

# Mini-Batched Online Incremental Learning Through Supervisory Teleoperation with Kinesthetic Coupling\*

Hiba Latifee<sup>1</sup>, Affan Pervez<sup>2</sup>, Jee-Hwan Ryu<sup>1</sup> and Dongheui Lee<sup>3</sup>

**Abstract**—We propose an online incremental learning approach through teleoperation which allows an operator to partially modify a learned model, whenever it is necessary, during task execution. Compared to conventional incremental learning approaches, the proposed approach is applicable for teleoperation-based teaching and it needs only partial demonstration without any need to obstruct the task execution. Dynamic authority distribution and kinesthetic coupling between the operator and the agent helps the operator to correctly perceive the exact instance where modification needs to be asserted in the agent’s behaviour online using partial trajectory. For this, we propose a variation of the Expectation-Maximization algorithm for updating original model through mini batches of the modified partial trajectory. The proposed approach reduces human workload and latency for a rhythmic peg-in-hole teleoperation task where online partial modification is required during the task operation.

## I. INTRODUCTION

Since Ray Goertz first proposed a pantograph-based mechanical teleoperation system [1], there has been a lot of progress in teleoperation. In early days, teleoperation systems have been limited to applications where human’s physical access is limited, which includes deep underwater, outer space, and nuclear power plants [2]. However, the recent advancements in human-computer interfaces and artificial intelligence technology have broadened its application domains [3, 4] to more general areas including telerobotic surgery, treatments and diagnosis [5], robotic gripping/grasping [6, 7], search and rescue (SAR) [8], and robotic rehabilitation [9].

Teleoperation tasks often require highly trained operators and high mental workload as the complexity of the teleoperation task increases and more general applications get introduced [10]. Shared teleoperation [11], which combines local autonomy with direct teleoperation, can relieve operator’s mental burden by shifting some of the workload, especially related to repetitive tasks, from the operator to the local autonomy of the slave.

There were efforts to shift repetitive tasks from a human operator to an artificial agent by Learning from Demonstra-

tions (LfD) through teleoperation [12, 13, 14]. Although the LfD through teleoperation showed possibility of relieving operator’s workload at least for repetitive tasks, a slight change in the task or the environment requires retraining from the beginning, which can be time consuming and computationally expensive. The human operator in parallel with an agent can handle these variations in a task or the environment [15], hence avoiding task failure. But, due to the active participation of the human operator on the control layer with the agent, the issue of operator’s heavy mental workload still remains prevalent.

In such a case, incremental learning using human demonstrations helps in updating the already learned task [16, 17, 18]. In [19], the authors update the model incrementally through an iterative Expectation-Maximization (EM) algorithm, and in [20] local data points were utilized for incrementally learning a new model at each time step in a simulated environment. However, even for partially modifying the behaviour of the initially learned agent, a completely new full trajectory has been required and the model was updated offline once the full demonstration is terminated [21]. Moreover, as can be noted in the aforementioned studies, both the robot and the human demonstrator were physically co-located to incrementally update the learned model [22]. While [23] allows learned model’s autonomous refinement through shared teleoperation, but it requires multiple task demonstrations from the human operator to assert a modification which can be both a bit costly and cumbersome.

One of the major limitations of the conventional incremental learning approaches (even though most of them are limited to kinesthetic teaching) is that *they do not support updating the agent’s behaviour through a partial demonstration without any obstructions*. Especially, for a partial modification in repetitive tasks, e.g. when one bolting position out of ten is changed, model update on the fly would provide large benefit; the task does not need to be obstructed especially when it is highly time critical, also we can maintain the quality of already well-trained model. However, the impact of partial demonstration on incremental learning has been understudied, particularly for rhythmic/repetitive tasks. Oftentimes, a complete new demonstration is not even needed (as only a part of the motion is needed to be updated) or even available [13]. Although there was a research effort to carry out an online model refinement through kinesthetic teaching [24, 25], there have been no prior investigations into an online model update especially through a partial teleoperated demonstration without obstructions.

In this paper, we propose an online incremental learn-

\*This research is partially supported by the project “Toward the Next Generation of Robotic Humanitarian Assistance and Disaster Relief: Fundamental Enabling Technologies (10069072)” and Helmholtz Association.

<sup>1</sup>Hiba Latifee and Jee-Hwan Ryu are with the Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea. h.o.latifee@gmail.com, jhryu@kaist.ac.kr

<sup>2</sup>Affan Pervez is with Intech Process Automation, Lahore, Pakistan. affan.pervez@intechww.com

<sup>3</sup>Dongheui Lee is with the Department of Electrical and Computer Engineering, Technical University of Munich (TUM), Munich, Germany and also with Institute of Robotics and Mechatronics, German Aerospace Center (DLR), Germany. dhlee@tum.de

ing approach using mini batches of the partial teleoperated demonstration to run a single round of a modified EM algorithm to incrementally update a learned model for rhythmic tasks. In order to provide a tool for the human operator to supervise and partially modify a learned model online through teleoperation, we use Dynamic Authority Distribution (DAD) along with kinesthetic coupling between operator and agent [26] and modify the EM for mini-batched, localized retraining through partial demonstration. This allows handling slight changes in the task or the environment on the fly without any task obstructions. To the best of authors' knowledge, there has been no prior research on a teleoperation-based mini-batched incremental learning approach via DAD and kinesthetic coupling.

In addition to DAD, kinesthetic coupling helps the operator to correctly perceive the precise instance for modifying the agent's behaviour. Moreover with our approach, the operator's role is elevated to that of a supervisor, where he/she intervenes only when a need arises to modify the agent's motion.

The major contributions of this work are:

- proposing a novel incremental learning approach for partially retraining a rhythmic task on the fly without any obstructions in the task execution using mini batches of a partial teleoperated demonstration trajectory,
- modifying the EM algorithm where the Expectation step (E-step) includes the probability along the phase variable in order to account for the correct assignment of the responsibilities of the relevant Gaussian components, and finally
- reducing human operator's workload through supervisory teleoperation using DAD and kinesthetic coupling between human and the artificial agent.

## II. GENERATING INITIAL LEARNED MODEL

This section describes how we remotely teach a robot a task through teleoperation for generating an agent's behaviour. In this work, we encode the agent's initial behaviour using Gaussian Mixture Model (GMM), from a single teleoperated demonstration. A GMM with  $k$  components is parameterized by  $\Theta_{(k)} = \{\pi_m, \mu_m, \Sigma_m\}_{m=1}^k$  where  $\pi_1, \dots, \pi_k$  are mixing coefficients (priors),  $\mu_1, \dots, \mu_k$  are mean vectors and  $\Sigma_1, \dots, \Sigma_k$  are covariance matrices. The probability density function of  $\mathbf{X}$  is said to follow the  $k$ -component GMM if it can be written as  $\mathcal{P}(\mathbf{X}|\Theta_{(k)}) = \sum_{m=1}^k \pi_m \mathcal{N}(\mathbf{X}; \mu_m, \Sigma_m)$ , subject to the constraints  $0 < \pi_m < 1$  and  $\sum_{m=1}^k \pi_m = 1$ . The dataset for each Degree-of-Freedom (DOF) is created by concatenating the teleoperated demonstration trajectory with the phase variable:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \dots & \mathbf{x}_T \end{pmatrix} = \begin{pmatrix} s_1 & \dots & s_T \\ y_1 & \dots & y_T \end{pmatrix}$$

The phase variable  $s$  is an exponentially decaying function from 1 to 0 for a discrete motion and a linearly increasing value with a single phase cycle from 0 to  $2\pi$  for a rhythmic

motion. A single phase variable is utilized for synchronization of multiple DOFs. The standard EM algorithm is utilized for fitting the data [27]. EM is an iterative optimization approach for estimating the maximum likelihood value of the parameters of a probabilistic model with hidden variables. In the E-step, the expected values of the incomplete terms appearing in likelihood maximization are calculated by using the old parameters estimate:

$$h_{i,k}^{t+1} = \frac{\pi_k^t \mathcal{N}(\mathbf{x}_i; \mu_k^t, \Sigma_k^t)}{\sum_{l=1}^K \pi_l^t \mathcal{N}(\mathbf{x}_i; \mu_l^t, \Sigma_l^t)}$$

where  $h$  is referred to as the responsibility term. For a rhythmic motion,  $s$  is mapped in the interval  $[\mu_k^i - \pi, \mu_k^i + \pi]$  for calculating valid probabilities for the  $k^{th}$  GMM component. These expectations are then utilized for updating the model parameters in the Maximization step (M-step):

$$\begin{aligned} \pi_k^{t+1} &= \frac{\sum_{i=1}^T h_{i,k}^{t+1}}{T}, & \mu_k^{t+1} &= \frac{\sum_{i=1}^T h_{i,k}^{t+1} \mathbf{x}_i}{\sum_{i=1}^T h_{i,k}^{t+1}} \\ \Sigma_k^{t+1} &= \frac{\sum_{i=1}^T h_{i,k}^{t+1} (\mathbf{x}_i - \mu_k^{t+1})(\mathbf{x}_i - \mu_k^{t+1})^\top}{\sum_{i=1}^T h_{i,k}^{t+1}} + \epsilon \mathbf{I} \end{aligned}$$

where  $\epsilon \mathbf{I}$  is the regularization term for avoiding overfitting/singularities. If the  $\mu_k^{t+1}$  moves outside the  $[0, 2\pi]$  range, it is wrapped back in the interval  $[0, 2\pi]$ . The EM iterations continue until the change in log-likelihood becomes very small.

Now for retrieving  $y_i$  for a given  $s_i$  value, Gaussian Mixture Regression (GMR) [28] is utilized, where  $y_i$  is the GMR output. In GMR, input and output variables for each component are represented separately:

$$\mu_k = \begin{bmatrix} \mu_k^I \\ \mu_k^O \end{bmatrix}, \Sigma_k = \begin{bmatrix} \Sigma_k^{II} & \Sigma_k^{IO} \\ \Sigma_k^{OI} & \Sigma_k^{OO} \end{bmatrix}$$

For a given input variable  $\mathbf{x}^I$ , the expected value of  $\mathbf{x}^O$  is calculated as:

$$\begin{aligned} E(\mathbf{x}^O | \mathbf{x}^I) &= \sum_{k=1}^K h_k \hat{\mathbf{x}}_k \\ \text{with } h_k &= \frac{\pi_k \mathcal{N}(\mathbf{x}^I; \mu_k^I, \Sigma_k^{II})}{\sum_{l=1}^K \pi_l \mathcal{N}(\mathbf{x}^I; \mu_l^I, \Sigma_l^{II})} \\ \hat{\mathbf{x}}_k &= \mu_k^O + \Sigma_k^{OI} (\Sigma_k^{II})^{-1} (\mathbf{x}^I - \mu_k^I) \end{aligned}$$

## III. MINI-BATCHED INCREMENTAL LEARNING THROUGH TELEOPERATION

This section describes our proposed online incremental learning method through teleoperation. In order to modify an already learned agent's behaviour using mini batches of a partial trajectory on the fly, we have implemented a modified version of DAD method with kinesthetic coupling.

### A. Dynamic Authority Distribution Between Human and Agent

This subsection explains the vital tool for human operator to modify a learned motion partially whenever it is required, by allowing human to take control from agent temporarily to define a partial demonstration through supervisory teleoperation. For this, we modify DAD [26] in a novel human-agent teleoperation setting. In DAD, control authority over the slave is dynamically distributed.

In this paper, the DAD asserts that the human drives the authority flow in case he/she wants to take authority or give it back to the agent. To identify the intention of human, we introduce a one-way energy-monitoring concept, observing inflows and outflows of energy from the operator. For each time interval  $\Delta T$ , we monitor:

$$E_i^{op}(k) = \sum_{n=1}^k \vec{F}_i^{op}(n) \vec{V}_i^{op}(n) \Delta T$$

$$\Delta E_i^{op}(k) = E_i^{op}(k) - E_i^{op}(k-1)$$

where  $\vec{F}_i^{op}$ ,  $\vec{V}_i^{op}$  are the force and velocity values of the master device respectively,  $i$  is the axis of motion,  $E_i^{op}(k)$  is the cumulative energy and  $\Delta E_i^{op}(k)$  is the energy at time step  $k$ . At each  $k$ ,  $\Delta E_i^{op}(k)$  determines whether the operator is generating energy or not. Whenever the operator starts generating energy, this translates to the physical sense, that the human wants to gain control in order to modify the current region of agent's motion through a partial demonstration. Alternatively, when the operator's energy starts reducing, this translates to his/her intention to give up authority over slave, thereby indicating the end of the modification:

$$n_i = \begin{cases} n_i + 0.01 & \text{if } \Delta E_i^{op} > 0 \\ n_i - 0.01 & \text{if } \Delta E_i^{op} < 0 \\ n_i & \text{otherwise} \end{cases}$$

To update the operator's authority based on the aforementioned energy-monitoring, adaptation law in [26] is applied:

$$\gamma_i = 0.5 + 0.4 \times \tanh(n_i) \quad (1)$$

where  $\gamma_i$  is the operator's dominance in the  $i^{th}$  axis. Equation (1) keeps  $\gamma_i$  within (0.1, 0.9) for a fast converging and slowly returning control transition between human and agent. The desired control input to the slave  $\vec{V}_{sd}$  having DAD is:

$$\vec{V}_{sd} = W_\gamma \vec{V}_{op} + (I - W_\gamma) \vec{V}_{Agent}$$

where  $W_\gamma = \text{diag}\{\gamma_x, \gamma_y, \gamma_z\}$ ,  $\vec{V}_{op}$  and  $\vec{V}_{Agent}$  represent velocities of operator and agent, respectively (see Fig. 1).

### B. Kinesthetic Coupling

This subsection explains the role of kinesthetic coupling. In order to modify the agent's original behaviour on the fly, the operator needs to understand the agent's intended motion correctly and to perceive the right instance to assert and assent control. For this, not only visual but also haptic feedback is essential. In our DAD, when the human operator is controlling the slave, then the interaction force from the

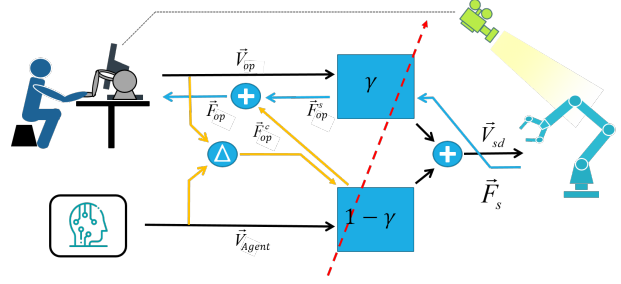


Fig. 1: DAD and kinesthetic coupling between human and agent. The black, blue and yellow arrows depict the flow of control, the feedback and kinesthetic coupling forces for the human, respectively. The red dotted line represents DAD.

slave is fully felt by the human. But when the human is simply following the agent's behaviour by yielding to its force, then the agent's coupling force is felt by the human and there is no significant interaction force felt by the human. This is shown in (2). This way, operator is able to assess how much force to assert to gain control of the slave when sharing control with the agent. In other words, the operator's dominance,  $W_\gamma$  is directly proportional to the interaction force from the slave and inversely proportional to the agent's kinesthetic coupling force:

$$\vec{F}_{op} = \vec{F}_{op}^c + \vec{F}_{op}^s = (\vec{V}_{Agent} - \vec{V}_{op})(I - W_\gamma) + \vec{F}_s W_\gamma \quad (2)$$

where the interaction force,  $\vec{F}_s$ , is fed back to operator as  $\vec{F}_{op}^s$  (see Fig. 1).  $\vec{F}_{op}^c$  is the kinesthetic coupling force that gives human substantial hints about agent's motion to allow him/her to modify the task precisely in regions where intended, without distorting the whole motion.

### C. Modified EM Algorithm for Partial Demonstration

This subsection describes our modified EM algorithm, in order to retrain the learned model based on the partial information of trajectory provided by the human whenever the dominance goes beyond a prescribed threshold (0.7).

As mentioned earlier, existing EM approaches for incremental learning require a complete new demonstration for updating the learned agent's parameters. EM is an iterative process which is computationally expensive. That is why existing EM-based approaches have been applied only for discrete tasks. In addition to these limitations of existing retraining approaches, there can also be numerical instability issues in encoding while using partial data. In conventional LfD, the GMM components are usually concentrated around the initial demonstration and it is possible that during the incremental learning, the new data points coming through the partial trajectory are far away from the already learned GMM. Since Gaussians are exponentially decaying functions, there's a possibility that for the new data points they have almost equal or near to zero probability density function (pdf) values for all Gaussians. This could lead to an equal responsibility assignment for new data points (modified trajectory) in the conventional incremental learning setting.

For our incremental learning algorithm, we propose a variation of the EM algorithm for taking partial teleoperated demonstration to update learned model parameters without any interruptions during task execution. Firstly, we only execute a single iteration of our modified EM algorithm for each small batch of newly recorded partial data points. Due to the single iteration of the proposed EM, we can perform the model updates in an online manner, while the system continues its operation. Secondly, we propose a modified responsibility term calculation in the E-step, where the probability along the input variable (phase variable) is also added in the responsibility term. Hence, the modified E-step considers the sum of 2 Gaussian distributions. This additional term - pdf value for the input variable  $x_i^I$  in (3) - helps in assigning the new data points to the relevant Gaussian components, which is closest in terms of the phase variable, and can also help to deal with the case of significantly deviated new demonstrations:

### E-step

$$h_{i,k}^{t+1} = \frac{\pi_k^t (\mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_k^t, \boldsymbol{\Sigma}_k^t) + \mathcal{N}(x_i^I; \boldsymbol{\mu}_k^{I,t}, \boldsymbol{\Sigma}_k^{I,t}))}{\sum_{l=1}^K \pi_l^t (\mathcal{N}(\mathbf{x}_i; \boldsymbol{\mu}_l^t, \boldsymbol{\Sigma}_l^t) + \mathcal{N}(x_i^I; \boldsymbol{\mu}_l^{I,t}, \boldsymbol{\Sigma}_l^{I,t}))} \quad (3)$$

where  $\mathcal{N}(x_i^I; \boldsymbol{\mu}_k^{I,t}, \boldsymbol{\Sigma}_k^{I,t})$  is the pdf of the phase variable  $s$ .

In the M-step, the model parameters are updated using the subsets of the new data points. We only update means and covariances of the learned model, since priors cannot be updated given only partial spatial information of the whole trajectory. The model parameters are updated by utilizing a weighted combination of the original model parameters and the additional information conveyed by the new data points:

### M-step

$$\boldsymbol{\mu}_k^{t+1} = \frac{\sum_{i=1}^T h_{i,k}^{t+1} \mathbf{x}_i + \alpha \boldsymbol{\mu}_k^t}{\sum_{i=1}^T h_{i,k}^{t+1} + \alpha}$$

$$\boldsymbol{\Sigma}_k^{t+1} = \frac{\sum_{i=1}^T h_{i,k}^{t+1} (\mathbf{x}_i - \boldsymbol{\mu}_k^{t+1})(\mathbf{x}_i - \boldsymbol{\mu}_k^{t+1})^\top + \alpha \boldsymbol{\Sigma}_k^t}{\sum_{i=1}^T h_{i,k}^{t+1} + \alpha} + \epsilon \mathbf{I}$$

where  $\alpha$  is the mixing weight for the original model.

## IV. EXPERIMENTS

For the experimental evaluation of our proposed approach, the setup consists of a 3-DOF SensAble PHANToM Omni as a master device and a 3-DOF PHANToM Premium 1.5A as a slave device. The pair of master and slave devices run on the same computer and are operated at a control frequency of 1kHz. A web-camera streams the visual feedback from the slave environment to the human operator, as illustrated in Fig. 2. The camera position and the lack of depth information in the two dimensional camera images introduce a perspective distortion which inhibits a clear visual perception for the

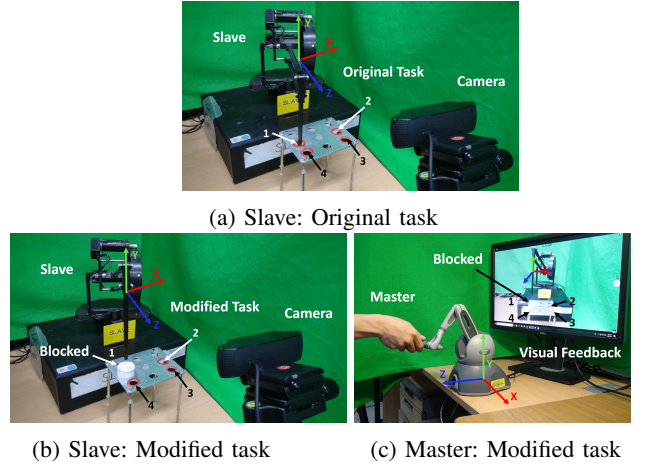


Fig. 2: (a) Slave side for the original task setting. The behaviour of the learned agent constitutes insertion into the four holes of the rig, from 1 through 4. (b) Slave side for the modified task setting with the first hole blocked and (c) master side for the modified task.

operator. Herein, the  $y$ -axis motion is DAD enabled for incremental modification on the fly, whereas the rest of the DOFs are fully controlled by the agent. The proposed incremental learning approach through teleoperation can scale to arbitrary number of DOFs, but for simplicity we show here the results for only one DOF (that is, for  $y$ -axis). The haptic feedback from the kinesthetic coupling between the human operator and the agent increases the awareness for the human about the agent's intended motion. An aluminum plate with 4 holes is placed under the slave to serve as the task rig for the peg-in-hole task. Two task settings, original and modified tasks (see Fig. 2) are considered for evaluating the performance of our proposed mini-batched incremental learning method:

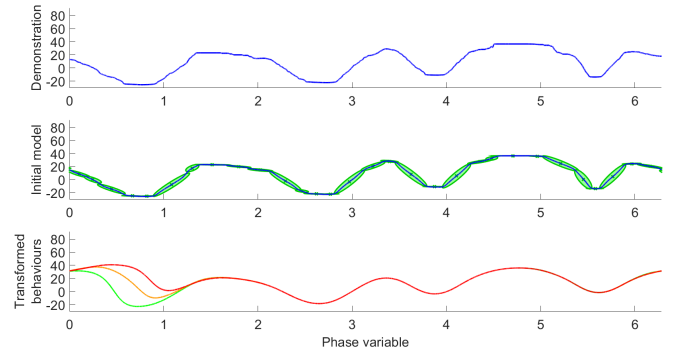


Fig. 3: Top: Teleoperated demonstration of  $y$ -axis for original task. Middle: Learned model along with its reproduced trajectory. Bottom: Behaviour of updated models after each incremental learning round for each batch of the new data points. The green and orange trajectories are the behaviours after the first and second increments and the red one indicates the behaviour after the last increment of the original model.

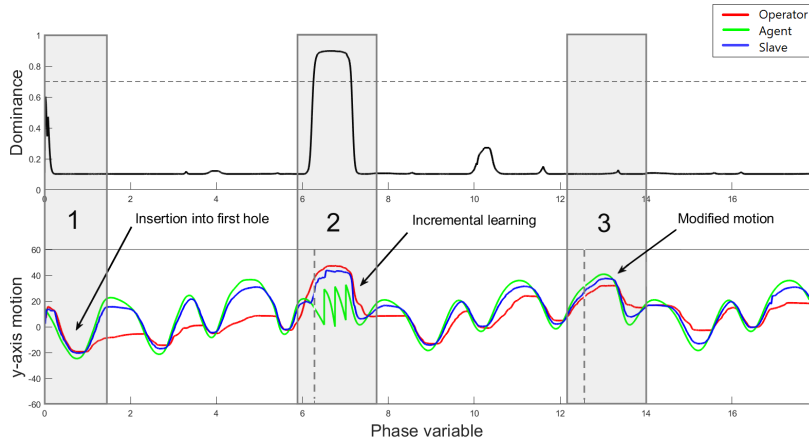


Fig. 4: Top: Operator’s dominance. The dashed horizontal line indicates the defined threshold (0.7) for acquiring new data points for incremental learning. Bottom: Operator’s, agent’s and slave’s motion trajectories in red, green and blue colours, respectively, for the consecutive cycles of motion. Gray regions 1, 2 and 3 indicate the regions of interest for task modification. Dashed vertical lines indicate the boundary of consecutive cycles of the peg-in-hole task.

1) *Original Task*: One execution cycle of the agent’s original behaviour constitutes insertion of the slave end-effector into four holes of the rig in clockwise direction, while starting and ending above the same hole (see Fig. 2a). The task repeats itself after each execution cycle (since it is rhythmic). Agent learns this four-hole insertion task first.

2) *Modified Task*: In order to introduce a slight modification in the learned agent’s original behaviour, the task setting is changed to three holes, with the first hole in the original task blocked (Fig. 2b, 2c). Now the operator has to intervene to take authority of the  $y$ -axis on the fly to restrain the insertion into the first hole to incrementally modify this region of agent’s behaviour.

### A. Results

For encoding the original behaviour of the agent, a human operator performs a single teleoperated demonstration of the peg-in-hole task. The position of the slave’s end-effector is utilized as an input to the learning algorithm. The  $y$ -axis of the teleoperated demonstration is shown in Fig. 3 (top). The fitted GMM (original model) along with the reproduced motion for the same axis is also shown in Fig. 3 (middle).

Fig. 4 shows the result of consecutive motion cycles of the human, agent and the slave device. The three gray regions of interest, where the human has to modify the original model’s behaviour to account for the first hole’s blockage, are marked as 1, 2 and 3, while the end and start of the adjacent motion cycles are marked with dashed vertical lines. In the first cycle, before the blockage is introduced, the human operator has minimal effect on the slave’s motion, as can be seen by his/her dominance value in Fig. 4. Thus the slave follows the agent’s original motion during the first cycle. Before the end of the first cycle, the first hole is blocked (modified task). Now at the start of the second cycle, within region 2, the operator asserts force against the agent to intervene and raises the end-effector above the first hole (see Fig. 2c). This results in an increase in the human’s energy and hence an increase

in the operator’s dominance in this region. Now this prohibits the insertion of slave into that hole while this partially modified behaviour is learned by the agent on the fly through our proposed incremental learning method. Whenever the human operator’s dominance crosses a predefined threshold (marked with the dashed horizontal line in Fig. 4), the spatial data points along with their corresponding phase variable values are utilized and fed into the incremental learning algorithm in mini batches. Hence, each batch of the non-contiguous modified data points is passed for performing a single round of our proposed EM. In our experiments, we’ve set the batch size to 3000 data points while  $\alpha$  and the dominance threshold have been set heuristically to 10 and 0.7, respectively. Investigating optimal values for these metaparameters has been left for future research.

Table I shows how updating GMM parameters using one batch of 3000 data points compares temporally to the conventional approach of full demonstration (demo.) based model update. In the table, *Total time* is the time taken since modified data points are recorded, up until the model parameters are updated. *Idle time* is the time taken by the respective EM algorithms to run their course. As can be seen, the proposed EM algorithm updates the model using a single batch of the partial demonstration during the same control loop (0.001 second), while for the full demonstration, the conventional EM takes 194 iterations before updating the model, thereby keeping the system continue with the originally learned behaviour until then.

During the incremental learning phase in the second cycle, agent’s erratic behaviour can be observed in region 2. This is because of the multiple updates of the model parameters for each subset of 3000 data points recorded whenever the dominance threshold is surpassed. The same erratic behaviour is not followed by the slave, since the human has high control authority (dominance threshold surpassed) during that period, and hence the effect on the slave device.

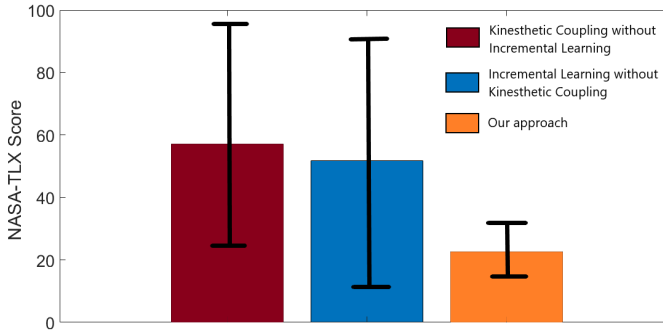


Fig. 5: Overall workload index of a modified peg-in-hole task with: Teleoperation without incremental learning, incremental learning without kinesthetic coupling, compared with the proposed incremental learning with kinesthetic coupling. Error bars indicate the minimum and maximum values.

Once the agent is modified, the human gives back his/her authority to the agent, letting agent again control the rest of the motion execution. Now, if we observe the agent’s behaviour in region 3, the insertion into the first hole is inhibited (through elevation), as it can be easily compared with the original behaviour in region 1 during the first cycle.

The behaviour of the generated models after each mini-batched update can be visualized in Fig. 3 (bottom). The effect of the first model update using the first batch of new 3000 data points is shown in green while the transition from green to red shows the behaviour of subsequent model updates that happened during the second cycle. It is clearly visible that the incremental learning only affects the agent’s behaviour locally in the region of modification, while keeping the other parts of the behaviour intact, thereby ensuring the desired partial modification of the original model using mini batches.

### B. Workload Index

In order to show the advantage of using our proposed incremental learning approach through partial teleoperated demonstration, we experimentally compare and analyze its performance against other settings to evaluate it for ease of task execution.

For this purpose, nine engineering students, both male and female, ages ranging from 22 to 32, participated in performing 3 cycles of each of the three types of experiments: Human-agent shared teleoperation without incremental learning [15, 10] with DAD and kinesthetic coupling, human-agent supervisory teleoperation with the proposed incremental learning without kinesthetic coupling ( $\bar{F}_{op}^c = 0$  in (2)), and human-agent supervisory teleoperation with the

TABLE I: Latency comparison - full demonstration vs. mini-batched incremental learning

	Total time (sec)	EM iteration(s)	Idle time (sec)
Full demo.	46.692	194	24.592
Mini-batch	3.000	1	0.001

proposed incremental learning and kinesthetic coupling (our proposed approach). The NASA-TLX subjective assessment in terms of the overall workload [29] is shown in Fig. 5 for each of the three above-mentioned experimental settings. The proposed approach ranked subjectively to have the lowest index of workload. The subjects reported that they felt complete ease at having to only intervene once for a small region of the task to merely modify it in case of a slight change. Also, they felt more at ease to just supervise over the agent with minimal interference. For the case of human-agent shared teleoperation without our proposed incremental learning, subjects commented on the difficulty of operation since they had to assert force and operate manually during each cycle of the motion, in order to account for a slight change in the original task. It is important to mention that the role of kinesthetic coupling during incremental learning made a large difference in the subjective study. For the proposed incremental learning approach without kinesthetic coupling between the human and the agent, the subjects had difficulty in modifying the task, as compared to our proposed approach with kinesthetic coupling. Subjects reported that there were no movement hints from the agent when they needed to exchange authority on the fly, therefore they had to assert unnecessarily large forces for taking authority. Also, for giving back authority to the agent, the subjects had to rely only on visual feedback of the slave in order to deduce the agent’s intended motion, which mostly resulted in abrupt jerks and oscillations. Hence, this confirms that our proposed incremental learning approach through teleoperation having kinesthetic coupling reduces operator’s workload while ensuring intuitive and precise modification of a learned behaviour online.

## V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an online incremental learning approach through a partial teleoperated demonstration. The proposed method allows an operator to partially modify the learned model online without stopping the rhythmic task execution. DAD between human and agent helps the operator to supervise the agent’s motion and intervene dynamically only when a partial change in the original behaviour is needed, given a slight change of the task. The proposed variation of EM for incremental learning enables mini-batched model updating on the fly. Results show significantly reduced operator’s workload and latency for the proposed mini-batched approach. The application of the presented approach can be expanded to many other useful areas, for example, where a repetitive motion of an industrial manipulator requires partial modifications during its execution without stopping the task. Therefore, this approach can potentially save both time and cost which would be incurred in case of a downtime retraining. Although we did not find our proposed approach very sensitive to the choice of metaparameters empirically, automatic tuning of the metaparameters could be investigated to find an optimized combination of them as an extension of the current work in the future.

## REFERENCES

- [1] R. C. Goertz. “Mechanical Master-Slave Manipulator”. In: *Nucleonics (US) (ceased publication)* (1954).
- [2] J. Cui et al. “A Review of Teleoperation System Control”. In: *Proceedings of the Florida Conf. on Recent Advances in Robotics*. Florida Atlantic University Boca Raton, FL. 2003, pp. 1–12.
- [3] H. Flemmer, B. Eriksson, and J. Wikander. “Control Design and Stability Analysis of a Surgical Teleoperator”. In: *Mechatronics* (1999), pp. 843–866.
- [4] A. Halme, J. Suomela, and M. Savela. “Applying Telepresence and Augmented Reality to Teleoperate Field Robots”. In: *Robotics and Autonomous Systems* (1999), pp. 117–125.
- [5] S. Avgousti et al. “Medical Telerobotic Systems: Current Status and Future Trends”. In: *BioMedical Engineering OnLine* (2016), p. 96.
- [6] B. Fang et al. “Robotic Teleoperation Systems Using a Wearable Multimodal Fusion Device”. In: *Int. Journal of Advanced Robotic Systems* (2017).
- [7] R. M. Pierce, E. A. Fedalei, and K. J. Kuchenbecker. “A Wearable Device for Controlling a Robot Gripper with Fingertip Contact, Pressure, Vibrotactile, and Grip Force Feedback”. In: *IEEE Haptics Symposium (HAPTICS)*. 2014, pp. 19–25.
- [8] J. Gwozdz, N. P. Morin, and R. P. Mowris. “Enabling Semi-Autonomous Manipulation on iRobot’s PackBot”. In: *Worcester Polytechnic Institute*. 2014.
- [9] S. F. Atashzar et al. “A Grasp-based Passivity Signature for Haptics-enabled Human-Robot Interaction: Application to Design of a New Safety Mechanism for Robotic Rehabilitation”. In: *The Int. Journal of Robotics Research* (2017), pp. 778–799.
- [10] A. Pervez et al. “Motion Encoding with Asynchronous Trajectories of Repetitive Teleoperation Tasks and its Extension to Human-Agent Shared Teleoperation”. In: *Autonomous Robots* (2019), pp. 1–15.
- [11] T. B. Sheridan. *Telerobotics, Automation, and Human Supervisory Control*. Cambridge, MA, USA: MIT Press, 1992. ISBN: 0-262-19316-7.
- [12] I. Havoutis and S. Calinon. “Learning from Demonstration for Semi-Autonomous Teleoperation”. In: *Autonomous Robots* (2018), pp. 1–14.
- [13] A. Pervez et al. “Novel Learning from Demonstration Approach for Repetitive Teleoperation Tasks”. In: *World Haptics Conf. (WHC)*. IEEE. 2017, pp. 60–65.
- [14] L. Peternel, T. Petrič, and J. Babič. “Human-in-the-loop Approach for Teaching Robot Assembly Tasks Using Impedance Control Interface”. In: *IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2015, pp. 1497–1502.
- [15] A. Pervez et al. “Human-Agent Shared Teleoperation: A Case Study Utilizing Haptic Feedback”. In: *Asia-Haptics*. 2018.
- [16] T. Kastritsi, F. Dimeas, and Z. Doulgeri. “Progressive Automation with DMP Synchronization and Variable Stiffness Control”. In: *IEEE Robotics and Automation Letters* (2018), pp. 3789–3796.
- [17] K. Kronander and A. Billard. “Learning Compliant Manipulation through Kinesthetic and Tactile Human-Robot Interaction”. In: *IEEE Transactions on Haptics* (2013), pp. 367–380.
- [18] M. Tykal, A. Montebelli, and V. Kyrki. “Incrementally Assisted Kinesthetic Teaching for Programming by Demonstration”. In: *ACM/IEEE Int. Conf. on Human Robot Interaction*. 2016, pp. 205–212.
- [19] S. Calinon and A. Billard. “Incremental Learning of Gestures by Imitation in a Humanoid Robot”. In: *ACM/IEEE Int. Conf. on Human-Robot Interaction*. ACM. 2007, pp. 255–262.
- [20] T. Cederborg et al. “Incremental Local Online Gaussian Mixture Regression for Imitation Learning of Multiple Tasks”. In: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 2010, pp. 267–274.
- [21] R. Wang et al. “Dynamic Movement Primitives Plus: For Enhanced Reproduction Quality and Efficient Trajectory Modification Using Truncated Kernels and Local Biases”. In: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 2016, pp. 3765–3771.
- [22] D. Papageorgiou, T. Kastritsi, and Z. Doulgeri. “A Passive Robot Controller Aiding Human Coaching for Kinematic Behavior Modifications”. In: *Robotics and Computer-Integrated Manufacturing* (2020), p. 101824.
- [23] F. Abi-Farraj et al. “A Learning-based Shared Control Architecture for Interactive Task Execution”. In: *IEEE Int. Conf. on Robotics and Automation (ICRA)*. 2017, pp. 329–335.
- [24] D. Lee and C. Ott. “Incremental Kinesthetic Teaching of Motion Primitives using the Motion Refinement Tube”. In: *Autonomous Robots* (2011), pp. 115–131.
- [25] D. Lee and C. Ott. “Incremental Motion Primitive Learning by Physical Coaching Using Impedance Control”. In: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 2010, pp. 4133–4140.
- [26] N. A. Usmani, Tae-Hwan Kim, and Jee-Hwan Ryu. “Dynamic Authority Distribution for Cooperative Teleoperation”. In: *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. 2015, pp. 5222–5227.
- [27] A. P. Dempster, N. M. Laird, and D. B. Rubin. “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1977), pp. 1–38.
- [28] A. Billard et al. “Robot Programming by Demonstration”. In: *Springer Handbook of Robotics*. 2008, pp. 1371–1394.
- [29] S. G. Hart and L. E. Staveland. “Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research”. In: *Advances in Psychology*. Elsevier, 1988, pp. 139–183.