Autonomous Active Recognition and Unfolding of Clothes using Random Decision Forests and Probabilistic Planning

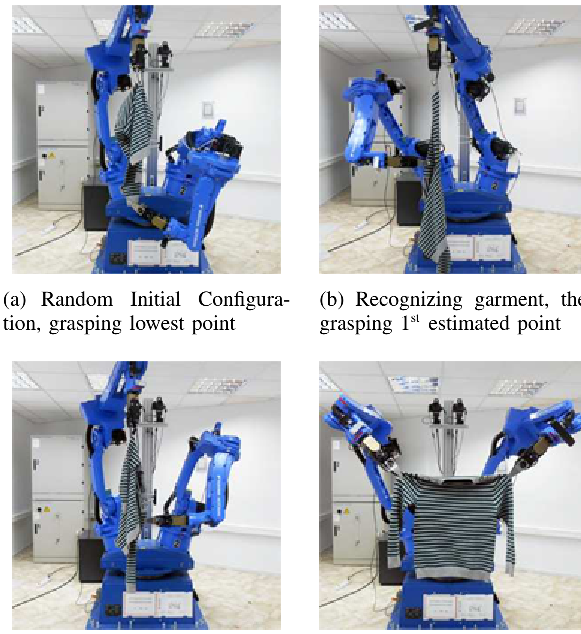
Andreas Doumanoglou, Andreas Kargakos, Tae-Kyun Kim, Sotiris Malassiotis

***Abstract*— We present a novel approach to the problem of autonomously recognizing and unfolding articles of clothing using a dual manipulator. The problem consists of grasping an article from a random point, recognizing it and then bringing it into an unfolded state. We propose a data-driven method for clothes recognition from depth images using Random Decision Forests. We also propose a method for unfolding an article of clothing after estimating and grasping two key-points, using Hough forests. Both methods are implemented into a POMDP framework allowing the robot to interact optimally with the garments, taking into account uncertainty in the recognition and point estimation process. This *active* recognition and unfolding makes our system very robust to noisy observations. Our methods were tested on regular-sized clothes using a dual-arm manipulator and an Xtion depth sensor. We achieved 100% accuracy in active recognition and 93.3% unfolding success rate, while our system operates faster compared to the state of the art.**

I. Introduction

Robots doing the housework have recently attracted the attention of scientists. Our interest is focused in the task of folding clothes and particularly in the first part of the procedure, which is the unfolding of an article of clothing. Starting from a crumbled initial configuration, we want to recognize the article and then bring it into an unfolded state so that it is ready for folding. One of the key challenges in clothes perception and manipulation is handling the variabilities in geometry and appearance. These variabilities are due to the large number of different configurations of a garment, self-occlusions and the wide range of cloth textures and colors.

Research on clothes perception and manipulation started in the middle 90s [1], presenting some first clothes recognition techniques with the help of a dual manipulator. Later, research has been made in garment modelling and feature extraction [2] [3], while only recently scientists were able to completely fold an article of clothing starting from a crumpled initial configuration [4] [5] [2]. The main limitations in the state-of-the-art are a) slow performance and b) difficulty to generalize to a variety of shapes and materials. This stems mainly from the model-driven approaches used and associated simplifying assumptions made. To address these limitations we propose a data-driven approach for clothes recognition and unfolding. We first recognize the type of the article from raw depth data using Random Forests. Based on the recognition result, a pair of key-points are identified such that the article will naturally unfold when held by these two points (Fig. 1). Point estimation is based on Hough Forests, a random forest framework with Hough voting. An active manipulation (perception-action) approach based on POMDPs is also proposed that accounts for uncertainty in



(a) Random Initial Configuration, grasping lowest point

(b) Recognizing garment, then grasping 1^{st} estimated point

(c) Grasping 2^{nd} estimated point

(d) Final unfolded configuration

Fig. 1. Robot unfolding a shirt

the vision tasks and thus leads to superior performance. In summary, our main contributions are:

- An active manipulation procedure for unfolding an unknown item of clothing with a minimal number of moves and only by means of gravity (previous approaches have to go through a flattening phase using a table).
- Fast data-driven machine learning algorithms for robust scale-invariant classification of the garment type and key-points estimation from noisy depth data.
- A probabilistic perception-action framework for optimal action policy accounting for uncertainty.

Compared to the state of the art, our system requires less movements and therefore can operate faster. Furthermore, to our knowledge, this is the first work that autonomously unfolds regular-sized clothes. While most researchers work on small or baby clothes for easier manipulation, regular-sized clothes allow higher degree of deformation and pose more challenges to the recognition and unfolding task.

II. Related Work

The first attempts in clothes manipulation have been made by Hamajima et al. [6] who tried to detect and grasp hemlines aiming to unfold clothes and Kaneko et al. [7] who used basic 2D shape analysis to recognize different types of clothing. Osawa et al. [1] dealt with the recognition task and were the first to introduce iterative grasping of the lowest point of a garment, converging to a finite set of possible configurations, an idea adapted to our work. The classification was based on template matching of the final state of the garment after re-grasping the lowest point several times. While they also unfolded the garment using the same procedure, it can be mostly considered as flattening rather than unfolding.

Later, researchers focused on clothes features and characteristics. Triantafyllou et al. [8] used 2D images to identify different types of corners inside a piece of fabric lying on a table, aiming to unfold it by a robotic arm. Willimon et al. [9] proposed an unfolding method based on clothes features for estimating and moving certain grasping points to gradually flatten a piece of clothing placed on a table. This method required a large number of movements to fully unfold a towel while performance on other types of clothes was not mentioned. The same authors [10] proposed a recognition method using features from two binary silhouettes taken from two vertical viewpoints and re-grasping the garment 10 times to improve accuracy. However the overall procedure became very time consuming. In the final work of the same authors [11], a multi-layer classifier was developed with features extracted from depth images. Results where not very accurate except for the case of considering only shirts, socks and dresses.

Closer to our approach is the work of Kita et al. [12] [13]. It is based on fitting a deformable model to the current pose of a shirt hanging from a random point. Using the model, they are subsequently able to estimate the next grasping point (e.g. the shoulders of a shirt) in order to unfold it. The authors show promising results, that are however limited to a single shirt.

Recently, Maitin-Shepard et al. [5] was the first to complete the whole task of folding a towel using the PR2 robot. The algorithm was based on a corner detector for appropriately grasping the towel, however the amount of time required for completing the task makes the approach intractable in a practical scenario. Bersch et al. [14] also used the PR2 robot to autonomously fold a T-Shirt with fiducial markers, again performing in a prohibited running time. The state of the art in unfolding articles of clothing is the work of Cusumano-Towner et al. [4]. Adopting the technique of sequentially grasping the lowest point they can estimate the resulting state of the garment using a cloth simulator in an HMM framework. After putting the garment on a table, they plan a sequence of grasps in order to bring it to the desired unfolded configuration. The method was applied on small or baby clothes and results were promising. However, applying this method to regular-sized clothes requires large working space (table) and long manipulators attached on a moving base for better results. In contrary, our approach does not require a table for unfolding and no movement of the base of the manipulators is necessary. Also, we have further reduced the number of moves a robot should make in order to unfold an article of clothing.

Fig. 2. Possible lowest points. Gray squares are the symmetric points of the red ones. Arrows show the desired grasping points for unfolding.

Concluding, our work is influenced by two other contributions not related to the laundry problem. The first is the work of Shotton et al. [15] who used random forests for human pose estimation from single depth images. The other work is published by Gall et al. [16] who introduced Hough forests for pedestrian detection, a Random Forest framework with the voting property. In this paper we demonstrate that the above machine learning techniques can produce robust results with challenging data such as various highly deformable clothes.

III. Clothes Recognition

The recognition is based on Random Forests, first introduced by Breiman et al. [17] as a method of classification and regression. They achieve state of the art performance compared to other classifiers like SVM [15] [16] [18] while they provide very fast inference appropriate for real-time applications.

To define our problem, we first assume that an article of clothing has already been isolated from a pile of clothes and the robot grasps it from a random point. The space of possible configurations of a garment hanging randomly is very large, therefore we reduce it by grasping the lowest point [1]. Fig. 2 shows the possible lowest points of four types of clothes considered in this paper: *shirts, trousers, shorts* and *T-shirts*. There is only one possible lowest point for shirts and trousers and two possible lowest points for shorts and T-shirts without counting the symmetric ones. Our classifier is able to distinguish between both garment types and hanging points. Thus, there are six classes defined as $\{Shirt, Trousers, Shorts_1, Shorts_2, Tshirt_1, Tshirt_2\}$ where subscripts $1,2$ indicate the different lowest points.

A set of trees is trained over a database of depth images captured from clothes of all the six classes. Each training sample is a pair $(\mathbf{I}, c)$ where $\mathbf{I}$ is a vector containing the depth image and c is the class of the garment labelled manually. Each tree is trained over a randomly chosen subset of the initial training set. At each node, a set of tests are randomly generated with each test containing the following parameters:

- $C_i, \quad i \in \{1,2\}$: the channel used.
- $\mathbf{V}$: a set of random vectors indicating random positions in the image.
- $f(\mathbf{V}, C_i) > t$: a binary test over the set $\mathbf{V}$ and channel C_i using threshold t

We have used two different channels. Channel C_1 corresponds to the *depth values* as captured from the sensor filtered by a bilateral filter and channel C_2 is the *mean*

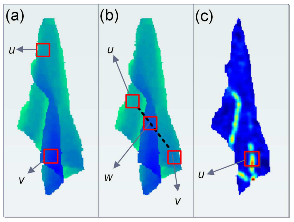


Fig. 3. Binary tests: a) 2 pixel test in depth channel, b) 3 pixel test in depth channel, c) 1 pixel test in curvature channel.

curvature H calculated from the depth data filtered by an average filter. The mean curvature at a point on a surface is defined as:

$$H = \frac{EN + GL - 2FM}{2(EG - F^2)} \tag{1}$$

where E, F, G and L, M, N are the First and Second Order Fundamental Coefficients respectively evaluated on the point [19]. Vectors in $\mathbf{V}$ are normalized to the width and height of the bounding box of the garment for scale invariance so that $v_x, v_y \in [0,1], \mathbf{v} \in \mathbf{V}$.

Three different types of binary tests were used:

- Two pixel test in the depth channel: $\mathbf{V} = \{\mathbf{u}, \mathbf{v}\}$, $f(\mathbf{V}, C_1) = d_u - d_v$, where d_x is the depth value at position $\mathbf{x}$.
- Three pixel test in the depth channel: $\mathbf{V} = \{\mathbf{u}, \mathbf{v}, \mathbf{w}\}$, with $\mathbf{w}$ being a random point on the line between $\mathbf{u}$ and $\mathbf{v}$, $f(\mathbf{V}, C_1) = (d_u - d_w) - (d_w - d_v)$.
- One pixel test in the curvature channel: $\mathbf{V} = \{\mathbf{u}\}$, $f(\mathbf{V}, C_2) = |c_u|$ where $|c_u|$ is the absolute value of the curvature at position $\mathbf{u}$.

Tests are illustrated in Fig. 3. Those simple features are extremely fast to compute and their combination along the path of the trees delivers high discriminative power. Pixel tests are not restricted inside a patch as in [16] and reveal global surface characteristics. At each node a random set of tests is generated and the best is chosen as the one that minimizes the Shannon Entropy of the samples at each child node, which is defined as:

$$H_{entropy} = \sum_{i=1}^{N_c} -\frac{N_i}{N} \ln\left(\frac{N_i}{N}\right) \tag{2}$$

N_c is the number of classes, N_i is the number of samples of class i and N is the total number of samples reached a child node. We declare a node as leaf and stop splitting it further when a minimum number of samples reached the node or the tree has grown to a maximum allowed depth.

Inference about a previously unseen item of clothing is made by traversing its depth image towards the leafs in every tree in the forest, going left or right according to the binary test f assigned to each node. The class of the sample will be the dominant class of the average class distribution of the leaf nodes reached. We should mention that the inference time of a tree in the forest is $O(logD)$ where D is the depth of the tree and therefore is very fast in real-time recognition.

IV. Grasp Point Estimation

Having recognized the garment, the objective is now to grasp it from two certain key-points in order to unfold it. Figure 2 shows the desired grasping points (arrows) for the four types of clothes we used. Our grasp point estimation is based on Hough Forests[16]. The idea is similar to random forests but with an additional property. Each training sample, apart from the label, contains some extra information which is in our case a vector containing the position of one of our desired grasping points. Thus, a training sample is now a triplet $(\mathbf{I}, c, \mathbf{p})$ where $\mathbf{p} = [p_x, p_y]$ is the position of the grasping point on the image $\mathbf{I}$. Coordinates p_x and p_y are normalized to the width and height of the bounding box of the garment for scale invariance so that $p_x, p_y \in [0,1]$. When the grasping point is not visible, $\mathbf{p}$ is undefined and not used. A separate Hough Forest is created for each type of garment, so the classes now become two: $c = 0$ represents images where the grasping point is not visible and $c = 1$ represents images where the grasping point is visible.

The binary tests used are the same as in clothes recognition with the difference that two objective functions should now be minimized for test selection: minimizing the uncertainty about the classes and minimizing the uncertainty about the location of the grasping point of the samples in each node. The uncertainty of the classes is again measured using the Shannon Entropy:

$$H_{entropy} = \sum_{c \in \{0,1\}} -\frac{N_c}{N} \ln \frac{N_c}{N} \tag{3}$$

where N_c is the number of samples of class c and N is the number of samples reached the node. To measure the uncertainty of the location, the following quantity was used:

$$D = \sum_{samples} d(\mathbf{p_s}, \mathbf{p_M}) \tag{4}$$

where d is the Euclidean distance, $\mathbf{p_s}$ is the vector of a sample and $\mathbf{p_M}$ is the average vector of all samples of a node.

While training, leaf nodes store the distribution $P(c)$ of the classes along with a list $\mathbf{L_p}$ of all the location vectors of the samples reached them. When a previously unseen image traverses the Hough forest, the location vectors stored in the leaf nodes will vote for the grasping point location. These votes are accumulated into a *Hough image* and the grasping point location is estimated as the point where the concentration of votes is high (Fig. 4). We estimate this point by picking the maximum of the Hough image after Gaussian filtering. This can also be done using the MeanShift algorithm. The localization of the grasping point only occurs when the class recognized is 1, i.e. the grasping point is visible. After a grasp point is detected, the surface orientation in its vicinity is estimated by locally fitting a plane. This is used to create a valid grasp by moving the gripper perpendicularly to the estimated direction.

Each Hough Forest is trained for a certain garment type in order to localize only one grasping point at a time. When this point is grasped, another Hough forests is used to estimate the second one and complete the unfolding. Therefore, we

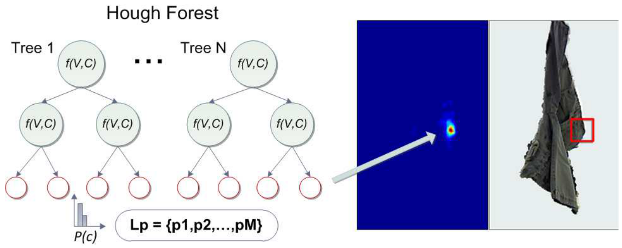


Fig. 4. Hough forest and grasp point estimation from Hough Image

have trained several Hough Forests for every occasion, i.e. for clothes hanging from the lowest or from one desired grasping point. The decision about which Hough forest should be used, is based on the recognition result.

V. Probabilistic Action Planning

Although the one-frame recognition accuracy of our Random Forests classifier is statistically high (Fig. 6(b)) there are some viewpoints of clothes where their type is hardly discernible. In order to eliminate the possibility of erroneous classification, we introduce an active recognition scheme. Furthermore, we want to make our system insensitive to noisy point estimations mainly caused by the noisy depth input, introducing an active point estimation scheme as well. The idea is that the robot will rotate the garment around the gravity axis until the uncertainty about recognition or point estimation is minimized. Instead of exhaustively searching over all viewpoints we employ a probabilistic framework that will select the best action policy jointly minimizing the uncertainty and cost of manipulation.

A widely used probabilistic framework is the Partially Observable Markov Decision Processes (MDP) which are capable of modelling the uncertainty about the current state and can find an optimal policy over the so called *belief state*. While other probabilistic planning approaches have been proposed [20] [21], we have adopted the POMDP framework because having only few states, our problem can be efficiently solved in reasonable time [22] while we take advantage of the representation power and ease of use of the framework. POMDPs have been also used in clothes manipulation by Monso et al. [23] who tried to isolate articles of clothing from a pile.

Fig. 5 shows the block diagram of the complete unfolding process. We have developed two different kinds of POMDPs, one for recognition and one for grasp point estimation described below. If one of those sub-tasks cannot be accomplished, the robot returns to its initial configuration by grasping the lowest point of the garment and thus enters a loop until it becomes unfolded.

A. Active Recognition

Our proposed POMDP is a tuple $(\mathbf{S},\mathbf{A},\mathbf{O},\mathbf{T},\mathbf{P},\mathbf{R},\gamma,\mathbf{b_0})$ where

- $\mathbf{S}$ is the set of states.
- $\mathbf{A}$ is the set of actions.
- $\mathbf{O}$ is the set of observations.
- $\mathbf{T}$ is the conditional transition probabilities.
- $\mathbf{P}$ is the conditional observation probabilities.
- $\mathbf{R}$ is the reward function over the actions and states.
- γ is the discount factor of rewards over time.
- $\mathbf{b_0}$ is the initial belief state.

The states $\mathbf{S}$ in the recognition phase are six and correspond to the six classes used in the classification $\mathbf{S} = \{S_1, S_2.., S_6\}$. The set of actions is $\mathbf{A} = \{A_{rotate}, A_1, A_2, ..., A_6\}$ where A_{rotate} means that the robotic gripper rotates the hanging garment by a degrees to take another observation, while $A_1 - A_6$ is the final recognition decision being at state $S_1 - S_6$ accordingly. The observations are collected from the Random Forests classifier and contain the inferred class c_{in} of the garment along with the probability $P(c_{in})$ from the averaged distribution of the leaf nodes. $P(c_{in})$ takes values in the interval $[0,1]$ but we quantise it into five equally spaced bars for reducing the observation dimensionality. Thus, there are 30 observations, five probability bars for each of the six classes ($\mathbf{O} = \{O_{S_1,P_1}...O_{S_1,P_5}, O_{S_2,P_1}...O_{S_2,P_5}, ..., O_{S_6,P_1}...O_{S_6,P_5}\}$). The transition probabilities taking action A_{rotate} are:

$$T(S_i|A_{rotate}, S_j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \qquad i,j \in \{1,..,6\} \tag{5}$$

All other actions finalize the recognition process and resets the state to its initial configuration:

$$T(S_i|A_j, S_k) = b_0(S_i) \qquad i,j,k \in \{1..6\} \tag{6}$$

Observation probabilities $P(O|S_i)$ are only dependent on the current state and are measured experimentally from previously unseen images. Rewards are assigned in the following way: a positive reward is given to the robot when being at state S_i takes action A_i and a very negative reward when being at state S_i takes action $A_{j \neq i}$. A small negative reward is given to the action A_{rotate} in order to avoid infinite rotation. Regarding the initial probabilities, each type of garment has equal probability to be selected by the robot, however shorts and T-shirts have two possible lowest point. Thus, the initial belief state is $\mathbf{b_0} = (0.25, 0.25, 0.125, 0.125, 0.125, 0.125)$ corresponding to states S_1 - S_6 accordingly. The discount factor γ is set to 0.99. Our objective is to increase the belief about a state before taking a final decision, taking into account the uncertainty about the result of the Random Forest classifier ($P(O|S_i)$).

At the time period t_n, the robot is in state $s \in \mathbf{S}$ and decides to take action $a \in \mathbf{A}$. The next state of the robot will now be s' with probability $T(s'|a,s)$ and will receive the reward $\gamma^n R(s',a)$. After each action of the robot, the belief about each state has to be updated. Let $b(s)$ be the probability of the robot being at state s and o the observation of the robot after taking action a. The belief state will be updated according to the following equation:

$$b'(s') = \frac{P(o|s',a) \sum_{s \in S} T(s'|s,a) b(s)}{\sum_{s' \in S} P(o|s',a) \sum_{s \in S} T(s'|s,a) b(s)} \tag{7}$$

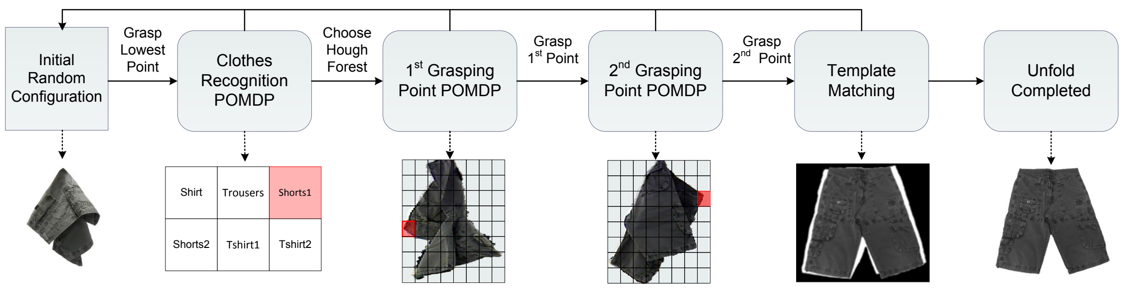


Fig. 5. Block diagram of the unfolding procedure

where s' is the next state of the robot, $b'(s')$ is the new belief state over the states s' and $P(o|s',a)$ is the probability of receiving observation o after taking action a and arriving at state s'. The denominator is a normalization factor.

Solving the above POMDP generates an optimal action policy for the robot. The belief state is updated after each observation of the classifier using (7) and the robot decides according to the policy whether to further rotate the garment to collect more observations or take a final decision and continue the unfolding process. In case the garment is rotated more than 360 degrees, the process is restarted by re-grasping the lowest point. As we will see (Table I), this active recognition dramatically increases the recognition accuracy with the cost of only a few rotations.

B. Active Grasp Point Estimation

The same idea is applied to the grasp point estimation procedure. The states now correspond to the different grasp point locations with an extra state indicating the *invisible* grasp point. Again, we quantize the image space to lower the problem dimensionality applying a 8x8 grid on the bounding box of the garment (Fig. 5). Thus, the set of states is $\mathbf{S} = \{S_0, S_1, ..., S_{64}\}$ where S_0 is the invisible-grasp point state and S_1 - S_{64} correspond to the location of the grasp point on the 8x8 grid. The set of actions is similarly defined as $\mathbf{A} = \{A_{rotate}, A_1, A_2 ..., A_{64}\}$ where A_{rotate} should be taken when grasp point is invisible and A_i is the action of grasping the estimated point located on the ith grid square (state S_i). The observations come from the Hough Forest and contain the location of the estimated point along with the probability of the class $c = 1$, i.e. the probability of the point being visible. Quantising this probability into 5 equally spaced bars we get 320 different observations, 5 probability bars for each of the 64 grid locations ($\mathbf{O} = \{O_{S_1,P_1} ... O_{S_1,P_5}, O_{S_2,P_1} ... O_{S_2,P_5}, ..., O_{S_{64},P_1} ... O_{S_{64},P_5}\}$). All the other variables ($\mathbf{T}, \mathbf{P}, \mathbf{b_0}$) are calculated experimentally from previously unseen images. A positive reward is given if the robot grasps the correct location, a smaller reward if it grasps a neighbour location and a very negative reward if it grasps any other location. In addition, if the robot decides to rotate the garment, a positive reward is given if the grasp point was invisible and a small negative reward is given if the grasp point was visible.

The solution of the unfolding POMDP gives the optimal action policy for estimating a desired point. The robot rotates the garment until a certain confidence about the location of the point is reached and then decides to grasp it. If the rotation exceeds 360 degrees the process is restarted by picking again the lowest point. This probabilistic planning makes point estimation very robust and insensitive to noisy estimations.

VI. Experiments

Robot Setup. We have tested our methods using a dual manipulator by YASKAWA[1]. We capture depth images from an Xtion depth sensor placed between the arms at a fixed height while grasping is based on custom “claw like” grippers [24].

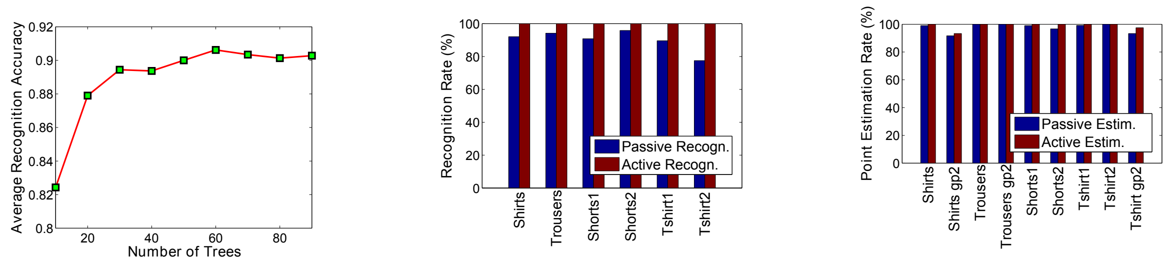
Training Set. Our training dataset was created using 24 regular-sized clothes of various sizes and fabric types, 6 of each category. In order to cover a variety of possible clothes configurations, each garment was grasped 20 times from each lowest point and 40 images were captured while the garment was rotating, covering the viewpoints of all 360 degrees. The final database contains 28,800 depth images. Image labelling was done manually using fiducial markers over key points to facilitate the task.

Testing Set. Testing was based on a dataset containing only novel items (not in the training dataset). The clothes used for testing were 3 per category (12 total). For measuring accuracy, we used 240 depth images for each category (1440 in total) while the same clothes were used for evaluating the whole unfolding procedure.

Random Forest configuration. For training Random Forest trees we used 5000 random candidate tests, 70 candidate thresholds per test and 4 minimum samples per node, while no restriction on the maximum depth of trees was imposed. Figure 6(a) shows the average recognition rate in relation with the number of trees in the forest. We see that above 60 trees recognition rate remains the same, therefore this is the number of trees used in our forests. We used the same configuration for both random and Hough forests. Training a tree for recognition takes about half an hour while a tree for point estimation takes about 10 minutes on an Intel i7 CPU. Inference of one frame takes less than 40ms.

POMDPs. For solving our POMDPs we used the point-based SARSOP algorithm [22]. All transition and observation probabilities needed were calculated experimentally

[1]Two M1400 Yaskawa arms mounted on a rotating base



(a) (b) (c)

Fig. 6. a) The average recognition rate as the number of trees increases, b) Recognition rate of each class for passive and active method, c) Grasp point estimation rate of each possible occasion. gp2 means the estimation of the 2nd grasping point while the garment is hanging from the 1st one.

Fig. 7. Examples of successful grasp point detection

from the training set of clothes from images not used for training the forests. The rotation angle a used is 10 degrees. Moreover, reward values affect the level of confidence required for the robot in order to take a decision about an action. Having a confidence level above 0.9999 we were able to achieve 100% active recognition accuracy requiring only few rotations (Table I), while also improving point estimation results (Fig. 6(c)).

Recognition Results. It is difficult to compare clothes recognition results with other approaches as each author makes different assumptions. Fig. 6(b) compares results of our passive and active recognition. Active recognition achieved 100% accuracy while the one-frame (passive) recognition had 90% success rate in average.

Grasp point estimation results. Fig. 6(c) shows the point estimation results, while there is no other similar work to compare with. Again, an improvement of active over passive estimation is observed. Some success and failure examples of point estimations are shown in Figures 7 and 8(a).

End-to-end unfolding results. We have conducted 120 full end-to-end unfolding experiments, using each test garment 10 times. We consider the unfolding successful when the grasped points are 10cm close to the desired points. We also implemented a shape matching algorithm [25] in order to automatically asses the final unfolded state by matching the unfolded garment to predefined unfolded templates. If the matching score is below a threshold, the process is restarted. The process is also restarted in case the garment is rotated more than 360 degrees and no recognition or point estimation occurs. If robot restarted the process more than once it was assumed as failure. 112 out of 120 experiments the unfolding was successful. Recognition achieved 100% accuracy requiring 2.4 rotations average. We encountered three types of errors caused by the gripper: a) the robot grasped two points on the garment because its opening is not adjustable, b) inaccurate alignment of the robot gripper because the plane fitting around the estimated point was affected by noisy depth sensor data, c) gripper couldn't grasp a surface because of its local shape. These errors are illustrated in Fig. 8(b) - 8(d). 15 out of 22 total errors (68%) were caused by the gripper while only 7 (32%) were caused by incorrect point estimations. However, in 14 erroneous situations, errors were perceived by the robot which restarted the process and successfully unfolded the garment. In the remaining 8 erroneous cases, robot required more than one restart and thus we considered them as failures. In no experiment the grasped point was more than 10cm away from the ground truth. Finally, the average unfolding time is 2min 23.5sec with the robot set in its safety speed mode. Results are summarized in table I[2]. In the state of the art work [4], their videos show a good case scenario of unfolding shorts in about 3:20 and a more complex scenario of unfolding a T-shirt in about 4:30, which is slower than our average time in every category. Although robot parameters like motion planning affects the execution time, making it a not very reliable metric, our approach have reduced the unfolding movements to three grasps and few rotations. Such actions are generally executed faster compared to lying the garment on a table and re-grasping it quite a few times.

VII. CONCLUSIONS

We have proposed a complete solution to the problem of autonomously unfolding an article of clothing. We used random forests for clothes classification and Hough forests for grasp point estimation in order to completely unfold four categories of clothes. Both were implemented into a POMDP framework for planning a dual manipulator optimally, enhancing the recognition and the unfolding procedure. We achieved very high recognition and unfolding success rate

[2]Supplementary video: `http://clopema.iti.gr/ICRA_2014/`

(a) (b) (c) (d)

Fig. 8. a) Failure examples of grasp point estimation (in green is the ground truth, when missing point is invisible), b) The opening of the gripper is not adjustable causing the grasp of two points, c) Inaccurate grasping because of noisy point cloud, c) Gripper cannot grasp some kinds of surfaces easily

TABLE I
UNFOLDING RESULTS

	Experiments	Successful Unfoldings	Successful Recognition	Average Rotations for recognition	Estimation Errors	Gripper Errors	Average time
Shirts	30	27	30	0,8	2	4	150 sec.
Trousers	30	30	30	1,1	0	3	136 sec.
Shorts	30	30	30	2,7	0	2	127 sec.
T-shirts	30	25	30	5	5	6	161 sec.
Total	120	112	120	2.4 (avg.)	7	15	143.5 (avg.)
%		**93.3%**	**100%**				

while our methods operate faster compared to the state of the art. The majority of our errors were caused by unsuccessful grasping of an estimated point. One reason is the noise of the Xtion depth sensor which causes inaccurate motion planning of the manipulators. The other reason is the lack of dexterity of the gripper, making the grasping of very thin or flat surfaces very difficult. The solution to the first problem would be a stereo camera with high resolution, which is planned to be used in the near future. On the other hand, a more humanoid gripper seems a more appropriate solution for clothes manipulation.

ACKNOWLEDGMENT

The authors were supported by the EC under the project FP7-288553 CloPeMa. T-K Kim was partially supported by EPSRC grant (EP/J012106/1) 3D intrinsic shape recognition.

REFERENCES

[1] F. Osawa, H. Seki, and Y. Kamiya, "Unfolding of Massive Laundry and Classification Types," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, pp. 457–463, 2007.

[2] S. Miller, M. Fritz, T. Darrell, and P. Abbeel, "Parametrized shape models for clothing," in *Proc. ICRA*, 2011.

[3] A. Ramisa, G. Alenya, F. Moreno-Noguer, and C. Torras, "Using depth and appearance features for informed robot grasping of highly wrinkled clothes," in *Proc. ICRA*, 2012.

[4] M. Cusumano-Towner, A. Singh, S. Miller, J. O'Brien, and P. Abbeel, "Bringing clothing into desired configurations with limited perception," in *Proc. ICRA*, 2011.

[5] J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, and P. Abbeel, "Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding," in *Proc. ICRA*, 2010.

[6] K. Hamajima and M. Kakikura, "Planning strategy for task of unfolding clothes," in *Robotics and Autonomous Systems*, vol. 32, 2000, pp. 145–152.

[7] M. Kaneko and M. Kakikura, "Planning strategy for putting away laundry-isolating and unfolding task," in *Proc. of the Int. Symposium on Assembly and Task Planning*, 2001, pp. 429–434.

[8] D. Triantafyllou and N. Aspragathos, "A vision system for the unfolding of highly non-rigid objects on a table by one manipulator," in *Intelligent Robotics and Applications*, 2011, vol. 7101, pp. 509–519.

[9] B. Willimon, S. Birchfield, and I. Walker, "Model for unfolding laundry using interactive perception," in *Proc. IROS*, 2011.

[10] B. Willimon, S. Birchfield, and Walker, "Classification of clothing using interactive perception," in *Proc. ICRA*, 2011.

[11] B. S. Willimon Bryan, Walker Ian, "Classification of Clothing Using Midlevel Layers," *ISRN Robotics*, 2013.

[12] Y. Kita, T. Ueshiba, E. Neo, and N. Kita, "Clothes state recognition using 3d observed data," in *Proc. ICRA*, 2009.

[13] Y. Kita, F. Kanehiro, T. Ueshiba, and N. Kita, "Clothes handling based on recognition by strategic observation," in *Proc. IRAS*, 2011, pp. 53–58.

[14] C. Bersch, B. Pitzer, and S. Kammel, "Bimanual robotic cloth manipulation for laundry folding," in *Proc. IROS*, 2011.

[15] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," in *Proc. CVPR*, 2011.

[16] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky, "Hough forests for object detection, tracking, and action recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 33, no. 11, pp. 2188–2202, 2011.

[17] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[18] T. Yu, T.-K. Kim, and R. Cipolla, "Unconstrained monocular 3d human pose estimation by action detection and cross-modality regression forest," in *Proc. CVPR*, 2013.

[19] B. Rosenfeld, "The curvature of space," in *A History of Non-Euclidean Geometry*, ser. Studies in the History of Mathematics and Physical Sciences. Springer New York, 1988, vol. 12, pp. 280–326.

[20] S. Prentice and N. Roy, "The belief roadmap: Efficient planning in linear pomdps by factoring the covariance," in *Robotics Research*, 2011, vol. 66, pp. 293–305.

[21] N. Dantam, P. Koine, and M. Stilman, "The motion grammar for physical human-robot games," in *Proc. ICRA*, 2011.

[22] D. Hsu, W. S. Lee, and N. Rong, "A point-based pomdp planner for target tracking," in *Proc. ICRA*, 2008.

[23] P. Monso, G. Alenya, and C. Torras, "Pomdp approach to robotized clothes separation," in *Proc. IROS*, 2012.

[24] T.-H.-L. Le, M. Jilich, A. Landini, M. Zoppi, D. Zlatanov, and R. Molfino, "On the development of a specialized flexible gripper for garment handling," *Journal of Automation and Control Engineering*, vol. 1, no. 3, 2013.

[25] I. Mariolis and S. Malassiotis, "Matching folded garments to unfolded templates using robust shape analysis techniques," in *Computer Analysis of Images and Patterns*, 2013, vol. 8048, pp. 193–200.