments is about 9.6 cm.

Table 1 also summarizes the method’s evaluation results. Each line corresponds to different garment type, whereas the last line corresponds to averages over all garments. In Table 1, CLRM column displays the correct localization rate (CLR) achieved after matching the folded garments to the original templates, whereas CLRf displays the correct localization rate achieved after matching the garments to the fitted template that produced the best match after five iterations of the method. It is clear that fitting the original templates improves correct localization rate with an overall increase from 75.9 to 90.7 %. The garment type that mostly benefits from updating the original template is “Towel” presenting an increase in CLR from 42.9 to 100 %. Although towels have a simple contour, the ratio between their sides varies significantly affecting severely the localization results. On the other hand, pants and shorts, which do not seem to benefit from the fitting, do not exhibit large intra-class shape variation.

In this study, a novel measure has been adopted for evaluating the accuracy of the folding axis localization. Namely, the convex hull defined by the end points of the estimated and the actual folding axes is calculated and its area is measured. The ratio dA of this area with respect to the unfolded garment’s total area provides an informative measure of the localization accuracy. In case of perfect localization dA is zero, whereas it can never get greater than one. A threshold of 0.1 has been selected for deciding which localization results are acceptable. Another measure that was used is the angle dθ formed between the estimated and the actual folding axes. The results are illustrated in the dA, dθ columns of Table 1, respectively. In the NMM column the number of mismatches caused by

verifying an erroneous hypothesis is shown for each garment type. Only shorts and T-shirts presented mismatches, with four total mismatches for the 54 examined scenarios. In the MCD column of Table 1, the decrease % of the final matching cost between the folded garment and the virtually folded template is presented. The decrease is calculated by (5), where MCm denotes the Match Cost between the folded garment and the virtually folded original template and MCf denotes the Match Cost between the folded garment and the virtually folded updated template that has been finally selected. An overall decrease of the matching cost of about 20 % has been achieved on average.

$$\text{MCD} = \frac{\text{MCm} - \text{MCf}}{\text{MCm}} \times 100 \% \quad (5)$$

In Fig. 5, an example of template fitting is presented illustrating the results of the first two iterations. In Fig. 5a, a folded pair of shorts is depicted, whereas in Fig. 5b the employed template and the estimated result of the matching stage are illustrated. Using this result, the initial template is updated and the contour of the new template is given in Fig. 5c by a blue dotted line. In Fig. 5d, matching is repeated using the updated template. The estimated folding axis is denoted by a red dashed line. Fitting is reapplied and the template is updated for a second time as demonstrated in Fig. 5e, whereas the new matching result is presented in Fig. 5f.

In case multiple templates of different garment types were employed, the method was always able to discriminate between them based on the template that generates the hypothesis with the minimum matching cost. Thus, the folded garments were correctly classified in all 54 cases.

To evaluate how well the produced polygon models fitted the folded garments, the error distance between manually annotated landmark points on the folded garments’ outline and the corresponding estimated landmark points of the model has been calculated. The results are presented in Table 2 averaged over each garment type, whereas in the last row the average is computed over all cases. The error distance is given both in pixels and centimeters, whereas the average standard deviations of the distances are also provided. Only cases with correct axis localization have been

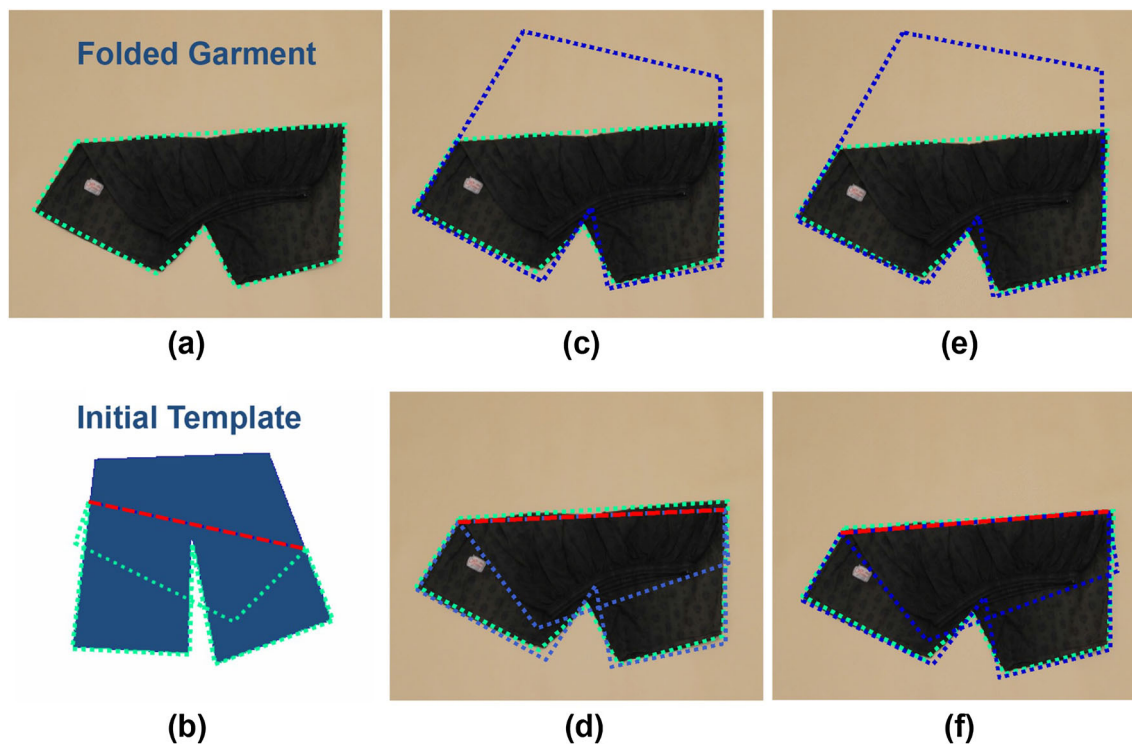


Fig. 5 Example images of template fitting for a folded pair of shorts

Table 2 Model fitting results

Type	Landmark point error distance	
	Pixels	cm
Pants	2.59 ± 1.46	0.65 ± 0.36
Shorts	2.82 ± 2.08	0.71 ± 0.52
Shirt	3.87 ± 4.4	0.96 ± 1.1
T-shirt	6.13 ± 8.32	1.53 ± 2.08
Skirt	3.55 ± 5.61	0.88 ± 1.4
Towel	3.2 ± 2.06	0.8 ± 0.51
Overall	3.64 ± 4.58	0.91 ± 1.14

examined, which corresponds to 90.7 % of our dataset (see Table 1). The reported errors span an acceptable range for locating suitable grasping points for unfolding, with an overall average error of 0.91 ± 1.14 cm. In about 92 % of the examined folding scenarios, all accessible point pairs have been correctly identified on the garment's contour within a margin of 5 cm (20 pixels) error distance. The reported accuracy is considered acceptable for re-grasping the garment from these points to unfold it. However, actual unfolding using a robotic manipulator will be addressed in future work.

5 Discussion

Modelling articles of clothing is a challenging task, largely due to their non-rigid nature. Especially in the case of not

fully unfolded garments, additional challenges are introduced because of self-occlusion. To cope with these challenges, the presented approach for modelling folded garments is proposed, since it has a series of attractive properties. It presents insensitivity to deformations, articulations, and self-occlusion by performing matching in both local and global scale using robust shape analysis techniques. It uses simple primitives that are invariant to rotation, translation, and reflection of the garment. It is not affected by the garments' variation in texture and colour, since it relies solely on contour information. It discriminates between different garment types, while the garments are still folded. Unlike previous works [12, 13], which employ the contour of the garment before folding, it allows the use of generic templates that are not necessarily very similar to the folded garments. Finally, since a geometric approach has been adopted, no training is required.

However, the proposed approach also presents certain limitations. The presented method deals only with approximately planar configurations. It is not invariant to the sampling distance used to extract the contour points. It cannot detect very small folds, or folds that are parallel and close to the sides of the garment, changing only the proportions of its shape. However, very small folds may be acceptable for fitting an unfolded model and re-grasping, whereas parallel folds is not very probable to arise by chance. Fitting is expected to improve folding axis localization accuracy and model estimation accuracy only in case a reasonable error

has been produced in the initial folding axis estimation. Otherwise, errors will propagate to the fitting stage as well. A straightforward modification for addressing this kind of failures is to use a larger (but still small) set of generic templates for each garment type. Another limitation is that the method is oblivious to the actual configuration of the garment part that lies inside the folded contour, although an estimate on this configuration is provided by the augmented model. It is also agnostic with respect to which part of the folded garment is on top. This has also been the case in [12]. However, in that study, a trial to grasp the candidate corner is made and in case of failure the process is repeated. In our case, since a complete model is extracted suitable outline points are selected for grasping to bring the garment in an unfolded configuration. We plan to address most of these challenges in future work.

Once the matching and fitting results of the method are produced, the proposed simplified polygon model can be derived in a straightforward fashion. However, the proposed method's results allow the extraction of more elaborative models, e.g. the parameterized shape model proposed in [13], which includes skeletal points and legality constrains. Such a parameterized model can be fitted to the updated template, whereas the estimated folding axis can be used to produce an augmented folded model as the one described in [13]. In that work, which addresses modelling of garments during folding the employed skeletal model produced better results than the polygon model that was used as baseline. However, whether that parameterized approach using skeletal points should also be preferred in case of garment unfolding remains to be investigated.

6 Conclusion and future work

In this work, a novel method modelling folded garments by matching them to unfolded templates has been proposed. The method is based on the simple but realistic assumption that after folding, the axis becomes part of the outer contour of the garment. Thus, each side of the folded contour is examined as potential folding axis. Using unfolded templates, hypotheses can be generated about the position and orientation of the axis with respect to the templates. The hypotheses are tested by virtually folding the templates and matching them to the folded garment. The hypothesis resulting to the best match is selected. However, matching results strongly depend on the similarity of the template to the garment shape before folding. Therefore, the initial matching results are used to update the template in an iterative fashion yielding a better fit.

To the authors' knowledge, the proposed method is the first one addressing the problem of matching folded shapes, whose unfolded contour is unknown, to generic unfolded

templates of different types. In case of folded garments, additional challenges are introduced due to their deformable properties and the large variation they present in shape and style. Thus, a template fitting approach has been proposed to address these challenges. The reported experimental results demonstrate a significant increase in the correct localization rate of the folding axis when the fitted templates were employed compared with the rate produced by the original templates. Namely, an overall increase from 75.9 to 90.7 % has been reported indicating the usefulness of the fitting stage. Using the matching results of the fitted template and the folded garment, a polygon model can be estimated and used for identifying suitable re-grasping points on the folded garment's contour for bringing it to an unfolded configuration. Experimental results reported a mean accuracy of about 1 cm in the estimated position of the model's landmark points on the folded contour, indicating the applicability of the proposed approach.

In future work, a larger dataset including garments with multiple folds will be acquired, and the method's performance will be evaluated. Moreover, the method will be integrated into a complete pipeline for autonomous robotic unfolding of real garments. Namely, the garment will be picked up from a random point and autonomously hanged by two outline points, to bring it into a planar but folded configuration. It will then be placed on a flat surface and the proposed method will be used to model it and identify re-grasping points suitable for unfolding. In this case, it will be investigated, whether the method can produce acceptable results despite the deformations introduced by hanging the garment before placing it flat. Another approach that could be investigated, is to use the model's estimate on the garment's configuration for further visual inspection and active exploration of the areas inside the folded contour. In that case, unfolding without lifting the bottom layer of the garment from the table could be considered.

References

1. Osawa, F., Seki, H., Kamiya, Y.: Unfolding of Massive Laundry and Classification Types. *J. of Adv. Comp. Intell. and Intell. Inf.* **11**, 457–463 (2007)
2. Hamajima, K., Kakikura, M.: Planning strategy for task of unfolding clothes. *Rob. and Aut. Syst.* **32**, 145–152 (2000)
3. Maitin-Shepard, J., Cusumano-Towner, M., Lei, J., Abbeel, P.: Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In: *ICRA*, pp. 2308–2315 (2010)
4. Cusumano-Towner, M., Singh, A., Miller, S., O'Brien, J.F., Abbeel, P.: Bringing clothing into desired configurations with limited perception. In: *ICRA*, pp. 3893–3900 (2011)
5. Bersch, C., Pitzer, B., Kammel, S.: Bimanual robotic cloth manipulation for laundry folding. In: *IROS*, pp. 1413–1419 (2011)

6. Willimon, B., Birchfield, S., Walker, I.D.: Model for unfolding laundry using interactive perception. In: IROS, pp. 4871–4876. IEEE Press, New York (2011)
7. Willimon, B., Birchfield, S., Walker, I.D.: Interactive perception of rigid and non-rigid objects. *Int. J. Adv. Robotic Syst.* **9**, 227–238 (2012)
8. Willimon, B., Hickson, S., Walker, I., Birchfield, S.: An energy minimization approach to 3D non-rigid deformable surface estimation using RGBD data. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, pp. 2711–2717 (2012)
9. Willimon, B., Birchfield, S., Walker, I.D.: 3D non-rigid deformable surface estimation without feature correspondence. In: Proceedings of the IEEE International Conference on Robotics and Automation, Karlsruhe, pp. 638–643 (2013)
10. Willimon, B., Walker, I.D., Birchfield, S.: Classification of clothing using mid-level layers. In: *ISRN Robotics*, pp. 1–17 (2013)
11. Triantafyllou, D., Aspragathos, N.A.: A vision system for the unfolding of highly non-rigid objects on a table by one manipulator. In: *ICIRA*, pp. 509–519 (2011)
12. Ono, E., Kita, N., Sakane, S.: Unfolding folded fabric using outline information with vision and touch sensors. *J. Robotics Mech.* **10**(3), 235–243 (1998)
13. Miller S., Fritz M., Darrell T., Abbeel P.: Parametrized shape models for clothing. In: *ICRA*, pp. 4861–4868 (2011)
14. Mariolis, I., Malassiotis, S.: Matching folded garments to unfolded templates using robust shape analysis techniques. In: *Lecture Notes in Computer Science, LNCS (part 2)*, vol 8048, pp. 193–200 (2013)
15. Hershberger, J., Snoeyink, J.: An $o(n \log n)$ implementation of the Douglas–Peucker algorithm for line simplification. In: Proceedings of the 10th Annual Symposium on Computational Geometry, SCG '94, pp. 383–384. ACM, New York (1994)
16. Riemenschneider, H., Donoser, M., Bischof, H.: Using partial edge contour matches for efficient object category localization. In: Proceedings of the 11th European Conference on Computer Vision, Part V, ECCV'10 (Berlin, Heidelberg), pp. 29–42. Springer, New York (2010)
17. Ling, H., Jacobs, D.W.: Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 286–299 (2007)
18. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), pp. 511–518 (2001)
19. Donoser, M., Riemenschneider, H., Bischof, H.: Efficient partial shape matching of outer contours. In: *ACCV* (2009)
20. Szeliski, R.: *Computer Vision: Algorithms and Applications*, 1st edn, pp. 35–38. Springer, Berlin (2011)
21. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**(6), 381–395 (1981)



Ioannis Mariolis received the Diploma degree and the Ph.D. degree in Electrical and Computer Engineering, in 2002 and 2009, respectively, both from the University of Patras (UoP) Greece. During 2010 and 2011, Dr Mariolis worked as a post-doctoral research assistant in the Medical Physics Laboratory, at School of Medicine (UoP). Since February 2012, Dr Mariolis has been working in the Information Technologies Institute, at the Centre of Research & Technol-



ogy Hellas, as a post-doctoral research fellow. His main research interests include statistical signal processing, machine vision, medical image analysis and pattern recognition.

Sotiris Malassiotis has received the diploma and Ph.D. degrees in Electrical Engineering from the Aristotle University of Thessaloniki, in 1993 and 1998, respectively. From 1994 to 1997 he was conducting research in the Information Processing Laboratory of Aristotle University of Thessaloniki. He is currently a senior researcher (associate professor equivalent) in the Information Technologies Institute, Thessaloniki and leader of the cognitive systems and robotics group. He has participated in more than 15 European and National research projects. He is the author of 30 journal publications and more than 60 papers in international conferences and book chapters. His research interests include 3D image analysis, machine learning, complex and adaptive systems, and computer graphics.