

# Structured Low-Rank Matrix Approximation in Gaussian Process Regression for Autonomous Robot Navigation

Eunwoo Kim, Sungjoon Choi, and Songhwa Oh

**Abstract**—This paper considers the problem of approximating a kernel matrix in an autoregressive Gaussian process regression (AR-GP) in the presence of measurement noises or natural errors for modeling complex motions of pedestrians in a crowded environment. While a number of methods have been proposed to robustly predict future motions of humans, it still remains as a difficult problem in the presence of measurement noises. This paper addresses this issue by proposing a structured low-rank matrix approximation method using nuclear-norm regularized  $l_1$ -norm minimization in AR-GP for robust motion prediction of dynamic obstacles. The proposed method approximates a kernel matrix by finding an orthogonal basis using low-rank symmetric positive semi-definite matrix approximation assuming that a kernel matrix can be well represented by a small number of dominating basis vectors. The proposed method is suitable for predicting the motion of a pedestrian, such that it can be used for safe autonomous robot navigation in a crowded environment. The proposed method is applied to well-known regression and motion prediction problems to demonstrate its robustness and excellent performance compared to existing approaches.

## I. INTRODUCTION

We are witnessing service robots appearing in public places, offices, hospitals and homes interacting with humans by performing routine tasks, such as performing household chores and delivering medicine and supplies. In the near future, more service robots will be assisting and cooperating with humans in many dynamic and complex real-world environments. In such environments, it is difficult to operate successfully without the exact prediction of dynamic obstacles or moving humans. Since the safe operation is an important requirement for the success of service robots, an ability to predict motions of pedestrians and moving objects is of paramount importance [1].

For safe navigation of a mobile robot under a dynamic and crowded environment, autonomous robot navigation has been studied extensively in recent years [2]–[9] and it is required to predict the trajectories of pedestrians or moving objects precisely. In [2], a probabilistic model based on a partially observable Markov decision process is used to predict trajectories of pedestrians for autonomous robot navigation. Fulgenzi et al. [3] proposed a motion pattern model of pedestrians using a Gaussian process (GP). Wang et

al. [10] proposed Gaussian process dynamical models and its applications to learning models of human motion. Henry et al. [4] proposed an inverse reinforcement learning based approach for human-like navigation in a crowded environment. Trautman et al. [7] developed a novel cooperative navigation approach using a GP and conducted the first trial of robot navigation in a crowded cafeteria. In general, it is assumed that the current positions of a robot and moving obstacles are available [8] or can be estimated from an external device [7]. However, such assumption makes existing approaches impractical in many practical environments since collecting exact locations using an external device can be a costly option and available only in a laboratory setting.

Recently, Choi et al. [9] proposed an autoregressive Gaussian process (AR-GP) to model a complex motion of a pedestrian from the egocentric view of a mobile robot. AR-GP is capable of capturing complex human motions using Gaussian process regression (GPR), a nonparametric regression method, whereas parametric models, e.g., a linear model, can only handle simple human motions [9]. This work was extended in [11] to a robust AR-GP motion model by removing the effects of measurement noises and outliers in the training set using low-rank kernel matrix approximation based on the  $l_1$ -norm. While the approximation method proposed in [11] shows the robustness against outliers, it can fail to find a feasible solution since the algorithm does not guarantee the positive semi-definiteness of its solution, which approximates the kernel matrix in AR-GP. Since an incorrect estimation of the kernel matrix can lead to an unstable situation when a robot navigates using AR-GP, it is necessary to approximate a kernel matrix while keeping its structure of positive semi-definiteness.

In this paper, we propose a novel structured low-rank matrix approximation, which finds a low-rank solution of a symmetric positive semi-definite matrix using nuclear-norm regularized  $l_1$ -norm minimization, for robust autoregressive Gaussian process regression (AR-GP). The proposed method approximates a kernel matrix used in AR-GP using its low-rank kernel approximation, assuming that the kernel matrix can be represented using a small number of dominating principal components, eliminating outliers and erroneous aspects in the training data set. The proposed method is applied to motion prediction problems to demonstrate its robustness against unwanted corruptions. Furthermore, the method is applied in a physical experiment using a Pioneer 3DX mobile robot and a Microsoft Kinect camera for motion prediction and autonomous robot navigation.

The remainder of this paper is organized as follows: In

This work was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2013R1A1A2065551) and by the ICT R&D program of MSIP/IITP (14-824-09-014, Basic Software Research in Human-Level Lifelong Machine Learning).

E. Kim, S. Choi, and S. Oh are with the Department of Electrical and Computer Engineering, ASRI, Seoul National University, Seoul 151-744, Korea (e-mail: {eunwoo.kim, sungjoon.choi, songhwa.oh}@cpslab.snu.ac.kr).

Section II, we briefly review low-rank matrix approximation and Gaussian process regression. In Section III, we propose a structured low-rank matrix approximation algorithm. Then, we present various experimental results to evaluate the proposed method in Section IV.

## II. PRELIMINARIES

### A. Low-rank matrix approximation

Low-rank matrix approximation is a minimization problem, in which the cost function measures the fit between an observation matrix and an approximating matrix, subject to the constraint that the approximating matrix has a reduced rank. The problem arises in a number of problems in machine learning and computer vision, such as image denoising, collaborative filtering, background modeling, and data reconstruction, to name a few [12], [13].

Let us consider the  $l_2$  approximation of matrix  $G$ . The problem is to minimize the following cost function for given  $G$ :

$$\arg \min_{P, X} \|G - PX\|_F, \quad (1)$$

where  $G \in \mathbb{R}^{m \times n}$ ,  $P \in \mathbb{R}^{m \times r}$ , and  $X \in \mathbb{R}^{r \times n}$  are the observation, projection, and coefficient matrices, respectively. Here,  $r$  is a predefined parameter less than  $\min(m, n)$  and  $PX$  is a low-rank approximation of  $G$ . However, the  $l_2$  based approximation is highly sensitive to non-Gaussian noises. To overcome the disadvantage, methods based on the  $l_1$ -norm have been emerged in many fields [12]–[14].

There is another family of approaches using the recent advances in nuclear-norm minimization which is also known as robust principal component analysis (RPCA) [14]. RPCA decomposes the observation matrix into a low-rank matrix and a sparse matrix by solving the  $l_1$ -norm regularized nuclear-norm minimization problem:

$$\begin{aligned} \min_{D, E} & \|D\|_* + \lambda \|E\|_1 \\ \text{s.t.} & G = D + E, \end{aligned} \quad (2)$$

where  $D$ ,  $E$ , and  $G$  are low-rank, sparse error, and observation matrices, respectively. Here, the nuclear-norm of a matrix is the sum of its singular values, i.e.,  $\|\Sigma\|_* = \sum_i \sigma_i$ , where  $\sigma_i$  is a singular value of  $\Sigma$ . RPCA has recently achieved many successful results in machine learning and computer vision [14], [15].

### B. Gaussian process regression

A Gaussian process (GP) is a collection of random variables which has a joint Gaussian distribution and is specified by its mean function  $m(\mathbf{x})$  and covariance function  $k(\mathbf{x}, \mathbf{x}')$  [16]. A Gaussian process  $f(\mathbf{x})$  is expressed as:

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \quad (3)$$

Suppose that  $\mathbf{x} \in \mathbb{R}^n$  is an input and  $y_i \in \mathbb{R}$  is an output. For a noisy observation set  $D = \{(\mathbf{x}_i, y_i) | i = 1, \dots, n\}$ , we can consider the following observation model:

$$y_i = f(\mathbf{x}_i) + w_i, \quad (4)$$

where  $w_i \in \mathbb{R}$  is a zero-mean Gaussian noise with variance  $\sigma_w^2$ . Then the covariance of  $y_i$  and  $y_j$  can be expressed as

$$\text{cov}(y_i, y_j) = k(\mathbf{x}_i, \mathbf{x}_j) + \sigma_w^2 \delta_{ij}, \quad (5)$$

where  $\delta_{ij}$  is the Kronecker delta function which is 1 if  $i = j$  and 0 otherwise.  $k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$  is a covariance function based on some nonlinear mapping function  $\phi$ . The function  $k$  is also known as a kernel function.

We can represent (5) in a matrix form as follows:

$$\text{cov}(\mathbf{y}) = K + \sigma_w^2 I, \quad (6)$$

where  $\mathbf{y} = [y_1 \dots y_n]^T$  and  $K$  is a kernel matrix such that  $[K]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ .

The conditional distribution of a new output  $y_*$  at a new input  $\mathbf{x}_*$  given  $D$  becomes

$$y_* | D, \mathbf{x}_* \sim \mathcal{N}(\bar{y}_*, \mathbb{V}(y_*)), \quad (7)$$

where

$$\bar{y}_* = k_*^T (K + \sigma_w^2 I)^{-1} \mathbf{y} = k_*^T \Lambda \mathbf{y}, \quad (8)$$

where  $\Lambda = (K + \sigma_w^2 I)^{-1}$  and the covariance of  $y_*$  is

$$\mathbb{V}(y_*) = k(\mathbf{x}_*, \mathbf{x}_*) - k_*^T (K + \sigma_w^2 I)^{-1} k_*. \quad (9)$$

Here,  $k_* \in \mathbb{R}^n$  is a covariance vector between the new data  $\mathbf{x}_*$  and existing data, such that  $[k_*]_i = k(\mathbf{x}_*, \mathbf{x}_i)$ . Note that when it comes to make a prediction given a collected training set, the computational cost of GP can be reduced by pre-computing the inverse of a kernel matrix [9].

An autoregressive Gaussian process (AR-GP) is a method to predict future positions of a moving object given a finite number of past positions [9], [11], which is applied in this paper for predicting future motions of pedestrians.

Kim et al. [11] proposed a low-rank kernel matrix approximation algorithm using the relationship between GPR and low-rank kernel matrix approximation and showed that the robustness of the proposed method in the presence of outliers and measurement noises. However, the approximated matrix in [11] is not a proper kernel matrix since the positive semi-definiteness is not assumed in the algorithm. In the next section, we propose a structured low-rank approximation algorithm which guarantees the positive semi-definiteness of its solution.

## III. THE PROPOSED METHOD

### A. Formulation

In this section, we propose a structured low-rank matrix approximation method for approximating a kernel matrix by making sure that the approximated matrix is positive semi-definite. For robustness of the proposed method in the presence of erroneous data, we use robust measures in a cost function. Instead of methods based on the  $l_2$ -norm, which as known to be highly sensitive to outliers, the proposed method is based on the robust principal component analysis (RPCA) framework [14] to reduce the effect of outliers with an automatic rank search. Hence, we approximate a kernel matrix using a nuclear-norm regularized  $l_1$ -norm minimization problem for robust approximation.

We formulate the problem of nuclear-norm regularized  $l_1$ -norm minimization as shown below:

$$\min_{P, M} \|K - PMP^T\|_1 + \lambda \|PMP^T\|_*, \quad (10)$$

where  $K \in \mathbb{R}^{n \times n}$  is a kernel or symmetric positive semi-definite matrix and  $P \in \mathbb{R}^{n \times r}$  and  $M \in \mathbb{R}^{r \times r}$  are optimization variables.  $\|\cdot\|_*$  denotes the nuclear-norm or trace-norm, and  $\lambda > 0$  is a regularization parameter. In the cost function, we use the nuclear-norm regularizer to minimize the rank of  $PMP^T$ , an approximation of  $K$ , to our desired one by adjusting the parameter  $\lambda$  since we do not know the exact rank. The nuclear-norm has been used as a convex surrogate for the rank in many rank minimization problems [14], [15]. This problem is non-convex and its solution can be obtained using the augmented Lagrangian framework [14].

To reduce the computational complexity and make the convergence faster, it is reasonable to enforce an orthogonality constraint to the basis matrix  $P$  by shrinking the solution space of  $P$ . Based on these observations, we reformulate the low-rank matrix approximation problem as follows:

$$\begin{aligned} \min_{P, M} \quad & \|K - PMP^T\|_1 + \lambda \|M\|_* \\ \text{s.t.} \quad & P^T P = I_r, M \succeq 0, \end{aligned} \quad (11)$$

where  $I_r$  is an  $r \times r$  identity matrix and  $M$  is a positive semi-definite matrix. By enforcing the orthogonal constraint on  $P$ , we can compute only small matrix  $M$  instead of  $PMP^T$  when computing the nuclear-norm. Figure 1 shows an overview of the proposed structured low-rank matrix approximation method. Due to the difficulty of solving the problem (11) directly, we introduce two auxiliary variables  $D$  and  $\widehat{M}$  and solve the following problem:

$$\begin{aligned} \min_{P, M, D, \widehat{M}} \quad & \|K - D\|_1 + \lambda \|M\|_* \\ \text{s.t.} \quad & D = P\widehat{M}P^T, \widehat{M} = M, P^T P = I_r, M \succeq 0. \end{aligned} \quad (12)$$

The augmented Lagrangian framework [14] is used to solve (12) by converting the constrained optimization problem into the following unconstrained problem:

$$\begin{aligned} \mathcal{L}(K, P, M, D, \widehat{M}) = & \|K - D\|_1 + \lambda \|M\|_* \\ & + \text{tr} \left( \Lambda_1^T (D - P\widehat{M}P^T) \right) + \text{tr} \left( \Lambda_2^T (\widehat{M} - M) \right) \\ & + \frac{\beta}{2} \left( \|D - P\widehat{M}P^T\|_F^2 + \|\widehat{M} - M\|_F^2 \right), \end{aligned} \quad (13)$$

where  $\Lambda_1, \Lambda_2 \in \mathbb{R}^{n \times n}$  are Lagrange multipliers and  $\beta > 0$  is a small penalty parameter. Here, we have not included the orthogonality constraint over  $P$ , but it is considered when we optimize (13) with respect to  $P$ . We apply the alternating minimization approach iteratively, which estimates one variable while other variables are held fixed. Each step of the proposed algorithm is described in the following section.

Proposed structured low-rank matrix decomposition

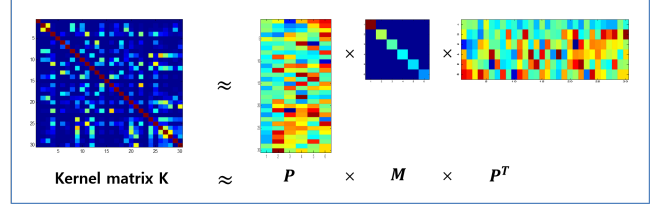


Fig. 1. A graphical illustration of the proposed method. A kernel matrix  $K$  can be approximated by multiplication of  $P$ ,  $M$ , and  $P^T$ . We can perform prediction of future motions of moving objects using AR-GP based on the rank reduced kernel matrix.

### B. Algorithm

To solve for  $M$ , we fix the other variables and solve the following optimization problem:

$$\begin{aligned} M_+ = & \arg \min_M \frac{\lambda}{\beta} \|M\|_* + \frac{1}{2} \left\| \widehat{M} - M + \frac{\Lambda_2}{\beta} \right\|_F^2, \\ = & \arg \min_M \frac{\lambda}{\beta} \|M\|_* + \frac{1}{2} \|M - A\|_F^2, \text{ s.t. } M \succeq 0, \end{aligned} \quad (14)$$

where  $A = \widehat{M} - \frac{\Lambda_2}{\beta}$ . If  $A$  is not a symmetric matrix, we make it a symmetric matrix by  $A \leftarrow \frac{A+A^T}{2}$  and find  $M_+$ . Then, this problem can be solved using eigenvalue thresholding (EVT) [17] and its solution is

$$M_+ = Q \text{diag} \left[ \max \left( \gamma - \frac{\lambda}{\beta}, 0 \right) \right] Q^T, \quad (15)$$

where  $Q$  and  $\Gamma$  are matrices, which contain eigenvectors and eigenvalues, respectively, from the eigenvalue decomposition of  $A$ , i.e.,  $A = Q\Gamma Q^T$  and  $\Gamma = \text{diag}(\gamma)$ .

For  $D$ , we solve the following problem:

$$\begin{aligned} D_+ = & \arg \min_D \|K - D\|_1 + \text{tr} \left( \Lambda_1^T (D - P\widehat{M}P^T) \right) \\ & + \frac{\beta}{2} \|D - P\widehat{M}P^T\|_F^2, \\ = & \arg \min_D \|K - D\|_1 + \frac{\beta}{2} \left\| D - P\widehat{M}P^T + \frac{\Lambda_1}{\beta} \right\|_F^2, \end{aligned} \quad (16)$$

and the solution can be computed using the shrinkage (soft-thresholding) operator [14]:

$$D_+ \leftarrow K - \mathcal{S} \left( K - P\widehat{M}P^T + \frac{\Lambda_1}{\beta}, \frac{1}{\beta} \right), \quad (17)$$

where  $\mathcal{S}(x, \tau) = \text{sign}(x) \cdot \max(|x| - \tau, 0)$  for a variable  $x$ .

With other variables fixed, we have the following optimization problem for finding  $P$ :

$$\begin{aligned} P_+ = & \arg \min_P \text{tr} \left( \Lambda_1^T (D - P\widehat{M}P^T) \right) + \frac{\beta}{2} \|D - P\widehat{M}P^T\|_F^2, \\ = & \arg \min_P \frac{\beta}{2} \left\| D + \frac{\Lambda_1}{\beta} - P\widehat{M}P^T \right\|_F^2, \text{ s.t. } PP^T = I_r. \end{aligned} \quad (18)$$

The above problem is a least square problem with an orthogonality constraint. Let  $R = D + \frac{\Lambda_1}{\beta}$  and  $L = P\widehat{M}$ ,

then  $L$  can be represented by  $L = R(P^T)^+ = R(P^T)^T = RP$ , where  $(P^T)^+$  is the pseudo-inverse of the matrix  $P^T$ . Therefore, from [18], we can obtain the orthogonal matrix  $P = \text{QR}(RP) = \text{QR}(L)$ , where  $\text{QR}(A)$  is the QR factorization of  $A$ .

To update  $\widehat{M}$ , we consider the following equation:

$$\begin{aligned} \widehat{M}_+ = \arg \min_{\widehat{M}} & \text{tr} \left( \Lambda_1^T (D - P\widehat{M}P^T) \right) + \text{tr} \left( \Lambda_2^T (\widehat{M} - M) \right) \\ & + \frac{\beta}{2} \left( \|D - P\widehat{M}P^T\|_F^2 + \|\widehat{M} - M\|_F^2 \right), \end{aligned} \quad (19)$$

and its solution is computed by taking a derivative as

$$\widehat{M}_+ = \frac{1}{2} \left( P^T D P + \frac{1}{\beta} P^T \Lambda_1 P + M - \frac{1}{\beta} \Lambda_2 \right). \quad (20)$$

Finally, we update the Lagrange multipliers  $\Lambda_1$  and  $\Lambda_2$  as follows:

$$\begin{aligned} \Lambda_1 &\leftarrow \Lambda_1 + \beta(D - P\widehat{M}P^T), \\ \Lambda_2 &\leftarrow \Lambda_2 + \beta(\widehat{M} - M). \end{aligned} \quad (21)$$

The proposed structured low-rank matrix approximation algorithm is summarized in Algorithm 1. Since it is a symmetric positive semi-definite matrix factorization algorithm, it is named as *factSPSD*. Similar to [11], we apply KPCA to the approximated low-rank kernel matrix after performing the algorithm for faster computation when GPR is applied, reducing the computational complexity from  $O(n^3)$  to  $O(rn^2)$ . In the algorithm, we have assumed a normalized observation matrix. Hence, the output matrices are obtained by rescaling them using the scaling factor. The alternating minimization order of optimization variables can be different, but we had empirically found that the order in Algorithm 1 showed better than other orders. We set the initial values to all zero matrices since the algorithm is not sensitive to the choice of initial values. We set the parameters in algorithm as  $\lambda = 10^{-3}$ ,  $\beta = 10^{-5}$ , and  $\rho = 2$ . The inner iteration of the algorithm (from line 5 to line 10) was set to 10 since it is enough to converge to a local solution. Although it is difficult to guarantee the convergence of the proposed method to a local minimum, the alternating optimization will converge to a finite limit since the cost function consists of non-negative functions. The stopping criterion (line 13 of Algorithm 1) is chosen as

$$\|D - P\widehat{M}P^T\|_1 < \epsilon \quad \text{or} \quad \|\widehat{M} - M\|_1 < \epsilon, \quad (22)$$

and  $\epsilon = 10^{-5}$ , which shows good results in our experiments.

#### IV. EXPERIMENTAL RESULTS

In this section, we evaluate the performance of the proposed method (*factSPSD*) by experimenting with various data sets and comparing with other sparse Gaussian process regression methods (SPGP<sup>1</sup> [19], PITC [20], GPLasso<sup>2</sup> [21], and PCGP- $l_1$  [11]) along with the standard full GP. In our

#### Algorithm 1 *factSPSD* for optimizing (12)

---

```

1: Input:  $K \in \mathbb{R}^{n \times n}$ , rank  $r$ ,  $\lambda$ ,  $\beta$ , and  $\rho$ 
2: Output:  $P \in \mathbb{R}^{n \times r}$ ,  $D \in \mathbb{R}^{n \times n}$ , and  $M \in \mathbb{R}^{r \times r}$ 
3: Initialization:  $M = P = D = \widehat{M} = 0$  and  $\beta_{max} = 10^{10}$ 
4: while not converged do
5:   while not converged do
6:     Update  $M$  by (15)
7:     Update  $P \leftarrow \text{QR}(RP)$  where  $R = D + \frac{\Lambda_1}{\beta}$ 
8:     Update  $\widehat{M}$  by (20)
9:     Update  $D$  by (17)
10:  end while
11:  Update the Lagrange multipliers  $\Lambda_1$  and  $\Lambda_2$  by (21)
12:  Update  $\beta = \min(\rho\beta, \beta_{max})$ 
13:  Check the convergence condition
14: end while

```

---

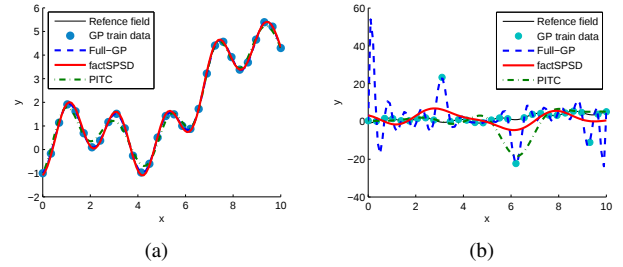


Fig. 2. Simulation results on a synthetic example. (a) No outliers. (b) 30% outliers.

experiments, we used the RBF Gaussian kernel for all GP methods and hyperparameters are learned using a conjugate gradient method [16]. The prediction or regression accuracy is measured by the root mean squared error (RMSE).

##### A. Regression problems

First, we tested on a synthetic example to compare the proposed structured low-rank matrix approximation with other Gaussian process methods [16], [20] in a regression problem with and without outliers. Since we are interested in how the proposed method performs in the presence of outliers, we have compared *factSPSD* to a sparse GP (PITC [20]) and the full GP [16].

Figure 2 shows the simulation results of the regression problem according to two outlier levels (no outliers and 30% outliers). As shown in Figure 2(a), full GP and the proposed method give nearly exact results when there are no outliers, whereas PITC does not fit the reference field exactly. So the proposed method shows its competitiveness compared with the other sparse GP methods. If we add some outliers as shown in Figure 2(b), the full GP and PITC try to fit outliers so they show fluctuations, but *factSPSD* is less affected by outliers than the full GP and PITC, showing its robustness against outliers. Although a kernel function can give an effect of smoothing, the effect of outliers still remain in the kernel matrix. From this experiment, we see the clear benefit of the proposed low-dimensional learning method to a regression problem when the train set contains outliers.

<sup>1</sup>Available at <http://www.gatsby.ucl.ac.uk/~snelson/>.

<sup>2</sup>Available at <https://www.cs.purdue.edu/homes/alanqi/softwares/softwares.htm>.

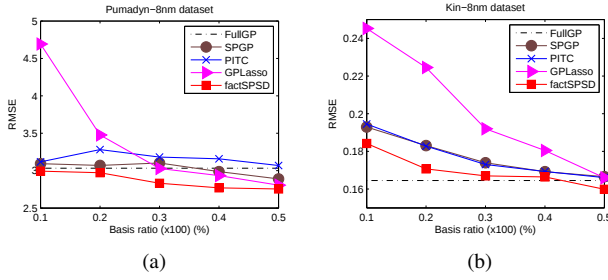


Fig. 3. Regression results of the proposed method compared with other GP methods according to basis conditions for two benchmark data sets: (a) Pumadyn-8nm, (b) Kin-8nm.

To verify the proposed algorithm for real-world data sets, we have tested algorithms using two well-known data sets, Pumadyn-8nm and Kin-8nm<sup>3</sup>, which are benchmark data sets in the Gaussian process regression literature [21]. Pumadyn-8nm is a data set which consists of puma forward dynamics of 8 inputs and Kin-8nm is a data set which consists of the forward kinematics of an 8 link robot arm. For each dataset, we randomly collected 1,000 training and 800 test samples from the data sets. To verify the robustness of the proposed method under the existence of various outliers, we added 30 percent outliers which are randomly selected from  $[-25, 25]$  in the data sets, whereas the data sets are in the range of  $[-2, 2]$ . The simulation results of the proposed method with other sparse GP methods (SPGP [19], PTC [20], and GPLasso [21]) for various basis ratios (10% ~ 50%) are shown in Figure 3. As shown in Figure 3(a), the proposed method gives the lowest error among the methods regardless of the basis conditions, especially, it shows better performance than full GP, whereas sparsity-based methods show lower error than full GP when the basis ratio is large. In Figure 3(b), the proposed method also gives lower errors than other sparsity-based methods. Although its performance is worse than full GP when the basis ratio is small, the difference is the smallest.

### B. Motion prediction of human trajectories

For an actual experiment, we collected trajectories of moving pedestrians using a Pioneer 3DX differential drive mobile robot and a Microsoft Kinect camera, which is mounted on top of the robot as shown in Figure 6. All algorithms are written in MATLAB using the mex-compiled ARIA package<sup>4</sup> on a notebook with a 2.1 GHz quad-core CPU and 8GB RAM. The position of a pedestrian is detected using the skeleton grab API for Kinect.

We performed experiments in our laboratory to predict the future position of a person. To model the future positions of a pedestrian, our algorithm is applied to an inversion of a kernel matrix in the autoregressive Gaussian process (AR-GP) motion model [9], which estimates the current position of a pedestrian based on  $p$  recent positions of the human

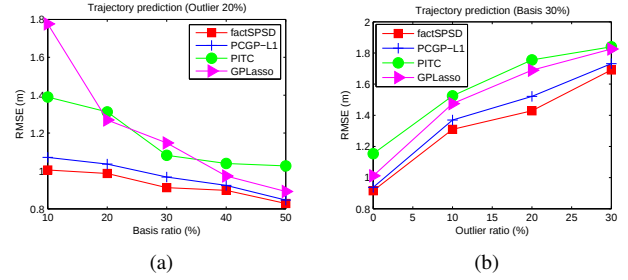


Fig. 4. Motion prediction results using Kinect-based human trajectories: (a) Various basis ratio with 30 percent outliers. (b) Various outlier ratio with 30 percent basis vectors.

by a nonlinear model of an autoregressive process under the Gaussian process framework. To make a training set from the collected trajectories, we uniformly sampled positions to have ten samples in a trajectory when a trajectory has many detected positions.

We compared the proposed method with state-of-the-art approaches (PCGP- $l_1$  [11], GPLasso [21], and PTC [20]) for the collected human trajectories. We divided the collected trajectories into training and test sets with autoregressive order  $p = 3$ . Using the data set, we have experimented for two cases: under various rank (basis) conditions with a fixed outlier level and under various outlier conditions with a fixed rank. We added outliers to randomly selected positions of collected trajectories from  $[-10, 10]$ , whereas the data sets are in the range of  $[-5, 5]$ . Figure 4 shows results for two cases. As shown in Figure 4(a), the proposed factSPSD shows the best results compared to other sparse GP methods for all cases. PCGP- $l_1$  gives the second best results regardless of the basis ratios. We can interpret that the proposed algorithm approximates a kernel matrix used in AR-GP better than PCGP- $l_1$ , since the proposed algorithm can guarantee the positive semi-definiteness, whereas PCGP- $l_1$  does not assume the positive semi-definiteness. The RMSE error results for a fixed rank ( $r/n \times 100 = 30\%$ ) under various outlier conditions are shown in Figure 4(b). As shown in the figure, the proposed method gives the best results regardless of outlier conditions. From two figures, we can see that the proposed method shows the robustness against outliers, by recovering from measurement noises and erroneous trajectories. Figure 5 shows some snapshots of a motion prediction experiment using two Microsoft Kinect cameras (about  $110^\circ$  field of view) in our laboratory.

### C. Motion control

We have applied the proposed method to a motion control problem. A non-parametric Bayesian motion controller can navigate through crowded dynamic environments [9] and the proposed method based AR-GP motion model is applied to the Gaussian process motion controller [9] for autonomous robot navigation. We performed the motion controller experiments using the Pioneer 3DX mobile robot and two Microsoft Kinect cameras, to enlarge the field of view of a robot, in various dynamic and crowded environments. The number of pedestrians varies from one to seven to verify

<sup>3</sup>Available at <http://www.cs.toronto.edu/~delve/methods/mars3.6-bag-1/mars3.6-bag-1.html>.

<sup>4</sup>Available at <http://robots.mobilerobots.com/wiki/ARIA>.

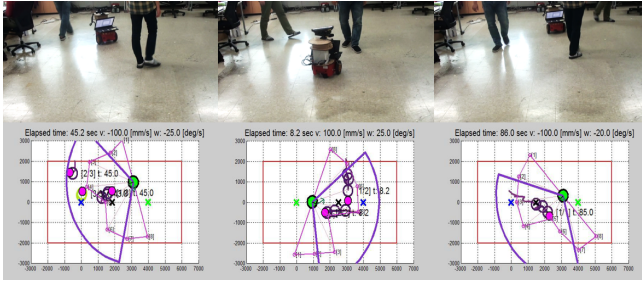


Fig. 5. A motion prediction experiment using the proposed algorithm. A pink circle represents a prediction of a pedestrian given the observed positions. Each experiment consists of a photo taken by camera and robot's internal state. Best viewed in color.



Fig. 6. Snapshots from a real experiment in a crowded school cafeteria. A pink circle represents a prediction of a pedestrian given the observed positions. Each experiment consists of photos taken by camera and robot's internal state. Best viewed in color.

the performance of the the proposed method under crowded environments. Figure 6 shows some snapshots from the experiments in a crowded school cafeteria. In the experiments, we have verified that the robot successfully avoided moving pedestrians and obstacles without any collisions and arrived at the goal position.

## V. CONCLUSION

In this paper, we have proposed a structured low-rank matrix approximation method using nuclear-norm regularized  $l_1$ -norm minimization and its application to robust autoregressive Gaussian process regression for autonomous robot navigation since modeling a complex pedestrian motion pattern is a difficult problem in the presence of measurement noises or outliers. To overcome the limitation of the state-of-the-art low-rank approximation method, we have presented a novel optimization formulation and its efficient algorithm to obtain a symmetric positive semi-definite matrix. The proposed method is applied to various well-known regression data sets and experiments using a Pioneer 3DX mobile robot and two Microsoft Kinect cameras. The experimental results show the robustness of the proposed method against outliers and sensor errors compared to existing methods.

## REFERENCES

- [1] C. H. Chen, Y. Weng, and S.-T. Sun, "Toward the human-robot coexistence society: on safety intelligence for next generation robots," *International Journal of Social Robotics*, vol. 1, no. 4, pp. 267–282, 2009.
- [2] A. F. Foka and P. E. Trahanias, "Predictive autonomous robot navigation," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2002.
- [3] C. Fulgenzi, C. Tay, A. Spalanzani, and C. Laugier, "Probabilistic navigation in dynamic environment using rapidly-exploring random trees and Gaussian processes," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2008.
- [4] P. Henry, C. Vollmer, B. Ferris, and D. Fox, "Learning to navigate through crowded environments," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2010.
- [5] R. Asaula, D. Fontanelli, and L. Palopoli, "Safety provisions for human/robot interactions using stochastic discrete abstractions," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010.
- [6] J. J. Park, C. Johnson, and B. Kuipers, "Robot navigation with model predictive equilibrium point control," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012.
- [7] P. Trautman, J. Ma, R. M. Murray, and A. Krause, "Robot navigation in dense human crowds: the case for cooperation," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2013.
- [8] G. S. Aoude, B. D. Luders, J. M. Joseph, N. Roy, and J. P. How, "Probabilistically safe motion planning to avoid dynamic obstacles with uncertain motion patterns," *Autonomous Robots*, no. 1, pp. 51–76, 2013.
- [9] S. Choi, E. Kim, and S. Oh, "Real-time navigation in crowded dynamic environments using Gaussian process motion control," in *Proc. of IEEE International Conference on Robotics and Automation (ICRA)*, 2014.
- [10] J. M. Wang and D. J. Fleet, "Gaussian process dynamical models for human motion," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 283–298, 2008.
- [11] E. Kim, S. Choi, and S. Oh, "A robust autoregressive gaussian process motion model using  $l_1$ -norm based low-rank kernel matrix approximation," in *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2014.
- [12] Q. Ke and T. Kanade, "Robust  $l_1$  norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2005.
- [13] E. Kim, M. Lee, C.-H. Choi, N. Kwak, and S. Oh, "Efficient  $l_1$ -norm-based low-rank matrix approximations for large-scale problems using alternating rectified gradient method," *IEEE Trans. on Neural Networks and Learning Systems (TNNLS)*, vol. 26, no. 2, pp. 237–251, 2015.
- [14] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, pp. 11:1–11:37, 2011.
- [15] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, pp. 471–501, 2010.
- [16] C. Rasmussen and C. Williams, *Gaussian process for machine learning*. MIT Press, 2006.
- [17] Y. Ni, J. Sun, X. Yuan, S. Yan, and L.-F. Cheong, "Robust low-rank subspace segmentation with semidefinite guarantees," in *Proc. of the IEEE International Conf. on Data Mining Workshops*, 2010, pp. 1179–1188.
- [18] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a non-linear successive over-relaxation algorithm," *Rice University CAAM Technical Report TR10-07*, 2010.
- [19] E. Snelson and Z. Ghahramani, "Sparse Gaussian processes using pseudo-inputs," in *Advances in Neural Information Processing Systems (NIPS)*, 2005.
- [20] J. Quinero-candela, C. E. Rasmussen, and R. Herbrich, "A unifying view of sparse approximate Gaussian process regression," *Journal of Machine Learning Research (JMLR)*, vol. 6, pp. 1939–1959, 2005.
- [21] F. Yan and Y. Qi, "Sparse Gaussian process regression via  $l_1$  penalization," in *Proc. of the International Conference on Machine Learning (ICML)*, 2010.