CLOTH EXTREMITY DETECTION FROM A CLUTTER OF CLOTHES USING BAYESIAN GP-LVM

*Vishal Gaurav[a], Nishanth Koganti[a,b], Riku Nakata [a], Ravi P. Joshi[a] and Tomohiro Shibata[a]
([a]Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology)
([b]Graduate School of Information Science, Nara Institute of Science and Technology)

1. INTRODUCTION

Clothing assistance and laundry handling are basic assistance activity in the daily life of the elderly and disabled. However handling highly non-linear objects such as clothes is a cumbersome task despite of the current technological advancements. This field has drawn quite attention from robotic researchers. But they focused on how to handle this non-rigid object and detect it in a scene[1][2]. Grasping clothes from a pile is challenging [3]. The success rate for such tasks becomes improbable on the account of the non-rigid nature, high dimensionality, occlusions, self-overlaps etc., can make detection and planning for robotic manipulators an intricate issue. Recent technological advancement has propelled the demand of intelligent robots for assisted living.

Manipulation of clothes for clothing assistance by robots is much complex than just picking [4][5]. It requires detection and tracking of cloth extremities e.g. collar, sleeves, hemline so forth. The initialization part was out of scope in both the studies. In this study, we consider a scenario where we have single or multiple clothing items on a flat surface, as depicted in Fig. 1. Detection and grasping of clothes from a scene is accomplished in many studies. But only a few studies have performed cloth extremity detection [6], even though it is a basic task for cloth state estimation for robotic manipulation of clothes.

2. PROPOSED METHODOLOGY

A framework of the proposed method is depicted in Fig. 2. We used the Bayesian GP-LVM [7] because it is data efficient and is used for nonlinear dimensionality reduction. Using Bayesian GP-LVM we tried to learn the non-rigid nature, occlusion, overlaps and so forth, for detection of extremities. We used three different inputs i.e. raw RGB, Graysclae and histogram of SIFT [8] features to train the Bayesian GP-LVM model(depth information was not used in training the classifier). We compared the performance of the classifier with these input features. In this study, we consider the clothing article to be non-rigid, occluded and entangled. We also assume if one or more clothing article is placed randomly on a surface then one or more cloth's extremities will be visible.

2·1 Nonlinear Latent Variable Models

For modelling the non-rigid nature, occlusions, overlaps and to detect the extremities present in the clutter of clothes. Conventional approaches e.g. SVM [9], all the features (dimension) are considered relevant. However, for Bayesian GP-LVM, the relevant features are learned

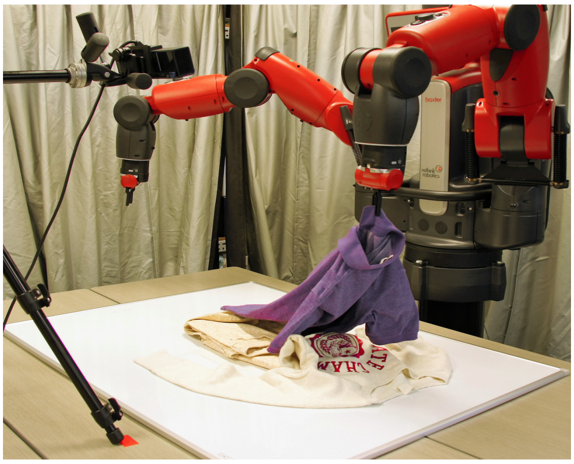

Fig.1 Depicts the grasping of detected cloth extremity in a clutter of clothes

in a Bayesian manner dependent on the training dataset. The extraction of such task-specific latent features is extremely challenging for clothes [6]. The generative model is as follows:

$$y = f(x) + \varepsilon \tag{1}$$

where y is observation space, f is a nonlinear GP mapping, ε is Gaussian noise which includes the uncertainty in the model. Assume $Y \in \mathbb{R}^{N\times D}$ be the observed data, where N is the number of observed samples and the observation space is D dimensional. Each data vector has an association with the latent variable $X \in \mathbb{R}^{N\times Q}$, for dimensionality reduction $Q \ll D$. From [10], assuming GPs to be i.i.d. across features, the likelihood is given as follows:

$$p(Y \mid X) = \prod_{d=1}^{D} p(y_d \mid X) \tag{2}$$

where y_d represents the d^{th} column of Y and

$$p(y_d \mid X) = \mathcal{N}(y_d \mid 0, K_{NN} + \beta^{-1} I_N) \tag{3}$$

where K_{NN} is the $N \times N$ covariance matrix derived from the kernel (or we can say the covariance function) $k(x,x')$ and β the inverse variance parameter. The marginal likelihood is calculated from Eq. (4):

$$p(Y) = \int p(Y \mid X) p(X) dX \tag{4}$$

In Bayesian training both latent mapping and latent space are integrated out. Exact Bayesian inference is intractable

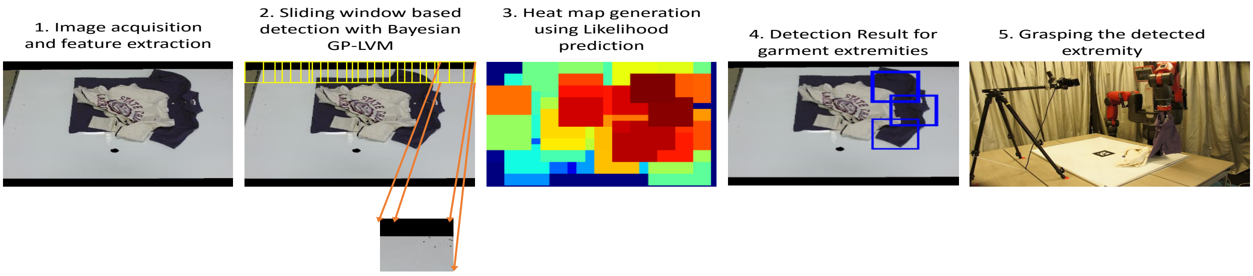

Fig.2 Depicts the Framework for cloth extremity detection in a clutter of clothes

[7] because $\log p(y_d \mid X)$ contains the latent variable X in highly nonlinear manner inside the inverse kernel matrix i.e. $K_{NN} + \beta^{-1} I_N$. So, in order to make this a tractable problem, Variational inference is used [11]. Maximizing variational lower bound formulates a Bayesian training technique that is robust to over-fitting and is able to autonomously select the dimension of the nonlinear latent manifold. Gaussian processes allows us to have a Bayesian non-parametric model for inferring highly nonlinear latent function from observation space. This model is used widely in field of Machine learning for solving complex task such as nonlinear regression or classification [7][10].

Automatic dimensionality reduction is performed using ARD Kernel[12] which is given by:

$$k(x,x') = \sigma_f^2 exp(-\frac{1}{2}\sum_{q=1}^{Q} w_q (x_q - x'_q)^2) \quad (5)$$

From Eq. (2), we can say that it is likelihood function of multiple-output Gaussian process regression. The various independent output have the same GP prior with latent input X. The density function for prior of X is given by

$$p(X) = \prod_{n=1}^{N} \mathcal{N}(x_n \mid 0, I_Q) \quad (6)$$

here x_n is n^{th} row of X. Therefore we can get the joint distribution for GP-LVM as follows:

$$p(Y,X) = p(Y \mid X)p(X) \quad (7)$$

Hyperparameters are the kernel parameters $\theta = (\sigma_f^2, w_1, ..., w_Q)$ and the inverse variance parameter β. We used the variational approach for marganalizing the latent variable X, for optimizing the resultant lower bound on the marginal likelihood with respect to the hyperparameters, as proposed in [7]. Use the lower bound for model comparison and dimensionality selection for latent space.

2·2 Test Inference

For predicting the class label of unseen data we compute the probability density $p(y^* \mid Y)$ of some unseen test data vector $y^* \in \mathbb{R}^D$, it can have missing data. For computing this probability function we can use the model as an estimator which can illustrate class conditional distribution in a classification system.

To approximate the probability function $p(y^* \mid Y)$, introduce the latent variable X and x^*, latent variable with respect to the output variable Y and y^*. From Eq. (4),

$$p(y^* \mid Y) = \frac{\int p(y^*, Y \mid X, x^*) p(X, x^*) dX dx^*}{\int p(Y \mid X) p(X) dX} \quad (8)$$

This equation gives a ratio between marginal likelihoods. The denominator term is already computed using the variational lower bound during training of Bayesian GP-LVM. We just need to approximate the integral in the numerator with the same procedure as performed in training. The difference will gives the maximum value and is selected for the particular class.

2·3 Cloth's Extremity Detection

We used the dataset provided by [6] for cloth extremity detection. The images contains 6 different clothing article (shirt, t-shirt, polo, jeans, sweater, and dress). They have manually annotated the bounding boxes for cloth extremities. The class labels for annotation are shirt color, shirt sleeves, t-shirt color, t-shirt sleeves, jeans hip, jeans hemline, polo color, polo sleeves, sweater hood, sweater sleeves, dress color. They have also provided a mask for each image but we have not used the mask because we want to develop a generalized model for cloth extremities detection. We divided the dataset in three categories: 1) Easy dataset containing images with single clothing item and low wrinkles and occlusions. 2) Hard dataset containing images with single clothing item but highly wrinkled and self-occlusion and non-canonical viewpoint. 3) Complete dataset containing all the images together with single and multiple clothes with high wrinkles, self-occlusion and entanglement.

As pointed out in [13], training a common model may not lead to learning a separable latent space. We can use Bayesian GP-LVM as Bayes classifier because it uses full Bayesian treatment. So, we trained a Bayes classifier with two separate Bayesian GP-LVM models for Positive (cloth extremity) and Negative (not extremity) class for each dataset. A likelihood of bounding box containing a cloth extremity was evaluated using Eq. (8). A comparison of likelihood was performed for each patch. The patches having higher values for positive class (i.e. having cloth extremity) were assigned to a positive class label. A heat map was generated for the input image using the compared likelihood. Choosing the number of peaks is difficult because it is dependent on what type of clothing item is present in a particular image, occlusion, overlaps, etc. So, we selected three peaks with less than 50 percent overlap as a detected part.

2·4 Grasping Detected Extremity

To validate our proposed method we performed an open loop grasping to the detected cloth extremity using Baxter dual arm robot. The 2D location of the selected peaks were then transferred to the world coordinates of the robot using the PCL library. A grasp was validated using the force sensor information after each grasping trial.

3. RESULTS AND DISCUSSION

3·1 Result for Cloth's Extremity Detection

Fig. 3 depicts the precision of the Bayes classifier. Fig. 4 depicts the confusion matrix for the Bayes classifier. For detection of cloth's extremities we used sliding window approach. Patches of 100×100 is generated with a step size of 20. Likelihood are calculated and comparison is performed for positive or negative classes as discussed in section 2.3. We selected highest three likelihood valued bounding boxes as candidate extremity. An example for cloth's extremity detection is shown in Fig. 5. The result depicts that the model trained for SIFT histogram was able to detect both the white polo collar and pink polo collar efficiently; whereas color and Grayscale features failed in this test image. For each input, image is scanned and a likelihood map is generated. In our experiment an extremity detection is considered true detection only if the centre of detected bounding box is inside the ground truth annotated bounding box. To evaluate the performance, we used recall at N(i.e.$R@N$), which gives us the understanding of the effectiveness of the method to correctly detect a cloth's extremity in an image if present by assessing only the N highest likelihood. This measure is pertinent for robotic application, as robots are capable of acknowledging only a few option in its planning. Also, the aspect (appearance) of the garment will change after each manipulation. The recall for garment extremity detection is depicted in Fig. 6. The recall for color feature depicts that the model was unable to learn an informative manifold to separate the non-linear high dimensional data. Grayscale features were more informative than the color features. Contrary to these features the SIFT histogram was able to find a separable manifold and detect the garment extremities. As we can see, the recall for easy, hard and complete is consistent for SIFT histogram. This is possibly because the latent space learnt by the model is same in all the scenario.

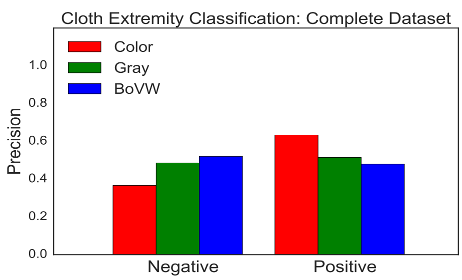

Fig.3 Depicts the precision of Binary Classification

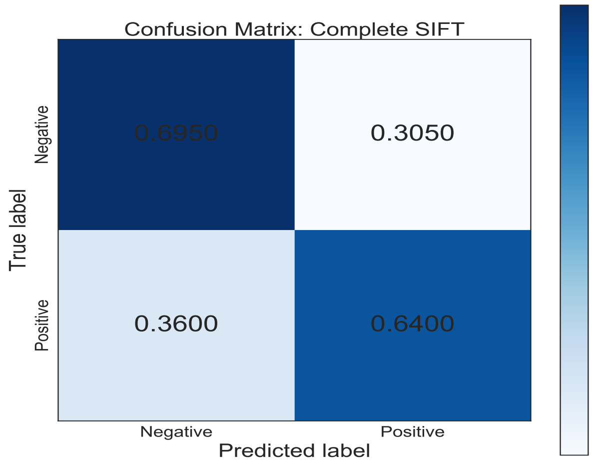

Fig.4 Depicts the confusion matrix for Binary Classification

4. CONCLUSION AND FUTURE WORK

In this study, we proposed a method to solve ambitious task of detecting cloth's extremities from a clutter of clothes under occlusion, overlaps and so forth using Bayesian GP-LVM. In conventional classification algorithms, every dimension is accounted for classification or detection task. Contrary to conventional algorithms, Bayesian GP-LVM learns relevant features in Bayesian fashion dependent on training dataset. Task specific feature extraction is extremely challenging for clothes [6]. We performed a comparative evaluation of three different input, i.e. raw color, Grayscale and histogram of SIFT features. From results Fig. 3, we devise that the task of classification is ambitious. Also in Fig. 4, we can say that

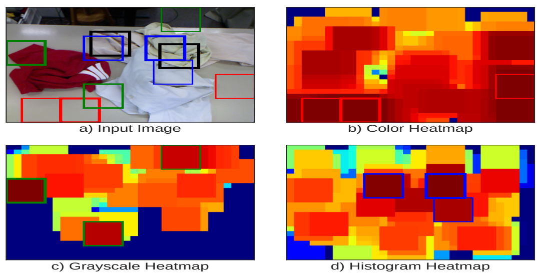

Fig.5 Detection example of cloth extremities for complete dataset. In a) The detected cloth's extremities are plotted. Black, Red, Green and Blue box denotes ground truth, color, Grayscale and SIFT histogram results respectively. We selected top three likelihood values. Thickness of bounding boxes depict the order of detection in decreasing order. Fig. b, c and d depicts the heat map for detection of cloth's extremities for color,Grayscale scale and SIFT histogram respectively. The three peak values in each heat map is shown in a coloured bounding box.

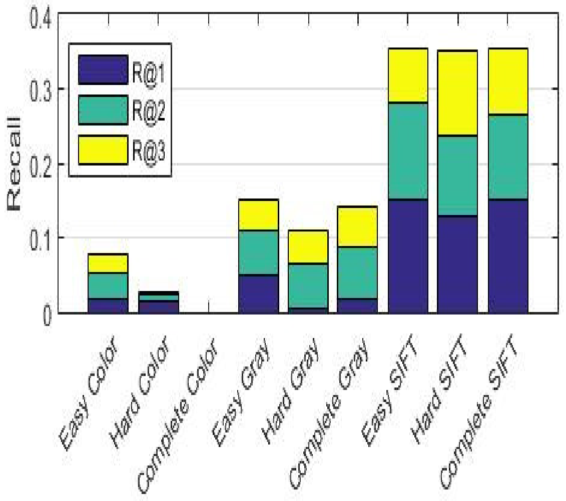

Fig.6 Depicts the recall at N(R@N). The recall is accumulated together (using different color) in a column for each dataset.

the model is being confused as the appearance of the positive and negative samples are akin. Form the results Fig. 6 we conclude that single shot detection of cloth's extremities is still a difficult task. The histogram of SIFT features showed an overall higher performance. The deformation affects the precision and recall values for classification.

Future work includes classification of the detected garment extremity by utilizing Manifold Relevance Determination(MRD) [14] to learn a shared manifold using the depth feature together with appearance features. Also, it includes use of robot proprioceptive information in MRD for interactive perception of clothing articles. The long term goal is manipulation of clothing articles initially on a flat surface to perform clothing assistance task.

ACKNOWLEDGEMENT

This work was supported in part by the Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science (No.16H01749).

References

[1] Jeremy Maitin-Shepard, Marco Cusumano-Towner, Jinna Lei and Pieter Abbeel: "Cloth Grasp Point Detection based on Multiple-view Geometric Cues with Application to Robotic Towel Folding" in *IEEE International Conference on Robotics and Automation (ICRA), pp. 2308-2315,* 2010.

[2] Marco Cusumano-Towner, Arjun Singh, Stephen Miller, James F. O'Brien and Pieter Abbeel: "Bringing Clothing into Desired Configurations with Limited Perception" in *IEEE International Conference on Robotics and Automation (ICRA), pp. 3893–3900,* 2011.

[3] David Fischinger, Astrid Weiss and Markus Vincze: "Learning Grasps with Topographic Features" in *The International Journal of Robotics Research, 34(9), pp. 1167–1194,* SAGE Publications, 2015.

[4] Tomoya Tamei, Takamitsu Matsubara, Akshara Rai and Tomohiro Shibata, "Reinforcement Learning of Clothing Assistance with a Dual-arm Robot" in *Proc. of the 11th IEEE-RAS Int. Conf. on Humanoid Robots, pp. 733-738,* 2011.

[5] Nishanth Koghanti, Jimson G. Ngeo, Tomoya Tamei, Takamitsu Matsubara and Tomohiro Shibata, "Cloth Dynamics Modeling in Latent Spaces and its Application to Robotic Clothing Assistance" in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3464–3469,* 2015.

[6] Arnau Ramisa, Guillem Alenyà, Francesc Moreno-Noguer and Carme Torras: "Learning RGB-D Descriptors of Garment Parts for Informed Robot Grasping" in *Engineering Applications of Artificial Intelligence, 35, pp. 246–258,* Elsevier, 2014.

[7] Michalis K. Titsias and Neil D. Lawrence: "Bayesian Gaussian Process Latent Variable Model" in *AISTATS, pp. 844–851,* 2010.

[8] David G. Lowe: "Distinctive Image Features from Scale-invariant Keypoints" in *International journal of computer vision, 60(2), pp. 91–110,* Springer, 2004.

[9] Corinna Cortes and Vladimir Vapnik: "Support-Vector Networks" in *Machine Learning, 20(3), pp. 273–297,* Springer, 1995.

[10] Neil D. Lawrence, "Gaussian Process Latent Variable Models for Visualisation of High Dimensional Data" in *Advances in Neural Information Processing Systems, 16(3), pp. 329-336,* 2004.

[11] Matthew James Beal: "Variational Algorithms for Approximate Bayesian Inference" in *PhD. Thesis, University of London,* 2003.

[12] Carl Edward Rasmussen: "Gaussian Processes for Machine Learning" in *MIT press,* 2006.

[13] Andreas C. Damianou, Michalis K. Titsias and Neil D. Lawrence: "Variational Inference for Latent Variable and Uncertain Inputs in Gaussian Processes" in *Journal of Machine Learning Research, 17(42), pp. 1–62,* 2016.

[14] Andreas C. Damianou, Carl Ek, Michalis K. Titsias and Neil D. Lawrence: "Manifold Relevance Determination" in *arXiv preprint arXiv:1206.4610,* 2012.