Learning to measure for preshipment garment sizing

Joan Serrat[a,*], Felipe Lumbreras[a], Idoia Ruiz[a]

[a] *Computer Vision Center and Dept. of Computer Science, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.*

Abstract

Clothing is still manually manufactured for the most part nowadays, resulting in discrepancies between nominal and real dimensions, and potentially ill-fitting garments. Hence, it is common in the apparel industry to manually perform measures at preshipment time. We present an automatic method to obtain such measures from a single image of a garment that speeds up this task. It is generic and extensible in the sense that it does not depend explicitly on the garment shape or type. Instead, it learns through a probabilistic graphical model to identify the different contour parts. Subsequently, a set of Lasso regressors, one per desired measure, can predict the actual values of the measures. We present results on a dataset of 130 images of jackets and 98 of pants, of varying sizes and styles, obtaining obtaining 1.17 and 1.22 cm of mean absolute error, respectively.

Keywords: apparel, computer vision, structured prediction, regression

1. Introduction

Automatically measuring objects in order to check whether they conform to a certain tolerance with respect to nominal values is a problem that frequently occurs in the industry, for instance for quality control. Machine vision systems obtain these measures by analyzing images of the inspected objects. On them it is possible to assess not only the dimensional quality of objects but also to quantify their shape attributes, position, orientation, alignment etc. [1] both in two and three dimensions [2].

Many and diverse manufacturing processes have benefited from the design of problem-specific automatic visual measurement methods [3–7]. The difficulty of the problem such methods have to face is often correlated to the degree in which objects shape and pose vary: from objects always at a fixed position with respect

*Corresponding author.

Email addresses: joans@cvc.uab.es (Joan Serrat), felipe@cvc.uab.es (Felipe Lumbreras), iruiz@cvc.uab.es (Idoia Ruiz)

to the camera and rigid, equal shapes, to free position, non-rigid and articulated shapes. In the first case, the detection of the points from which to take longitudinal or area measures is in general easier than in the second, where the shape has to be first understood before localizing the points of interest along its contour. The problem we address is the automatic visual measurement of apparel. It lies in the later group, that requires shape understanding.

Unlike mass manufacturing of other products, which often uses precise machinery and automation, clothing is still largely done by hand. This results in discrepancies between the actual and the nominal dimensions. This adds to the problem of deciding how much smaller the dimensions of the garment with respect to the body need to be in order to achieve a good fit.

Difference in dimensions is a problem in ready-to-wear because when a customer buys the same garment shape in the same size several times, he or she wants them to have the same dimensions. Moreover, new players in the apparel industry specialize on the manufacture of custom tailored garments through web sites. Typically, they offer the possibility to potential clients to enter their body measures so that the product they receive fits them perfectly. Again, and now even more critical, it may be the case that the garment finally made does not satisfy this key goal. Ill-fitting garments don't just imply extra costs in the form of returns and unsellable products but can turn customers to a competitor. Thus, the final checking of garments dimensional measures before they are dispatched to customers is highly desirable. But manually obtaining them presents several drawbacks:

- Mainly, the inspection of all the produced garments is costly, so perhaps not all the garments are inspected but just some random sample. The reason is that even taking and writing down 10 measures per clothing piece takes some minutes.
- Measurements have some degree of subjectivity not only because they are collected by humans but because, obviously, textiles fold.
- Another source of subjectivity is that whereas some measures are distances between characteristic points like seams or profile corners, others are between ill-defined points (figure 2), their location being left to the criteria of the inspector.

1.1. Objective

In this paper we describe an automatic, fast and generic method to obtain dimensional measures from garments at preshipment time, exemplified with two common types : jackets and pants. Each one has its own set of measures and of

course a different shape and pattern of shape variation. One possible —though cumbersome— approach is to build an ad-hoc, independent method for each type.

Our goal is more far reaching: a single method able to cope with these two types and easily extensible to new types coming in the future, with their own shapes and list of measures. The solution is, as we will see, a method that *learns how to measure* a certain type of garment just from a training set of samples, being the samples pairs of image and corresponding groundtruth measures values. That is, the method does not depend on the particular shape of a type of garments. Learning means being able to understand the apparel contour in order to then obtain an estimation for the value of the sought measures —normally distances— without knowing the particular points from which they are defined.

1.2. Related work

The computer-assisted visual inspection within the apparel industry covers several of the phases of the production pipeline [8]. However, most of the works concentrate in the early stage of fabric inspection, where the problem is the detection of a number of defects like stains, pilling, weaving and knitting errors, etc. [9–11]. From the computer vision point of view, these are problems of texture analysis, for which [12] offers a through review and categorization of techniques. The problem of automated inspection of manufactured garments has received comparatively less attention, perhaps because of the larger variability of the kind of objects to inspect.

As for the problem we address, obtaining dimensional measures of garments, the literature is scarcer. To the best of our knowledge there are only four works coincident to ours. The oldest one dates back to thirty years ago [13], where the authors proposed a simple algorithm to find key points along the contour of garments from which later obtain dimensional measures. The goal was to assess the shrinkage of knitted garments, but they already mention the possibility of "routine inspection prior to dispatch to customers". After a large gap, next two works [14], [15] compute the Freeman chain code of the contour in order to locate its corners which are the same type of key points. Like in the previous case, no extensive results are presented but just a few examples. Last and most recent work is [16], which follows the same strategy of key point detection prior to measurement. An interesting contribution of this work is that they apply their method to several types of garments (shirt, pants, vest, skirt) plus the ability to recognize it from the image. Each key point among a list of corner points is selected by comparing relative position, concavity and convexity with those of a template garment. Like before, no analysis of the results is provided, globally, per garment type or per dimensional measure. Furthermore, no details are provided on the specific parameters and procedure to select the key points beyond that they are the most similar to those in the template.

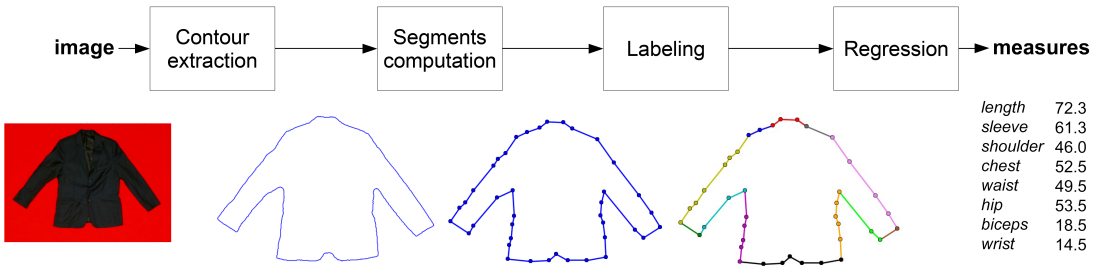

Figure 1: Scheme of the method.

Coincident with our goal of obtaining dimensional measures from ill-defined points along the contour of a varying shape, the authors of [17] propose a method to automate the task of a tailor by image analysis of human body profiles. Besides needing infra-red markers, this method relies on heuristics to locate the endpoints of the dimensional measures, similarly to [16].

As we will see in next section, differently from all of these works we strive for a generic method whose parameters are learned from a training set. Hence, it will be easy to extend from one type of garment to new types. Moreover, we do not classify corner or high curvature points into some key point class, in order to avoid local decisions that are prone to error. Instead our method will consider the contour globally and from it will obtain the dimensional measures by regression, not by computing any kind of distance.

1.3. Method overview

Figure 1 illustrates the idea of the method. The first step is to extract the contour of a garment from the image and then approximate it by a number of straight segments. We have previously decided a division of the contour of each type of garment into a set of meaningful parts. For example in jackets these are the neck, left shoulder, left wrist, outer and inner left sleeve, left trunk, etc. (see figures 4 and 11 for the complete list). The second step is to classify each segment as belonging to one of these parts. We cast this task into labeling a number of unknown variables of a chain conditional random field (CRF), being the variables the segments, and the labels the former types of contour parts. Probabilistic graphical models like CRFs allow us to properly combine

- the vector of features describing the segments, like their position or orientation, which play the role of likelihood, and
- the strong dependencies or relationships that exist between labels of consecutive variables, namely, a prior: after an sleeve segment often comes another sleeve segment, less often a wrist segment, and never trunk, for instance.

Labeling results from inference on this CRF, that is, finding the configuration of labels that maximizes the joint a posteriori probability, which is proportional to the likelihood times the prior. However, in order to succeed it is critical to learn in advance the CRF parameters, which we achieve through the structured support vector machine (S-SVM) technique.

Once labeled all the segments, the last step estimates the measures defined for that type of garment. Now we could try to locate the end points for each measure of interest from the labeled segments. However, this would need an ad-hoc, apparel-dependent algorithm (one for jackets, another for pants, etc.), opposite to our goal of a completely learnable system. Moreover, some measures are defined between hidden points, like the chest in jackets, whose endpoints are at the joint of the sleeves and the trunk (figure 2), and thus can not be obtained in this way. Our approach is, instead of computing distances, to regress them from distances among the labeled segments. Thus we again end up learning parameters, now of a set of regressors. We will show that this indirect approach yields very good estimations for all measures. Furthermore, being able to estimate all the measures from a single image saves the time to place the garments in different poses in order to make all the endpoints visible.

To summarize, the contributions of this work are

- a new method for apparel visual measurement with a sound theoretical basis, that is fast and performs well, as shown by processing a relatively large number of samples
- easily extensible to new types (that is, shapes) of garment, as we will show when switching from jackets to pants, because it learns to measure from examples
- with a key component, a CRF for segment labeling, which is highly accurate in spite of being based on few and very simple features
- able to obtain accurate longitudinal dimensions for which their end points are not even visible, because they are regressed from other distances

The rest of the paper is organized as follows. Next section presents the details of the dataset we have used, composed of two subsets, one per type of apparel. In section 3 we explain the two first components of the method, namely, the extraction of the garment contour and its segments, and their labeling. For the sake of comprehensiveness, we include background on CRFs and learning with S-SVM. Also, we report there the results for this component and compare it to a well known non-structured prediction technique, linear SVM. Section 4 continues with the last component, the one computing all the measures from the labeled

Figure 2: Eight measures on jackets and five on pants that form the groundtruth for these two garment types. Some of them are distances between hidden points (chest, crotch) or points not on a corner or a seam (biceps, waist, thigh), in both cases shown as dashed lines.

segments through regression. Section 5 presents several statistics comparing the estimated and real measures. Finally, in section 6 we draw the conclusions and avenues of future work.

2. Dataset

The dataset used for learning the CRF and the regressors parameters is composed of 130 images of jackets and 98 pants. Images were captured by an off-the-shelf USB web camera at a resolution of 1920×1080 under controlled LED lighting. As for the imaged garments, their axes of variations are not only color and size but also style (i.e, shape), slight rotations plus, more importantly, fabric folds and the positioning of the "articulated" parts, sleeves and legs. For the sake of speed, garments are placed under the camera without enforcing strict rules, only trying to avoid excessive folding and to keep sleeves and legs apart from the trunk and each other, respectively. Despite around 20% of samples do not fulfill this constraint we have still kept them in the dataset. Top and bottom rows of figure 3 show three examples of well placed jacket and pants, whereas middle row shows typical a priori problematic samples.

Along with the images, we have the groundtruth values for the measures corresponding to each type of apparel (figure 2). They were manually collected with a regular metric tape with marks every 0.5 cm, so that measures are quantized to multiples of this number. These values are destined to train the regressors and then to evaluate the estimated measures, according to a cross-validation procedure that we will introduce in the results section.

In addition, in order to previously train and asses the CRF that labels the contour segments we also need groundtruth. This time it consists on the class of contour part for each of the segments of all available jackets and pants, totaling more than 8000 segments. Their manual annotation is not as long as it may seem, since we only need to provide the number of the first segment for each contour part, because segments are sorted clockwise. It amounts to just 12 and 7 numbers per sample for jackets and pants, respectively. Figure 4 shows a number of such annotated samples.

Figure 3: Examples of jackets and pants. Top and bottom rows depict well laid out garments, keeping sleeves and legs apart. Middle row are problematic yet considered samples.

3. Segment processing

3.1. Extraction of contour segments

Images were recorded by laying garments on a red background, a color so distinct from those of garments, that simple thresholding already segments well the garment region. From the binary segmented image it is then also straightforward to extract the contour.

Now, the most direct way to obtain the sought measures is to find out the two end points for each one of them. For instance in jackets, as figure 2 shows, measure "sleeve" is the distance from the high curvature point that separates shoulder and sleeve, to the left corner of the wrist, "hip" is the distance between the two lower corners of the trunk, etc. Many such endpoints are corners, namely, points where the contour distinctly changes direction (figures 5a and 5b). This suggests a way to find the endpoints : select a fixed number of candidates among the extrema of curvature by ranking them according to their magnitude, and then apply a set of heuristic rules that combine the magnitude, sign of curvature and relative position in order to identify the endpoints among those candidates. This approach, though feasible, has two drawbacks. First and foremost, it's a long and adhoc procedure : for each endpoint of each measure of each type of garment we have to define a certain heuristic rule, which can be rather complicated. Second, it implicitly assumes the garment is well placed, thus requiring the user to lay it down carefully under the camera. Accidental textile folds may provoke wrong detections, as the decisions (the rules) are made locally and applied in some predefined order.

In order to avoid the former drawbacks, we resort to an indirect but more general approach. For each garment type we define:

1. a list of parts in which the whole contour is divided, as shown in figures 4 and 11 for jackets and pants, respectively
2. the list of measures to take (but not from where) by regression of all possible distances from the starting point of each part to the rest

These two will be the only apparel-dependent information the proposed method will demand, in form of training data. Furthermore, we approximate the contour by straight segments, thus simplifying its representation. There exist many algorithms to obtain a polygonal approximation of contours into a given number of segments, as reviewed in [18, 19] and references therein. For our particular problem, a sensible way to do it is to not simply perform a polygonal approximation but to somehow enforce the segments to end at extrema of curvature, because they often coincide with the endpoints of longitudinal measures. Therefore, from the contour points $(x_i, y_i), i = 1 \ldots$ we compute the curvature as [20]

$$\kappa_i = \frac{x'_i y''_i - x''_i y''_i}{(x'^2_i + y'^2_i)^{3/2}} \tag{1}$$

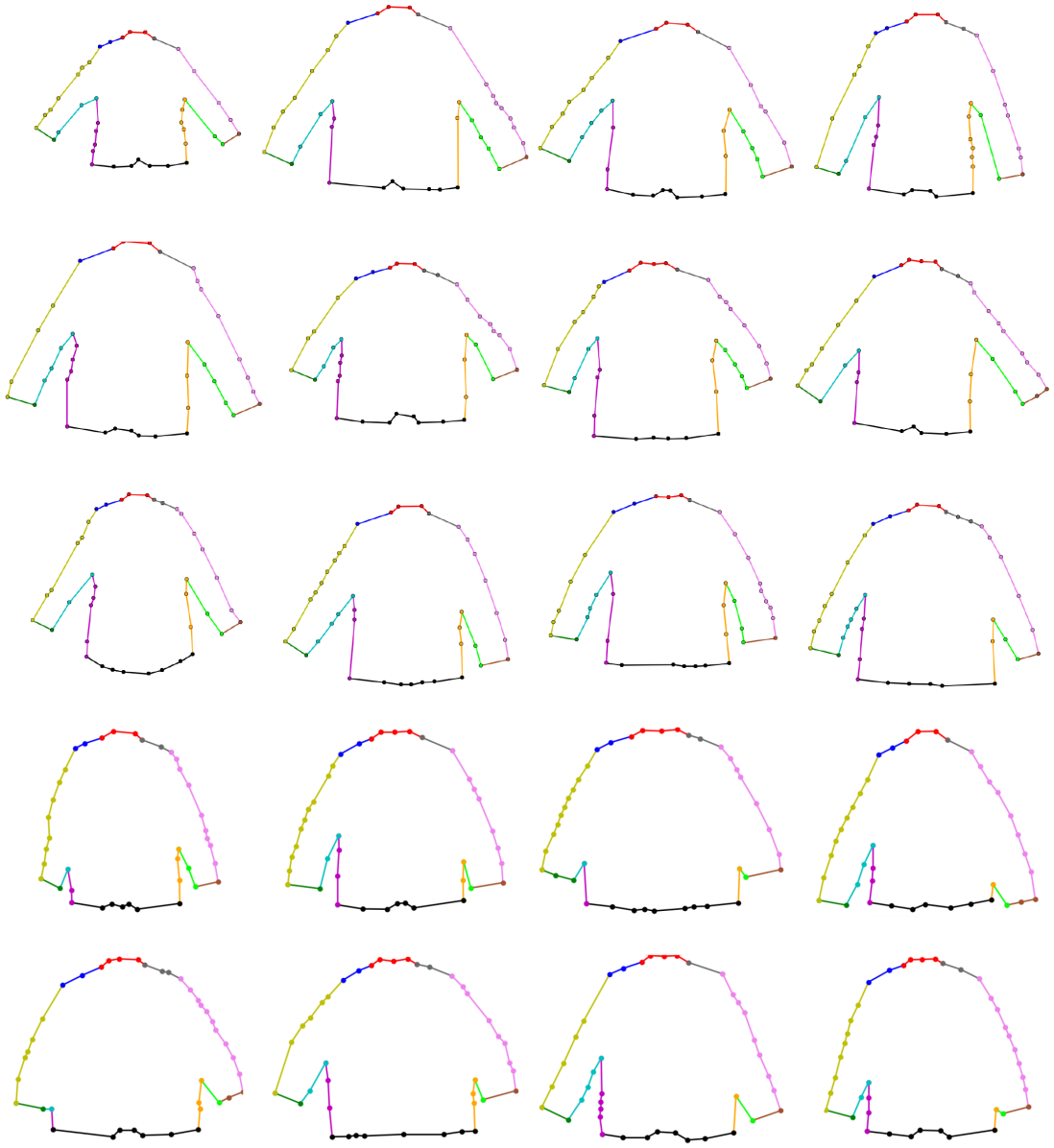

Figure 4: Groundtruth of segment labels for 20 samples. Color denotes the segment class, that is, the type of contour part. These are the following: neck at top (red), waist at the bottom (black), and left and right shoulder (blue, gray), outer sleeve (yellow, magenta), wrist (green, brown), inner sleeve (cyan, light green) and trunk (pink, orange). Last two rows show cases of badly placed garments. Best viewed in color.

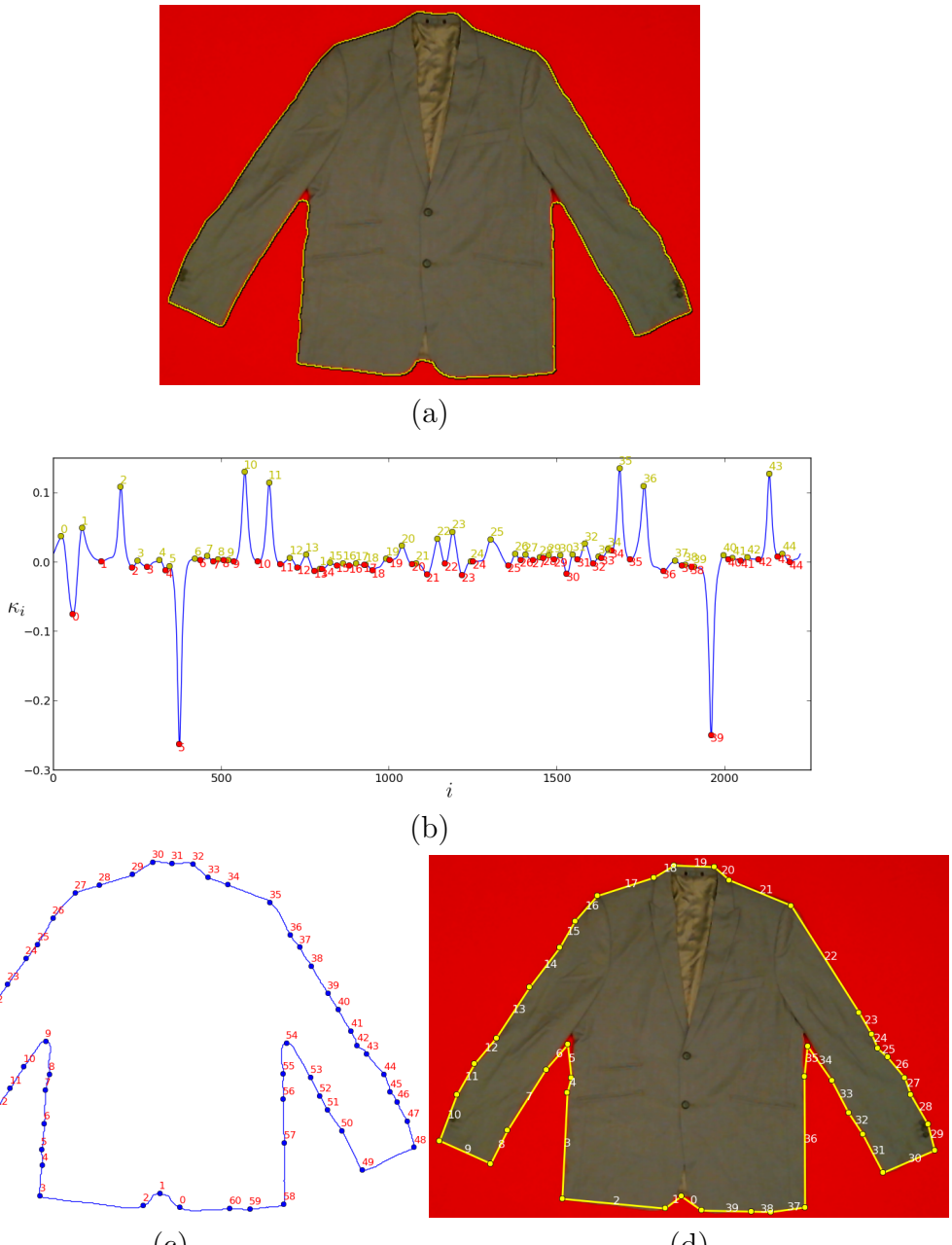

Figure 5: Process for the extraction of segments. (a) Contour from segmentation, (b) curvature of the contour and its extrema, (c) location of extrema, which are the input to the process which extracts the (d) segments.

where the derivatives are approximated by convolving the x and y vectors with first and second order differences of discretized Gaussian kernels,

$$\begin{aligned} x'_i &\approx x * (G_\sigma(i) - G_\sigma(i-1)) \\ x''_i &\approx x * (-G_\sigma(i-1) + 2G_\sigma(i) - G_\sigma(i+1)) \end{aligned} \tag{2}$$

with σ controlling the degree of regularization. The largest l extrema in magnitude are kept, giving rise to the same number of segments, with $l = 40$ for jackets and $l = 30$ for pants (figures 5c and 5d). The next step now is to classify each segment as belonging to a certain contour part, that is, a labeling problem.

3.2. Segment classification

Structured prediction consists in assigning a class label not to a single instance but to each of the instances of an aggregate or structure —a chain, tree, graph, etc. depending on how a problem is modeled—, with modeled interdependencies among them. Before the surge of deep learning, probabilistic graphical models were the technique of choice for structured prediction.

In this framework, and following the notation of [21], $y = (y_1 \dots y_n)$ denotes the n unknown variables of the aggregate. The space of all possible labellings is $\mathcal{Y} = Y^n$, being $Y = \{1 \dots c\}$ with c the number of possible classes or labels. $x \in \mathcal{X}$ is the set of known variables or observations, for instance one per unknown variable. $\mathcal{X}$ is the space of all possible observations for the known variables. Each observation x_i is actually a vector of d features or measurements of different quantities which are somehow related to some of the unknown labels y. It is assumed that the conditional probability distribution $p(y|x)$ has the following shape :

$$p(y|x) = \frac{1}{Z(x)} \exp[-E(x,y)] \tag{3}$$

$$Z(x) = \sum_y \exp[-E(x,y)] \tag{4}$$

being $Z(x)$ a normalizing constant known as the partition function. E is the energy function, and maximizing $p(y|x)$ with respect to y is equivalent to energy minimization (or maximization of $-E$), for which several algorithms have been proposed. Many of them leverage the fact that, for maximum conditional probability inference to be computable, E has to have a certain structure, namely it factorizes as

$$E(x,y) = \sum_{F \in \mathcal{F}} E_F(x_F, y_F) \tag{5}$$

where $\mathcal{F}$ is a set of factors, groups of known and unknown variables in which we partition the model graph. For instance, in the popular pairwise models, each factor F consists of two directly related (neighbor) variables y_i, y_j plus all the observations x_k also directly related to them.

Let the prediction function of the graphical model $f : \mathcal{X} \to \mathcal{Y}$ parametrized by coefficients w be

$$f(x; w) = \arg\min_{y \in \mathcal{Y}} \sum_{F \in \mathcal{F}} E_F(x_F, y_F) = \arg\max_{y \in \mathcal{Y}} g(x, y; w) \tag{6}$$

As for g, for tractability reasons and also because in practice it allows to obtain good solutions to many real problems, it is assumed that $g(x, y)$ is linear in the factor functions

$$g(x, y; w) = \sum_{F \in \mathcal{F}} \langle\, w_F, \psi_F(x, y) \,\rangle = \langle\, w, \psi(x, y) \,\rangle \tag{7}$$

where w and ψ are the concatenation of coefficients and factor functions, respectively.

Now, returning to our problem of segment labeling, the CRF model we adopt is the linear chain or sequence shown in figure 6. In it, $x = (x_1 \dots x_n)$, being n is the number of segments, d the number of features per segment, $x_i \in \mathbb{R}^d, i = 1 \dots n$ the observations or value of the features for the i-th segment, and $y_i \in 1 \dots c$ the label assigned to it. The small solid squares are factor functions which model the dependencies of y_i on both the i-th observation and the following (unknown) label y_{i+1}. x_i will be a vector of $d = 5$ very simple features : the normalized coordinates of its endpoints plus the angle it forms with the horizontal axis.

Prediction is then performed according to the following classification rule :

$$\begin{aligned} y^\star &= \arg\max_{y \in \mathcal{Y}} \langle\, w, \psi(x, y) \,\rangle \\ &= \arg\max_{y \in \mathcal{Y}} \sum_{i=1}^{n} \sum_{p=1}^{c} \sum_{j=1}^{d} w_{pj} x_{ij} \mathbf{1}_{y_i = p} + \sum_{i=1}^{n-1} \sum_{p=1}^{c} \sum_{q=1}^{c} w_{pq} \mathbf{1}_{y_i = p,\, y_{i+1} = q} \end{aligned} \tag{8}$$

where $y_i \in Y$ and $\mathbf{1}_{y_i = p,\, y_{i+1} = q}$ evaluates to 1 if $y_i = p$ and $y_{i+1} = q$. Therefore, there are cd unary coefficients w_{pj} and c^2 pairwise coefficients w_{pq}. The former weight the features while the later encode the compatibility between pairs of successive labels. Even though $|\mathcal{Y}| = c^n$ is huge, prediction $y^\star$ can be computed with a complexity of $O(cn)$ by means of the max-product inference algorithm [21]. Moreover, it is not approximated but exact because the absence of loops in a chain CRF.

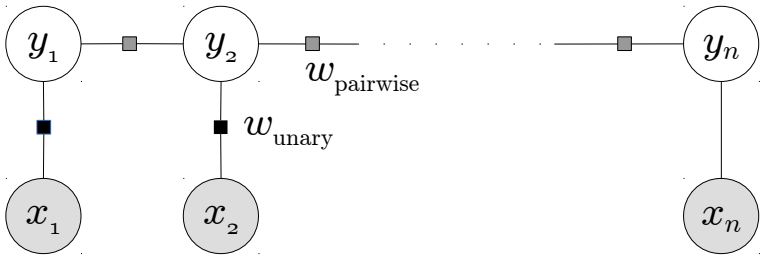

Figure 6: Graphical model, a chain CRF. Solid squares are factor functions, in our case the dot products of equation (8).

Note that an alternative formulation is possible, namely, classifying the end-points of segments —extrema of curvature— into either one of the key points from which longitudinal measures are taken, or to non-key point. This is a far less convenient formulation as the former one based on segment labels, because it does not leverage the constraints the context provides. In effect, since there are much more irrelevant curvature extrema than key points, what we would learn then is just that 1) after a non-key point comes almost always another non-key point, and 2) after each key point most often comes a non-key point, not another certain key point.

Table 1 shows the accuracy of segment labeling when assessed with leave-one-out cross-validation : 99.67% for jackets and 100% for pants. The key to the success of structured prediction is not as much that inference is exact but having properly learned the coefficients of the model. We have resorted to a minim-risk learning technique known as structured support vector machine (S-SVM) [22, 23] that we summarize following the notation of [21].

3.2.1. Structured Support Vector Machines

Let $\Delta(y, f(x; w)) : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ be a loss function that returns the cost of getting an answer $y' = f(x; w)$ instead of the true labeling y. Depending on the graphical model, the loss function is different. In the case of a single label, it is normally the 0-1 loss $\Delta(y, y') = \mathbf{1}_{y \neq y'}$. However for sequence prediction it makes more sense the Hamming loss, that counts the number of wrong labels,

$$\Delta(y, y') = \frac{1}{N} \sum_{i=1}^{n} \mathbf{1}_{y_i \neq y'_i} \tag{9}$$

that we adopt here. S-SVM seeks to minimize the regularized empirical risk of

the training set $(x^n, y^n), n = 1 \dots N$, being N the number of samples :

$$w^\star = \arg\max_w \frac{1}{2}||w||^2 + \frac{C}{N}\sum_{i=1}^{N} l(x^n, y^n; w) \tag{10}$$

with the loss function l being

$$l(x^n, y^n; w) = \max_{y \in \mathcal{Y}} \Delta(y^n, y) - g(x^n, y^n; w) + g(x^n, y; w) \tag{11}$$

It can easily be shown that l is a convex upper bound of $\Delta(y^n, f(x^n; w))$, the true loss. C for its part is a hyperparameter to weight the contribution of the regularization term, set experimentally by cross-validation.

Equation (10) involves a non-differentiable operator so plain gradient descent can not be applied. However, it can be numerically optimized through subgradient descent minimization. Alternatively, it can be formulated in an equivalent way very close to classic SVMs, by replacing the max operator by N slack variables :

$$w^\star = \arg\min_{w,\, \xi^n \in \mathbb{R}} \frac{1}{2}||w||^2 + \frac{C}{N}\sum_{n=1}^{N} \xi^n \tag{12}$$

subject to, for all $n = 1 \dots N$,

$$g(x^n, y^n; w) - g(x^n, y; w) \geq \Delta(y^n, y) - \xi^n, \quad \forall y \in \mathcal{Y}$$
$$\xi^n \geq 0$$

Even though the objective function is quadratic in w under linear constraints, a well studied case in optimization theory, the sheer number of constraints, $N|\mathcal{Y}|$, seems to preclude its optimization in practice. Actually, it is not even possible to store them in a computer program. However, it turns out that only a tiny subset of these constraints are relevant because they subsume the remaining ones. There exist algorithms to iteratively find this subset, like cutting plane and Frank-Wolfe, after which the solution can be readily found. The explanation of these rather complex algorithms is out of the scope of this paper and we refer the interested reader to [22, 24] for an in depth treatment. What is more relevant here is that they are based on repeatedly performing inference, that is, solving equation (6). In our case this means making inference on a chain, equation (8), which not only is fast but also exact.

3.2.2. Comparison with non-structured prediction

One may wonder what is the benefit of performing structured prediction, where context is taken into account, compared to simpler non-structured prediction, where the labeling just depends on the actual observations of each segment. That

is, does it pay to formulate and solve the problem in a more complex way? To this end, we have performed the experiment of labeling the segments with a classic Support Vector Machine classifier based on the same features and with the same folds of leave-one-out cross-validation. Table 1 shows the accuracy for the two methods, both for jackets and pants. Looking at the accuracy, one may think the two approaches are almost the same, being the difference a scarce 1.6% and 1%, respectively. We argue the success of SVM is due to having selected highly discriminant features. Figure 7 supports this argument in that normalized end-point coordinates and angle are quite well separated for the different contour parts. However, the error distribution tells a different story. In either case wrong labeled segments are not concentrated on a few samples. For almost all samples with errors these count to 1 or 2 wrong labels. The number of samples with some labeling error is four times larger for SVM than for CRF in jackets. As for pants, structured prediction gets a striking no errors score, whereas SVM produces at least one error in one out of every five samples. Moreover, the type of errors is distinct. In CRF classification most wrong labels extend or shorten some contour part *but the right order is preserved* while in SVM it is not and this has an impact on the measures regression step. The influence of these two factors with regard the error in the final longitudinal measures will be further analyzed in the results section.

We have conducted an experiment to compare the robustness to noise of both approaches. A random Gaussian noise with different standard deviations was added to the features, prior to learning and prediction. Figure 8 shows the evolution of the number of incorrectly labeled segments as σ increases. We can appreciate that whereas SVM accuracy rapidly decreases, that of CRF degrades at a much slower pace.

A final experiment in defense of the superiority of structured prediction is removing features from the observations. What if only the angle is kept, can we still expect to classify well the segments when their position is unknown? With this unique feature CRF still produces an accuracy of 79% and 72% for jackets and pants, respectively, while SVM drops to 33% and 55%. We attribute the results to the fact that modeling and learning the context, that is, the dependencies between successive label pairs, is a powerful cue to the prediction.

An interesting side effect of modeling the problem with a CRF is that, differently from SVM, its parameters are interpretable. Unary parameters $w_{p,j}, p = 1 \dots c, j = 1 \dots d$ tell us how much the j-th feature weights with regard to classifying a segment into class p. For instance, figure 9a shows that the class of neck segments weights negatively the vertical x coordinates because they are at the top of the image. On its part, $w_{p,q}, p, q = 1 \dots c$ are the weights for the compatibility of two successive segments to have labels p and q, that is, which class of segment comes after each other class. Thus, we expect a kind of bi-diagonal matrix, telling

Table 1: Labeling accuracy for 130 jackets and 98 pants, with 40 segments per jacket and 30 per pants.

Apparel	Total segments	Classifier	Wrong labels	Accuracy in %	Samples with wrong labels	Percent samples
Jacket	5200	CRF	**17**	**99.67**	**14**	**10.8**
		SVM	100	98.08	59	45.4
Pants	2940	CRF	**0**	**100.00**	**0**	**0.0**
		SVM	31	98.95	21	21.4

us that after some label we will have the same label again or the one for the next part in the contour. Figure 9 confirms this intuition.

A last experiment explores the dependence of the classification accuracy to the size of the training set. Figure 10 shows that SVM obtains a similar result independent of the number of folds in the cross-validation, from 8 folds (16/12 jackets/pants for test, 114/86 for training) to leave-one-out (a single sample for test). It seems that SVM does not need many samples to learn a good classifier, at least with the present highly discriminative features. CRF, on its side, benefits from more samples per training fold, specially for jackets possibly because they exhibit more shape variations than pants in the placement of the sleeves.

4. Measurements

Now that we have labeled the segments, thus identifying the different contour parts, we can proceed to taking the dimensional measurements which are the true outcome of the method. In principle, we could just compute distances between certain pairs of contour part endpoints, as most measures are defined so. For example, in figure 11, pants bottom is the distance between endpoints 2 and 3 or 5 and 6, and pants length is distance between 1 and 2 or 0 and 6. However, this direct approach does not take into account the uncertainty in segment labeling: it may not always be perfect, and a single error could induce some completely wrong measures. Moreover, some measures are defined between points which are not any of the endpoints of the contour parts: chest, waist and biceps in jackets, hip, crotch, thigh in pants.

We propose instead to regress each measure from the list of all the distances between the starting points of each part. This makes $c(c-1)/2$ distances if c is the number of contour parts or segment labels. In this way, even if any segment is misclassified, it is possible that its effect is reduced because regression depends on a certain number of distances between contour parts, most of which will be right.

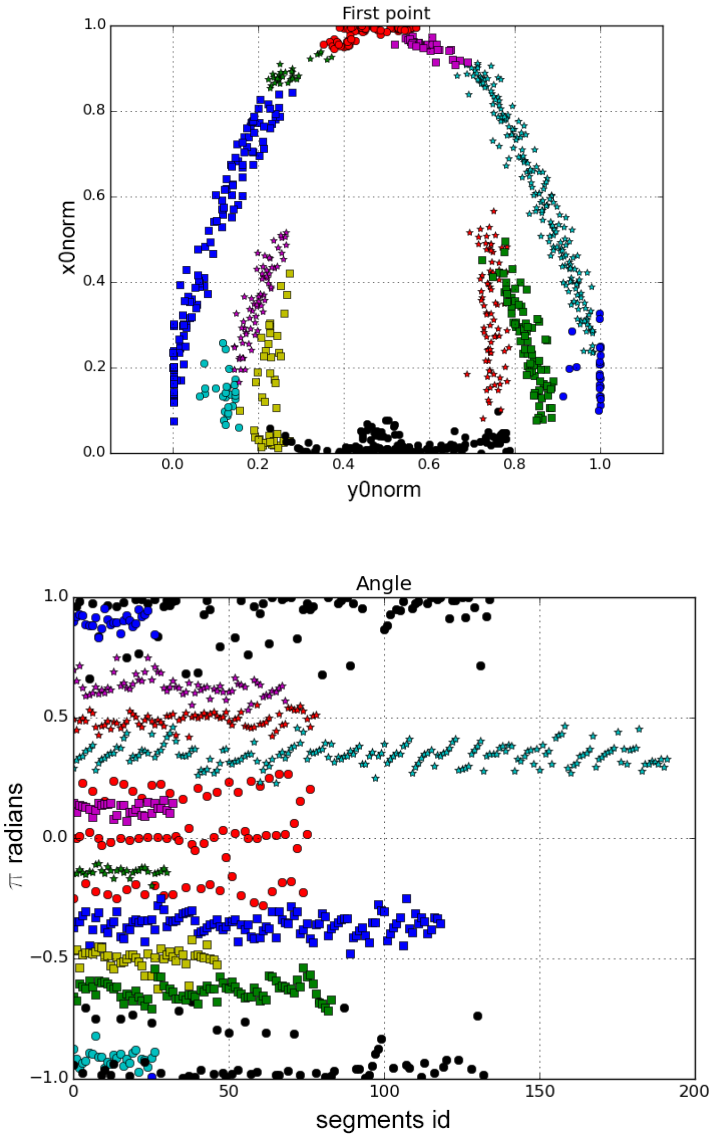

Figure 7: Normalized coordinates of first point of each segment (top) and segment angle (bottom) for a subset of 23 jackets. Each symbol represents one of the 12 classes of segments. Best viewed in color.

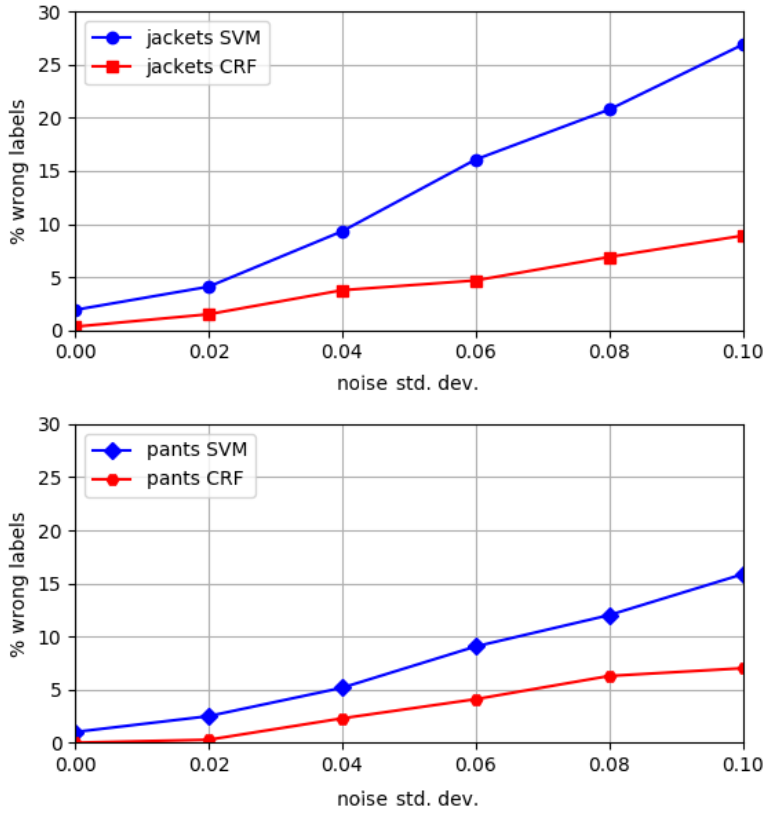

Figure 8: Rate of wrong labels when Gaussian noise is added to the features rescaled to the $[0, 1]$ range.

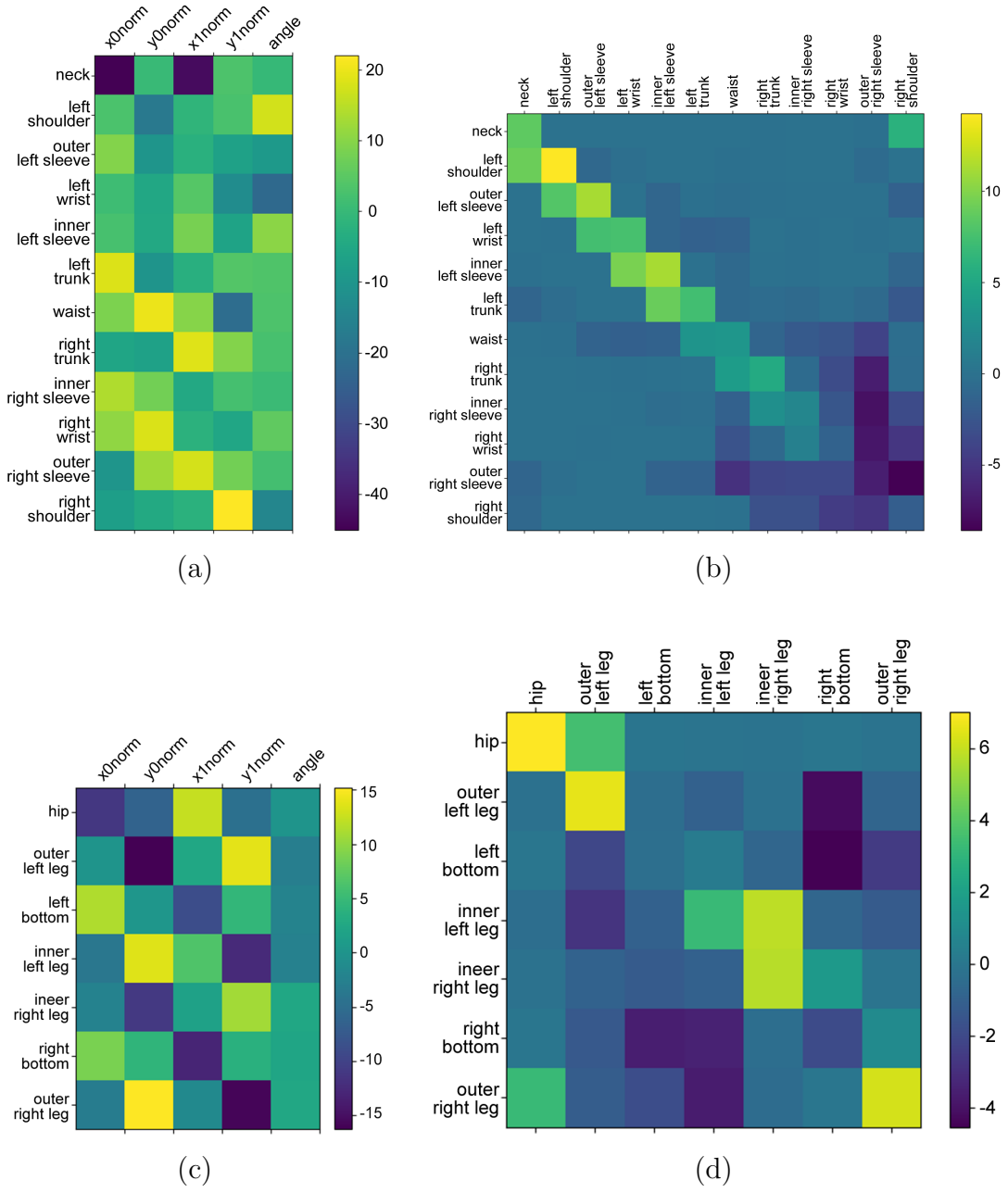

Figure 9: Learned unary (a), (c) and pairwise (b), (d) coefficients for jackets (top) and pants (bottom). `x0norm`, `y0norm`, `x1norm`, `y1norm` stand for the vertical and horizontal coordinates of the segment endpoints rescaled to $[0, 1]$.

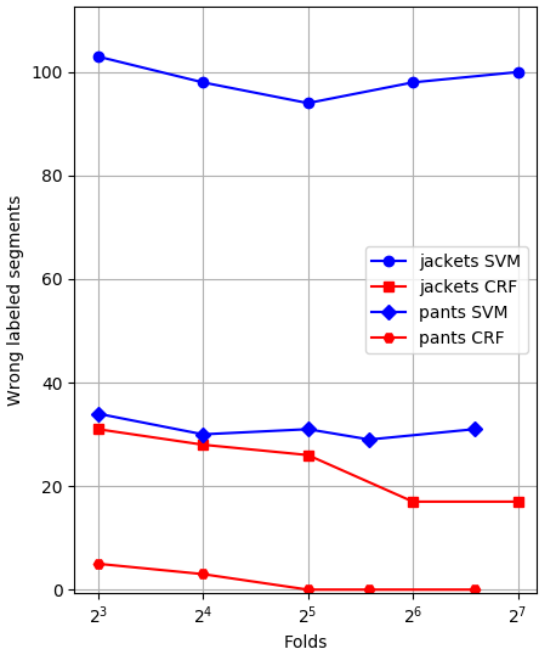

Figure 10: Accuracy of labeling depending on the number of folds. Last points correspond to leave-one-out.

Table 2: Regression methods and their objective functions.

Method	Norm of data loss	Coefficients regularizer	Objective function
Least squares	l_2	none	$\|\|y - Xw\|\|_2^2$
Ridge regression	l_2	l_2	$\|\|y - Xw\|\|_2^2 + \alpha\|\|w\|\|_2^2$
SVR	ε-insensitive l_1	l_1	$\max\{0, \|\|y - Xw\|\|_1 - \varepsilon\} + \alpha\|\|w\|\|_1$
Lasso	l_2	l_1	$\|\|y - Xw\|\|_2^2 + \alpha\|\|w\|\|_1$

But then, which type of regression is better to apply? According to the former reasoning, classic least squares regression is a good candidate. However, we have come across two problems. One is the well known sensitivity of least squares to outliers, which do happen in our problem when segments are misclassified sometimes. The other one is that it is not reasonable to impose, as least squares does, a Gaussian prior on the coefficients which implies that all of them should have more or less similar magnitude. On the contrary, for all measures only a few distances matter. Even a unique distance, like the length in pants or the wrist and sleeve in jackets, because of the way we have divided the contour.

What we want to avoid, for the sake of generality, is to impose from which few distances to regress each measure. Instead, we want to automatically learn it. This is possible through the addition of a L1 regularization term to the regression cost that imposes sparsity to the learned coefficients. Hence, Lasso regression fits well into our problem. We have assessed the accuracy of four regression methods: least squares with no regularization, ridge regression, Lasso [25] and support vector regression [26], whose objective functions are shown in table 2. Table 3 compares their accuracy in terms of the uncertainty (standard deviation of error) for each measure independently. We see there our intuition confirmed: Lasso regression is the best method for the majority of measures and also in average. In addition, in figure 11 we note that for each pants measure its regressor employs just a few non-negligible coefficients, and that these are those weighting the distances the most related to the measure in question.

5. Results

The final outcome of the method is an estimation of a list of dimensional measurements, each provided by a learned regressor. In order to make the most of the dataset, consisting on 130 jackets and 98 pants, we have adopted a leave-one-

Table 3: Uncertainty in cm for several regression techniques. For least squares we first removed the outliers from the regressed values.

Apparel	Measure	Least squares	Ridge regression	SVR	Lasso
Jacket	chest	3.12	2.51	2.46	**2.38**
	length	2.51	1.70	2.00	**1.63**
	waist	1.71	1.66	1.49	**1.48**
	hip	2.37	2.14	**2.11**	2.13
	shoulder	2.16	1.75	1.66	**1.56**
	sleeve	1.45	1.07	1.09	**0.99**
	wrist	0.91	**0.70**	0.77	0.71
	biceps	1.90	1.27	1.36	**1.19**
Pants	length	1.06	0.95	1.02	**0.93**
	bottom	1.14	1.03	1.00	**1.00**
	hip	2.54	2.43	**2.09**	2.20
	waist	1.95	1.66	**1.56**	1.63
	crotch	**1.98**	2.04	2.21	2.07
	thigh	2.06	**1.96**	2.27	2.10
	mean	1.92	1.63	1.65	**1.57**

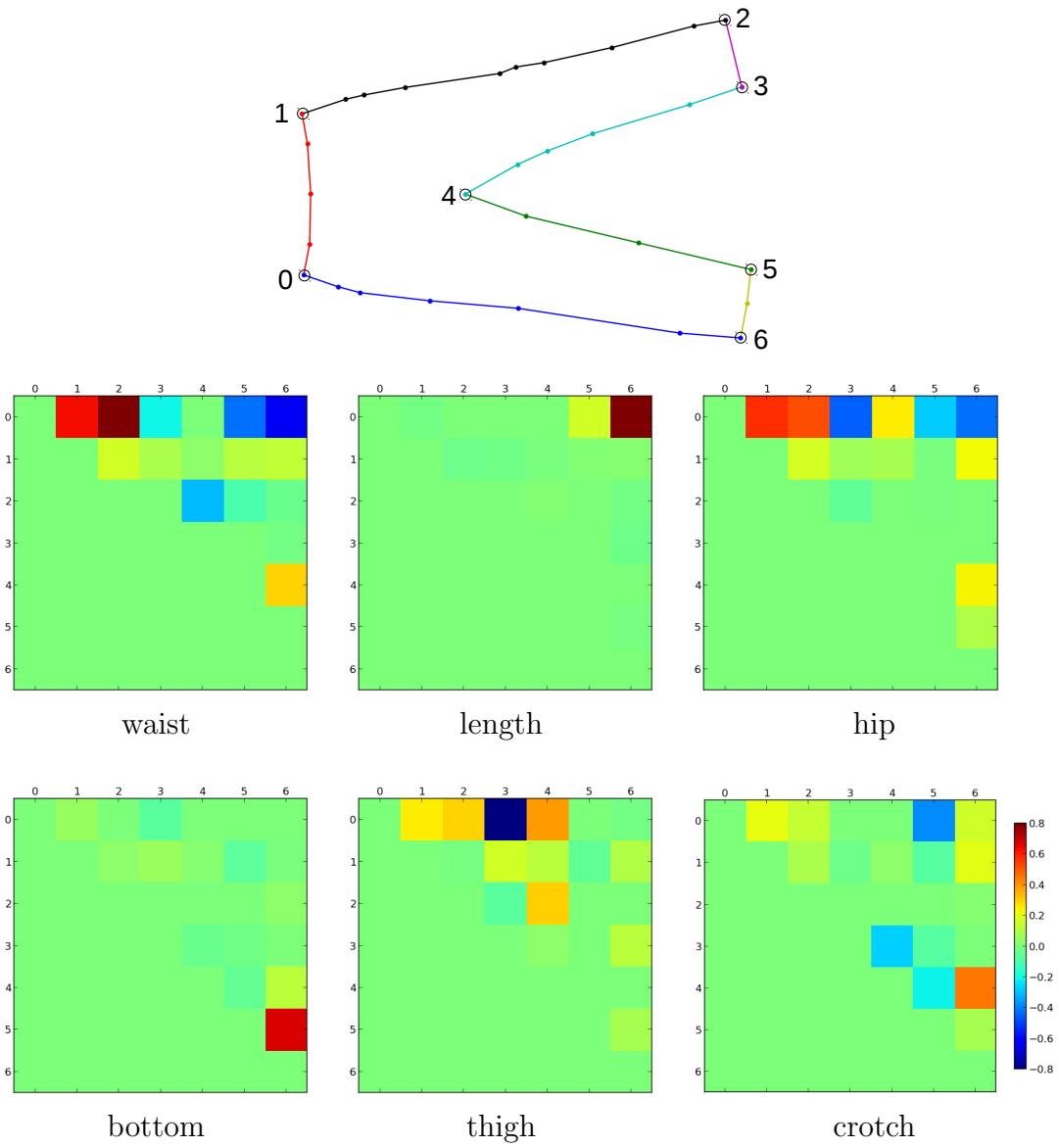

Figure 11: Lasso coefficients for each measure of pants. Each coefficient weights the distance between the starting points of two parts. On the axis, point numbers as enumerated on the top plot.

out cross-validation strategy: all the samples but one are used to train the CRF for segment labeling and then the Lasso regressors. The remaining one is used for testing, that is, to assess the labeling and compute the regression error. And this process is repeated for every sample.

We assess the accuracy by computing four kinds of statistics on the error, the difference between the groundtruth measure obtained by hand and that estimated by the regressor. The first two are the percentage of samples with an absolute and relative error below certain margins. The third and fourth ones are the uncertainty and the mean absolute error (MAE), with which we try to summarize the accuracy into a single number. Table 4 shows them all, for each measure and garment type. Overall, the estimation performs well both for jackets and pants. We can appreciate that for all measures but three, the relative error is less than 10% in more than 95% of the samples. For the remaining three (biceps, crotch, hip) it is around 90% of samples. For a tighter margin of less than 5% error, we still get the vast majority of samples, being the worst case the wrist, with about 70% of samples. One possible reason is that it is the shortest one, and the groundtruth has a resolution of 0.5 cm for many samples.

One may wonder if all the errors are equally important with regard to how well the garment will fit to the customer. It turns out that no, being the chest, sleeve and shoulder the most important measures for jackets, and waist and length for pants. All of them are well estimated but chest, for which we get only 77% of samples with an error less than 5%, and the largest uncertainty and MAE of jackets measures, 2.38 and 1.83 cm, respectively. The reason is that this longitudinal dimension has no endpoints coincident with any contour part being thus more complex to estimate than others, but also it is the more ambiguous measure when one gets it manually.

Finally, and as mentioned above, we want to compare these results with those obtained with the non-structured classification of segments, that is, using the SVM classifier and the same regression technique. Even though the classification accuracy was close to that of the CRF, we can see when comparing tables 5 and 4 that for SVM *both* the uncertainty and MAE for all measures are worse by a significant margin. Similarly, all means of absolute and relative error percentages under all margins are also worse. This comes to support our hypothesis that structured prediction pays in the end.

Table 4: Results of Lasso regression for 130 jackets and 98 pants after labeling with the CRF classifier.

Apparel	Measure	Percent samples with absolute error less than			Percent of samples with relative error less than			Uncer-tainty	Mean absolute error
		2 cm	4 cm	6 cm	2.5%	5%	10%	(cm)	(cm)
Jacket	chest	66.2	92.3	97.7	46.9	76.9	98.5	2.38	1.83
	length	76.2	97.7	100.0	76.9	96.9	100.0	1.63	1.28
	waist	81.5	99.2	99.2	65.4	93.8	100.0	1.48	1.14
	hip	68.5	92.3	99.2	53.1	80.0	99.2	2.13	1.66
	shoulder	81.5	97.7	100.0	63.1	87.7	98.5	1.56	1.18
	sleeve	96.2	100.0	100.0	93.1	100.0	100.0	0.99	0.81
	wrist	99.2	100.0	100.0	40.8	66.9	96.2	0.71	0.57
	biceps	87.7	100.0	100.0	34.6	68.5	89.2	1.19	0.92
	mean	82.1	97.4	99.5	59.2	83.8	97.7	**1.51**	**1.17**
Pants	length	96.9	100.0	100.0	99.0	100.0	100.0	0.93	0.72
	bottom	95.9	98.0	100.0	44.9	82.7	96.9	1.00	0.70
	hip	69.4	90.8	98.0	58.2	81.6	96.9	2.20	1.66
	waist	83.7	96.9	100.0	67.3	86.7	98.0	1.63	1.17
	crotch	70.4	90.8	100.0	35.7	67.3	90.8	2.07	1.61
	thigh	73.5	94.9	96.9	43.9	69.4	92.9	2.10	1.46
	mean	81.6	95.2	99.2	58.2	81.3	95.9	**1.66**	**1.22**

Table 5: Same as in table 4 but for the SVM segment classifier. 6 jackets (4.6%) could not be processed because some parts of the contour were not detected at all in them.

Apparel	Measure	Percent samples with absolute error less than			Percent of samples with relative error less than			Uncer-tainty	Mean absolute error
		2 cm	4 cm	6 cm	2.5%	5%	10%	(cm)	(cm)
Jacket	chest	63.8	88.5	93.1	46.9	73.1	93.8	2.43	1.83
	length	65.4	83.8	86.2	65.4	83.1	92.3	3.75	2.08
	waist	73.1	90.0	93.8	56.9	85.4	93.8	2.40	1.47
	hip	60.8	83.1	90.8	47.7	71.5	90.0	3.02	2.04
	shoulder	71.5	93.1	94.6	55.4	80.8	93.8	1.75	1.34
	sleeve	87.7	95.4	95.4	85.4	94.6	95.4	1.11	0.88
	wrist	94.6	95.4	95.4	36.2	61.5	92.3	0.73	0.60
	biceps	86.9	95.4	95.4	33.1	64.6	85.4	1.16	0.90
	mean	75.5	90.6	93.1	53.4	76.8	92.1	2.04	1.39
Pants	length	92.9	98.0	98.0	96.9	98.0	99.0	1.62	0.90
	bottom	93.9	96.9	99.0	40.8	78.6	94.9	1.32	0.84
	hip	67.3	86.7	94.9	57.1	78.6	92.9	2.69	1.89
	position	79.6	90.8	92.9	65.3	82.7	91.8	2.80	1.61
	crotch	63.3	84.7	94.9	32.7	60.2	84.7	3.23	2.11
	thigh	71.4	88.8	92.9	39.8	65.3	88.8	3.19	2.66
	mean	78.1	91.0	95.4	55.4	77.2	92.0	2.48	1.67

6. Conclusions

We have presented a method to obtain measures from a single image of a garment for size checking. Its main advantage is that, differently from past works, it is not highly dependent on the garment shape and the good localization of key points from which the measures are defined. Instead, it is able to understand the garment contour, that is, it classifies its segments into a set of meaningful contour parts from which a set of regressors estimate the measures. The parameters of both processes, contour segment labeling and regression, are learned from training samples and thus the method is directly extensible to new shapes other than jackets and pants. The only thing it needs is a partition of the contour into classes and groundtruth in the form of segment class annotations and values for the new measures.

We have analyzed the contribution on the result of the structured prediction through a CRF and the Lasso regression, and compared it to non-structured prediction with SVM and other types of non-sparse regressors in order to justify our method. The results show a high accuracy for all measures for the two types of garment considered, with a mean absolute error of 1.17 and 1.22 cm for jackets and pants, respectively.

The present success of deep learning for structured prediction may rise the question of what would be the performance with such types of architectures. We believe that graphical models are better suited to our problem that deep neural networks (DNN), at least for the segment labeling part. DNNs typically have a large number of parameters and part of their success relies on the availability of large datasets to learn them well. Ours is not that large, in the order of thousands of segments that means only hundreds of samples. Another reason for their success is that they learn good representations for the data and the task at hand. However, we show here that just five very simple and sensible features —segment endpoint coordinates and angle— lead to a very high labeling accuracy. Hence, as future work we do not envisage learning better representations but to apply the present method on new types of garments common in the apparel industry like shirts, skirts, coats etc. and also to other kinds of objects, like human profiles.

Acknowledgements

We kindly thank H. Codina and A. Benítez for providing samples, groundtruth of measures and feedback.

References

[1] E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, J.-D. Legat, A survey on industrial vision systems, applications and tools, Image and Vision Computing 21 (2) (2003) 171–188. doi:10.1016/S0262-8856(02)00152-X.

[2] Z. Bi, L. Wang, Advances in 3D data acquisition and processing for industrial applications, Robotics and Computer-Integrated Manufacturing 26 (5) (2010) 403–413. doi:10.1016/j.rcim.2010.03.003.

[3] N. Gonalves, V. Carvalho, M. Belsley, R. M. Vasconcelos, F. O. Soares, J. Machado, Yarn features extraction using image processing and computer vision a study with cotton and polyester yarns, Measurement 68 (2015) 1–15. doi:10.1016/j.measurement.2015.02.010.

[4] G. Peng, Z. Zhang, W. Li, Computer vision algorithm for measurement and inspection of o-rings, Measurement 94 (2016) 828–836. doi:10.1016/j.measurement.2016.09.012.

[5] E. Gadelmawla, Computer vision algorithms for measurement and inspection of external screw threads, Measurement 100 (2017) 36–49. doi:10.1016/j.measurement.2016.12.034.

[6] R. Xiang, W. He, X. Zhang, D. Wang, Y. Shan, Size measurement based on a two-camera machine vision system for the bayonets of automobile brake pads, Measurement 122 (2018) 106–116. doi:10.1016/j.measurement.2018.03.017.

[7] X. Wang, J. Liu, S. Liu, P. Jin, T. Wu, Z. Wang, Accurate radius measurement of multi-bend tubes based on stereo vision, Measurement 117 (2018) 326–338. doi:10.1016/j.measurement.2017.12.009.

[8] W. Wong (Ed.), Applications of Computer Vision in Fashion and Textiles, Elsevier, 2018.

[9] C. Keist, Quality control and quality assurance in the apparel industry, in: R. Nayak, R. Padhye (Eds.), Garment Manufacturing Technology, Woodhead Publishing Series in Textiles, Woodhead Publishing, 2015, pp. 405–426. doi:10.1016/B978-1-78242-232-7.00016-3.

[10] A. Goyal, Automation in fabric inspection, in: R. Nayak, R. Padhye (Eds.), Automation in Garment Manufacturing, The Textile Institute Book Series, Woodhead Publishing, 2018, pp. 75–107. doi:10.1016/B978-0-08-101211-6.00004-5.

[11] T. Gries, V. Lutz, V. Niebel, M. Saggiomo, K. Simonis, Automation in quality monitoring of fabrics and garment seams, in: R. Nayak, R. Padhye (Eds.), Automation in Garment Manufacturing, The Textile Institute Book Series, Woodhead Publishing, 2018, pp. 353–376. doi:10.1016/B978-0-08-101211-6.00014-8.

[12] H. Y. Ngan, G. K. Pang, N. H. Yung, Automated fabric defect detection—A review, Image and Vision Computing 29 (7) (2011) 442 – 458. doi:10.1016/j.imavis.2011.02.002.

[13] L. Norton-Wayne, A. Mackellar, C. Nicklin, Measurement of garment dimensions using machine vision, in: Third International Conference on Image Processing and its Applications, IET, 1989, pp. 197–201.

[14] K. Chen, Image analysis technology in the automatic measurement of garment dimensions, Asian Journal of Information Technology 4 (9) (2005) 832–834. doi:ajit.2005.832.834.

[15] L. Cao, Y. Jiang, M. Jiang, Automatic measurement of garment dimensions using machine vision, in: Computer Application and System Modeling (ICCASM), 2010 International Conference on, Vol. 9, IEEE, 2010, pp. V9–30. doi:10.1109/ICCASM.2010.5623093.

[16] C. Li, Y. Xu, Y. Xiao, H. Liu, M. Feng, D. Zhang, Automatic measurement of garment sizes using image recognition, in: Proceedings of the International Conference on Graphics and Signal Processing, ACM, 2017, pp. 30–34.

[17] T. Uhm, H. Park, J.-I. Park, Fully vision-based automatic human body measurement system for apparel application, Measurement 61 (2015) 169–179. doi:10.1016/j.measurement.2014.10.044.

[18] E. Aguilera-Aguilera, A. Carmona-Poyato, F. Madrid-Cuevas, R. Medina-Carnicer, The computation of polygonal approximations for 2d contours based on a concavity tree, Journal of Visual Communication and Image Representation 25 (8) (2014) 1905–1917. doi:10.1016/j.jvcir.2014.09.012.

[19] D. K. Prasad, M. K. Leung, C. Quek, S.-Y. Cho, A novel framework for making dominant point detection methods non-parametric, Image and Vision Computing 30 (11) (2012) 843–859. doi:10.1016/j.imavis.2012.06.010.

[20] A. Gray, E. Abbena, S. Salamon, Modern Differential Geometry of Curves and Surfaces with Mathematica, 3rd Edition, Chapman and Hall, 2006.

[21] S. Nowozin, C. H. Lampert, Structured learning and prediction in computer vision, Foundations and Trends in Computer Graphics and Vision 6 (34) (2011) 185–365. doi:10.1561/0600000033.

[22] I. Tsochantaridis, T. Joachims, T. Hofmann, Y. Altun, Large margin methods for structured and interdependent output variables, Journal of Machine Learning Research 6 (2005) 1453–1484.

[23] T. Joachims, T. Hofmann, Y. Yue, C.-N. Yu, Predicting structured objects with support vector machines, Communications of the ACM 52 (11) (2009) 97–104. doi:10.1145/1592761.1592783.

[24] S. Lacoste-Julien, M. Jaggi, M. Schmidt, P. Pletscher, Block-coordinate Frank-Wolfe optimization for structural SVMs, in: S. Dasgupta, D. McAllester (Eds.), Proceedings of the 30th International Conference on Machine Learning, Vol. 28 of Proceedings of Machine Learning Research, PMLR, Atlanta, Georgia, USA, 2013, pp. 53–61.

[25] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, 2nd Edition, Springer-Verlag, 2009.

[26] A. J. Smola, B. Schölkopf, A tutorial on support vector regression, Statistics and Computing 14 (3) (2004) 199–222. doi:10.1023/B:STCO.0000035301.49549.88.