Individual Human Behavior Identification Using an Inverse Reinforcement Learning Method

Jairo Inga, Florian Köpf, Michael Flad, Sören Hohmann
Institute of Control Systems, Karlsruhe Institute of Technology, Karlsruhe, Germany
jairo.inga@kit.edu; florian.koepf@kit.edu

Abstract—Shared control techniques have a great potential to create synergies in human-machine interaction for efficient and safe applications. However, an optimal interaction requires the machine to consider the individual behavior of the human partner. A widespread approach for modeling human behavior is given by optimal control theory, where the movement trajectories of a human arise from an optimized cost function. The aim of the identification is thus to determine parameters of a cost function which explains observed human motion. The central thesis of this paper is that individual cost function parameters which describe specific behavior can be determined by means of Inverse Reinforcement Learning. We show the applicability of the approach with a tracking control task example. The experiment consists in following a reference trajectory by means of a steering wheel. The study confirms that optimal control is suitable for modeling individual human behavior and demonstrates the suitability of Inverse Reinforcement Learning in order to determine the cost function parameters which explain measured data.

Index Terms—Human Behavior Identification, Inverse Optimal Control, Inverse Reinforcement Learning, Shared Control.

I. INTRODUCTION

Recent trends in human-machine collaboration have led to increased interest in shared control systems, where both human and machine simultaneously interact with a dynamic system. The shared control paradigm allows to keep the human "in the control loop" [1], [2]. In this way, potential issues in highly critical situations (see [3]–[5]) due to the lack of situational awareness are prevented. While literature has shown that shared control can increase overall performance while reducing control effort, performance and user acceptance can be further improved by taking individual human behavior models into account [6]. Thus, suitable models for human behavior are needed which allow the adaption of the machine to the partner it is sharing control with. An identification method for individual model parameters hence allows the development of shared control systems based on individual human skill [7].

The last decades have seen a growing trend towards optimality principles and dynamic optimization in order to describe human behavior. Within this theory, human motion arises from the optimization of a cost function, a process which is believed to take place in the central nervous system [8]. Various kinds of cost functions have been evaluated with regard to the suitability to describe human motion. Some of these include minimum jerk [9], minimum joint torque change [10] or a combination of several cost functions [11]. These studies motivate the use of optimal control theory in order to describe

human motor control intent in specific tasks [12]. In this context, the human control trajectories are available and the aim is to recover a cost function with respect to which the observed movement patterns are optimal. This is known as the *inverse optimal control* problem. Recently, literature has grown around applying inverse optimal control for human behavior identification, especially within the class of linear quadratic regulator (LQR) theory [13]–[15]. Additionally, the retrieved cost function is used in *Learning by Demonstration* in order to teach a robotic device to act in a similar way as the humans [16], [17].

Different approaches for the solution of the inverse optimal control problem have been proposed. One possibility is an iterative adaption of the cost function parameters based on a comparison of the measurements with the trajectories produced by the cost function with the current parameters [15], [16]. However, the computation time is in general huge since several forward optimal control problems have to be solved in the identification process. Other studies introduce the minimization of residual functions which represent the extent to which first-order optimality conditions are violated [18]–[20]. Further methods either assume the control law as a function of the state to be given [12] or propose a solution based on the Riccati equation which leads to a limitation to LQR problems [13].

A related class of methods is given by Inverse Reinforcement Learning (IRL) techniques. These have been developed in the field of computer science and in the context of Markov Decision Processes (MDP). Such models assume discrete and finite control and state variables. More recently, IRL methods for continuous variables were proposed, making them similar to inverse optimal control and therefore suitable for modeling human motion which is of continuous nature. Observed human motion trajectories might slightly deviate from optimality due to measurement noise, control noise and variability of human motor control [21]. To account for these effects during identification, a probabilistic IRL approach is reasonable. Therefore, we propose for human behavior identification the application of a Maximum Entropy (MaxEnt) based IRL method. Using the principle of maximum entropy (a concept introduced in [22]) also helps to avoid issues concerning the ill-posedness of inverse optimization.

In this paper, we present an approach for identification of human behavior by means of inverse optimization. While the structure of the cost function is a-priori defined, its parametrization can describe the different movement patterns

of humans performing the same task. The results demonstrate that specific cost function parameters can be found for each individual performing a control task and hence lead to characteristic models for each person.

The paper is organized as follows. In the next section, we formulate the inverse optimization problem for the identification of cost function parameters. Section III introduces MaxEnt IRL for the identification problem. We further show in Section IV the applicability of the approach for human behavior modeling by means of an experiment where several subjects perform a specific control task. Finally, conclusions are drawn in Section V.

II. PROBLEM DEFINITION

In this paper, we regard human behavior in the control of a specific dynamic system. We assume the system to be time-invariant and to be given by

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{f}(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)}) \tag{1a}$$
$$\boldsymbol{x}^{(1)} = \boldsymbol{x}_1, \tag{1b}$$

with discrete states $\boldsymbol{x}^{(k)} \in \mathbb{R}^n$ and controls $\boldsymbol{u}^{(k)} \in \mathbb{R}^m$, where $(\cdot)^{(k)}$ indicates the k-th time step. Furthermore, $\boldsymbol{x}_1$ denotes the initial state value. We define the control and state trajectories generated by the human as

$$\underline{\boldsymbol{x}} = \left[\left(\boldsymbol{x}^{(1)}\right)^\top \quad \ldots \quad \left(\boldsymbol{x}^{(T)}\right)^\top \right]^\top \in \mathbb{R}^{nT} \tag{2a}$$

and

$$\underline{\boldsymbol{u}} = \left[\left(\boldsymbol{u}^{(1)}\right)^\top \quad \ldots \quad \left(\boldsymbol{u}^{(T)}\right)^\top \right]^\top \in \mathbb{R}^{mT} \tag{2b}$$

for T time steps.

Following the line of optimal control theory for modeling human behavior, the control and state sequence in (2) arise from the minimization of a cost function. In other words, the human determines a control strategy to minimize the individual cost function $J(\underline{\boldsymbol{x}}, \underline{\boldsymbol{u}})$ which leads to the optimization problem

$$\underline{\boldsymbol{u}}^* = \arg\min_{\underline{\boldsymbol{u}}} J(\underline{\boldsymbol{x}}, \underline{\boldsymbol{u}}), \tag{3}$$

w.r.t. (1), where $(\cdot)^*$ denotes the optimal solution. The control strategy $\underline{\boldsymbol{u}}^*$ leads to corresponding trajectories $\underline{\boldsymbol{x}}^*$.

Within the scope of this paper, we assume the cost function to be a linear combination of N given features $\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{u})$, i.e.

$$J = \sum_{k=1}^{T} \boldsymbol{\theta}^\top \boldsymbol{\phi}(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)}), \tag{4}$$

with $\boldsymbol{\theta} \in \mathbb{R}^N$.

The problem of identifying J is then formulated as follows.

Problem 1 [Identification of J]

Let measured state and control trajectories $\underline{\tilde{\boldsymbol{x}}}$, $\underline{\tilde{\boldsymbol{u}}}$ of human movement and the system dynamics (1) *be given. Furthermore, assume the observed states and controls to be optimal w.r.t. a cost function with the structure given in* (4).
Determine parameters $\hat{\boldsymbol{\theta}}$ of the cost function $\hat{J}$ which is minimized by $\underline{\tilde{\boldsymbol{u}}}$.

III. IDENTIFICATION METHOD

The problem of recovering cost functions from observations of expert's demonstrations assumed as optimal is ill-posed. In this case this means that multiple possible cost functions may provide identical optimal trajectories. Moreover, as mentioned previously, observed trajectories might slightly deviate from optimality [21]. To account for these effects during identification, a probabilistic approach is reasonable, where observed trajectories are assumed to be sampled by a density $p(\zeta), \zeta = \{\underline{\boldsymbol{x}}, \underline{\boldsymbol{u}}\}$ and denoted as expert trajectories $\zeta_{E_l}, l \in \{1, ..., d\}$. Therefore, in order to resolve this ambiguity, we base the identification method upon the MaxEnt IRL formulation first introduced by [22] for discrete and finite values, which is based on the principle of maximum entropy. This method leads to the least biased estimate possible on the given information [23].

In order to describe human behavior, continuous valued states and controls need to be considered. This leads to an uncountable infinite set of possible trajectories. Maximizing differential entropy leads to a probability distribution

$$p(\zeta_{E_l}) = \frac{\mathrm{e}^{-J(\zeta_{E_l})}}{\int_{-\infty}^{\infty} \mathrm{e}^{-J(\tilde{\zeta})} \, \mathrm{d}\tilde{\zeta}}. \tag{5}$$

Here, the integral is over all possible trajectories considering (1).

Since we regard deterministic system dynamics, the expert trajectories ζ_{E_l} are uniquely defined by the initial state $\boldsymbol{x}^{(1)}$ and the control strategy $\underline{\boldsymbol{u}}$. Therefore, $p(\zeta_{E_l}) = p(\underline{\boldsymbol{u}}_{E_l} | \boldsymbol{x}^{(1)}, \boldsymbol{\theta})$. Following the formulation of [24] for deterministic, fixed-horizon control tasks, rewriting (5) and using a quadratic Taylor series expansion around $\underline{\boldsymbol{u}}_{E_l}$ for approximating the cost function J in the integral leads to

$$p\left(\underline{\boldsymbol{u}}_{E_l} \,\middle|\, \boldsymbol{x}^{(1)}, \boldsymbol{\theta}\right) = \frac{\mathrm{e}^{-J\left(\underline{\boldsymbol{u}}_{E_l} \middle| \boldsymbol{x}^{(1)}, \boldsymbol{\theta}\right)}}{\int_{-\infty}^{\infty} \mathrm{e}^{-J\left(\underline{\tilde{\boldsymbol{u}}} \middle| \boldsymbol{x}^{(1)}, \boldsymbol{\theta}\right)} \, \mathrm{d}\underline{\tilde{\boldsymbol{u}}}} \tag{6}$$
$$\approx \mathrm{e}^{\left(-\frac{1}{2} \boldsymbol{g}_{E_l}^\top \boldsymbol{H}_{E_l}^{-1} \boldsymbol{g}_{E_l}\right)} \det\left(\boldsymbol{H}_{E_l}\right)^{\frac{1}{2}} (2\pi)^{-\frac{nT}{2}}.$$

Here, $\boldsymbol{g}_{E_l} \in \mathbb{R}^{nT}$ and $\boldsymbol{H}_{E_l} \in \mathbb{R}^{nT \times nT}$ are defined as

$$\boldsymbol{g}_{E_l} := \frac{\partial J}{\partial \underline{\boldsymbol{u}}} + \boldsymbol{D} \frac{\partial J}{\partial \underline{\boldsymbol{x}}} \bigg|_{\substack{\underline{\boldsymbol{u}} = \underline{\boldsymbol{u}}_{E_l} \\ \underline{\boldsymbol{x}} = \underline{\boldsymbol{x}}_{E_l}}}, \tag{7}$$

$$\boldsymbol{H}_{E_l} := \frac{\partial^2 J}{\partial \underline{\boldsymbol{u}}^2} + \boldsymbol{D} \frac{\partial^2 J}{\partial \underline{\boldsymbol{x}}^2} \boldsymbol{D}^\top + \frac{\partial^2 \underline{\boldsymbol{x}}}{\partial \underline{\boldsymbol{u}}^2} \frac{\partial J}{\partial \underline{\boldsymbol{x}}} \bigg|_{\substack{\underline{\boldsymbol{u}} = \underline{\boldsymbol{u}}_{E_l} \\ \underline{\boldsymbol{x}} = \underline{\boldsymbol{x}}_{E_l}}} \tag{8}$$

The matrix $\boldsymbol{D} = \frac{\partial \underline{\boldsymbol{x}}}{\partial \underline{\boldsymbol{u}}}^\top$ represents the influence of changes of $\boldsymbol{u}$ on the state $\boldsymbol{x}$ at each time step $k = 1, ..., T$ [25]. The entries of $\boldsymbol{D}$ are given by

$$D_{k_1,k_2} = \frac{\partial \boldsymbol{x}^{(k_2)}}{\partial \boldsymbol{u}^{(k_1)}}^{\top}$$

$$= \begin{cases} \left. \dfrac{\partial \boldsymbol{x}^{(k+1)}}{\partial \boldsymbol{u}^{(k)}}^{\top}\right|_{k=k_1}, & k_2 = k_1 + 1 \\[2ex] \left. D_{k_1,k_2-1}\dfrac{\partial \boldsymbol{x}^{(k+1)}}{\partial \boldsymbol{x}^{(k)}}^{\top}\right|_{k=k_2-1}, & k_2 > k_1 + 1 \end{cases} \quad (9)$$

The partial derivatives of the states with respect to the control values are part of the numerical solution of the initial value problem which approximates the next state. The matrix $\boldsymbol{D}$ employs the partial derivatives with respect to $\boldsymbol{u}$ in each time step for the whole corresponding interval ΔT. Contrary to this approach, we propose a modification of the matrix $\boldsymbol{D}$ in order to improve the approximation. Inspired by the trapezoid method for solving initial value problems [26], we propose an approximation of the effect of $\boldsymbol{u}^{(k_1)}$ at $k_1 \Delta T$ on $\boldsymbol{x}^{(k_2)}$ by means of

$$\tilde{\boldsymbol{D}}_{k_1,k_2} := \frac{1}{2}\left(\frac{\partial \boldsymbol{x}^{(k_2)}}{\partial \boldsymbol{u}^{(k_1)}} + \frac{\partial \boldsymbol{x}^{(k_2+1)}}{\partial \boldsymbol{u}^{(k_1)}} \right)^{\top}. \quad (10)$$

The modified matrix $\tilde{\boldsymbol{D}}$ takes in to account the effect of the control value $\boldsymbol{u}^{(k_1)}$ on the interval of $\boldsymbol{x}^{(k_2)}$ until $\boldsymbol{x}^{(k_2+1)}$. This allows a better approximation of the influence on the states by changes in the control applied by the human.

After determining (10) and hence (7) and (8), the approximation for the probability distribution in (6) can be calculated. Let d observed expert demonstrations be given by $\mathcal{D} = \{\zeta_{E_1}, \ldots, \zeta_{E_d}\}$ with $\zeta_{E_l} = \{\underline{\boldsymbol{x}}_{E_l}, \underline{\boldsymbol{u}}_{E_l}\}$, $l = 1, \ldots, d$. Now we determine

$$\hat{\boldsymbol{\theta}}_E = \arg\max_{\boldsymbol{\theta}} \prod_{l=1}^{d} p\left(\zeta_{E_l}\right) = \arg\max_{\boldsymbol{\theta}} \sum_{l=1}^{d} \ln\left(p\left(\zeta_{E_l}\right)\right), \quad (11)$$

where $\hat{\boldsymbol{\theta}}_E$ is an estimate for the parameters of the expert's cost function. This is a parameter optimization problem which can be solved using standard algorithms. Maximizing $p\left(\zeta_{E_l}\right)$ leads to a high probability for the expert's demonstrated trajectories $\mathcal{D}$. This means, trajectories with similar costs as the expert's are favoured. At the same time, it requires that the demonstrations receive minimal costs from the function parameterized by $\boldsymbol{\theta}$. The solution $\hat{\boldsymbol{\theta}}_E$ of (11) solves the identification problem defined in Section II.

IV. EXPERIMENTAL RESULTS

In this section we demonstrate the suitability of the previously presented IRL method for identification by means of an experiment. Our aim is to determine individual cost function parameters from measured data which describe the behavior of each test person.

A. Experimental Framework and Participants

For the measurements, we used an experimental framework consisting of three main components: an active steering wheel, a monitor with a visualization window and a real-time environment which contains the communication structure between

Fig. 1. Steering wheel and visualization monitor used for the experiment

TABLE I
STEERING WHEEL PARAMETERS

Parameter	Value		Description
Θ_L	0.04	kg m^2	Steering wheel rotational inertia
c	1.146	Nm/rad	Spring constant
d	0.286	Nm·s/rad	Damping constant

all components. Fig. 1 shows the components of the conducted experiment.

The measurements of the steering wheel angle φ are done by an incremental encoder of 40000 increments per full rotation at a sampling frequency of $f_s = 40\,\text{Hz}$. The dynamics of the steering wheel is given by

$$\dot{\boldsymbol{x}} = \begin{bmatrix} 0 & 1 \\ -\frac{c}{\Theta_L} & -\frac{d}{\Theta_L} \end{bmatrix} \boldsymbol{x} + \begin{bmatrix} 0 \\ \frac{1}{\Theta_L} \end{bmatrix} \boldsymbol{u}, \quad (12)$$

where $\boldsymbol{x} = \begin{bmatrix} \varphi & \dot{\varphi} \end{bmatrix}^{\top}$ and $u = M$ is the steering torque applied by the human. The parameters of the system are given in Table I. This system was discretized using a sample time of $T_s = 1/f_s = 0.025\,\text{s}$ in order to get the discrete-time representation given in (1).

Fig. 2 shows a sequence of screen captures from the experiment visualization. The participants are able to control the horizontal position of the marker (square) in order to follow a reference trajectory. A reference point crosses the complete visualization window in $2\,\text{s}$. The vertical position of the marker is fixed at a height of 25%.

The control task consists in following a reference trajectory using the steering wheel. We further introduce the velocity and the torque as possible features:

$$\boldsymbol{\phi}(\boldsymbol{x}^{(k)}, \boldsymbol{u}^{(k)}) = \begin{bmatrix} \left(x_1^{(k)} - x_{1,\text{ref}}^{(k)}\right)^2 \\ \left(x_2^{(k)}\right)^2 \\ \left(u^{(k)}\right)^2 \end{bmatrix}. \quad (13)$$
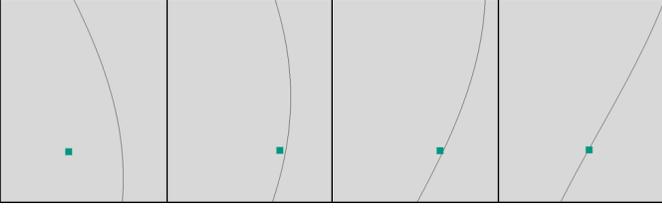
Fig. 2. Visualization of the control task. The picture sequence is from left to right.

Hence, the cost function is given by (4), with the features (13) and is parameterized by a vector $\boldsymbol{\theta} \in \mathbb{R}^3$.

Five people (age 24.6 ± 2.19) participated in the study. They did not have prior knowledge of the investigation subject and also never participated in a study concering the experimental design described previously. The test persons did a first run for 2 minutes with an arbitrary reference composed of constant, steps and linear parts. They were told to move the steering wheel freely and not necessarily with the aim of matching the reference. In this way, they were able to get familiar with the steering wheel and the visualization system. Afterwards, before starting the test to be measured, they were told this time to move the steering wheel in such a way that the marker matches the reference trajectory. This reference had similar components as the one the participants saw in the test run. The final experiment had a duration of 90 seconds.

B. Evaluation Procedure

The identification algorithm demands an expert demonstration ζ_E which consists of all states and the control variable for all time steps k. In the experiment, only a sensor for measuring the steering angle $\underline{x}_1$ is available. Therefore, the steering angle velocity $\underline{x}_2$ and the steering torque $\underline{u}$ were obtained by differentiating the angle signal and calculated by means of (12), respectively.

Afterwards, the expert demonstration of each test person was used to determine their individual cost function parameters $\boldsymbol{\theta}$ by solving (11) with a standard sequential quadratic programming method. Since no ground truth exists for the cost function parameters, it is not possible to evaluate the obtained cost function parameters directly. Therefore, we use them to calculate the trajectories $\underline{x}$ and $\underline{u}$ which are optimal with respect to the identified cost function. The trajectories are determined by solving the dynamic optimization problem numerically. These are then compared to the expert demonstrations. This evaluation procedure is depicted in Fig. 3. We calculated the root mean squared (RMS) error between model trajectory i and measured trajectory j

$$\delta_\xi(i,j) = \sqrt{\frac{1}{T}\sum_{k=1}^{T}\left(\tilde{\xi}_i^{(k)} - \hat{\xi}_j^{(k)}\right)^2} \qquad (14)$$

for $\xi \in \{x_1, x_2, u\}$ and then build the mean RMS error

$$\delta_{\mathrm{MRMS}}(i,j) = \frac{1}{3}\left(\delta_{x_1}(i,j) + \delta_{x_2}(i,j) + \delta_u(i,j)\right) \qquad (15)$$

for all combinations $i, j \in \{1, ..., 5\}$.

C. Results

The identification algorithm determined different cost function parameters for each test person. The parameters are given in Table II. We normalized all parameters with respect to θ_3 in order to ensure comparability. It is worth mentioning that this does not affect the trajectories $\hat{\underline{x}}$ and $\hat{\underline{u}}$ since the ratio between cost function parameters determines the course of the trajectory and not their absolute values. We show exemplarily the identification results for the whole experiment of subject 1. The model trajectories generated by the identified parameters, the expert demonstration and the reference trajectories are depicted in Fig. 4. It can be seen that the model is capable of describing the behavior of the subject while performing the control task. However, several jumps can also be observed in the model trajectory of the steering torque u. These occur at the same time where the reference trajectory of the steering angle has a step. These steps could be observed in all identification results. Furthermore, we compared the identification results of the steering angle of test subject 1 and 2. This is shown in Fig. 5 for the time between $t = 0$ and $t = 18$. The form of the trajectory of both subjects is similar. However, some differences are noticeable. Test subject 2 is less smooth and did not follow the reference with the same accuracy as subject 1. The following Table III shows the mean RMS error between model and measured trajectories. For example, model 3 has the least mean RMS error to the corresponding subject 3. However, models 1 and 5 show similar values. A comparable result is observed with subjects 2 and 4 which are very alike. This similarities are also noticeable in the identified parameters in Table II.

D. Discussion

The experiments indicate that different movement patterns between humans can be observed, even though the control
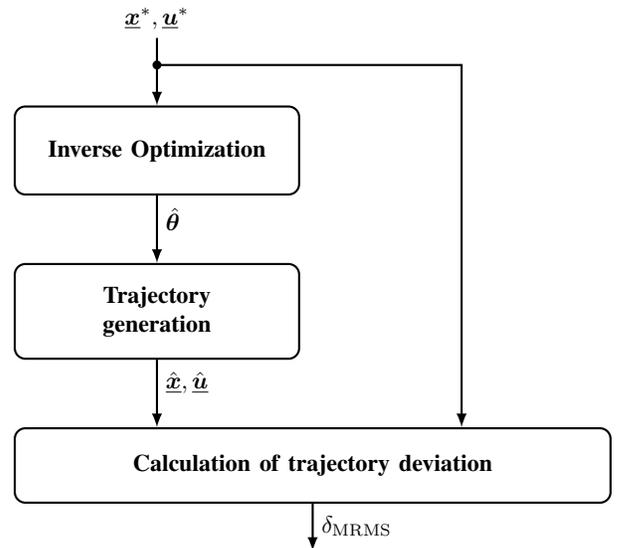
Fig. 3. Evaluation process for the identification results
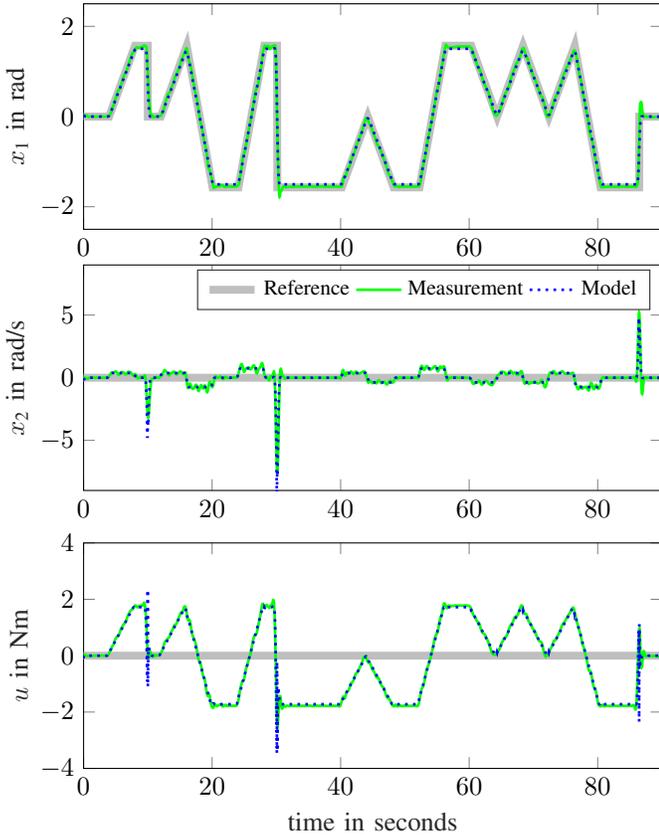
Fig. 4. Identification results for subject 1. The reference trajectories are plotted in gray and the expert demonstration in green. The blue dotted trajectories are optimal with respect to the cost function with the identified parameters $\hat{\theta}$.

Fig. 5. Expert demonstrations (red) and model trajectories (orange, dashed) for subject 1. Expert demonstrations (dark green) and model trajectories (green, dotted) for subject 2.

TABLE II
IDENTIFICATION RESULTS

Test person	$\hat{\theta}$
1	$\begin{bmatrix} 28.7939 & 0.2092 & 1 \end{bmatrix}^{\top}$
2	$\begin{bmatrix} 14.2072 & 0.5758 & 1 \end{bmatrix}^{\top}$
3	$\begin{bmatrix} 24.3028 & 0.2566 & 1 \end{bmatrix}^{\top}$
4	$\begin{bmatrix} 14.7202 & 0.6295 & 1 \end{bmatrix}^{\top}$
5	$\begin{bmatrix} 22.8732 & 0.2231 & 1 \end{bmatrix}^{\top}$

TABLE III
MEAN RMS ERROR δ_{MRMS} FOR ALL SUBJECTS

		Data			
Model	**1**	**2**	**3**	**4**	**5**
1	0.1282	0.1965	0.1472	0.2292	0.1505
2	0.1799	0.1624	0.1728	0.1914	0.1820
3	0.1274	0.1826	0.1411	0.2153	0.1467
4	0.1806	0.1617	0.1732	0.1899	0.1824
5	0.1278	0.1835	0.1419	0.2171	0.1473

task was clearly specified by a trajectory. Considering higher-level goals (e.g. reaching a certain final position) would lead to an even greater variability as shown by several studies

[27], [28]. The observed differences in human behavior can be described by the optimal control approach. Fig. 5 shows that the steering angle trajectories of subject 1 and 2 are best explained by the corresponding models. Table III confirms this trend. The identification algorithm does not directly minimize this metric. Therefore, in some cases the corresponding model is outperformed by another similar model by a small margin. These minor differences may also arise due to signal noise. Table II shows that subjects 1, 3 and 5 as well as subjects 2 and 4 have similar parametrization. Since the parameters θ_1 of the first group are bigger, it can be asserted that this group was more skilled in the tracking task than the second group. This can also be seen in the comparison of subject 1 and 2 in Fig. 5.

On the whole, the model is able to describe human behavior adequately. Nevertheless, it is noticeable that some aspects cannot be fully taken account of. For example, the final value of a reference (e.g. after a step) is generally not reached. One difficulty lies within the fact that the test persons sometimes reached the final value and sometimes did not. This effect could not be taken into account by the cost function. Moreover, high-frequency components of signals cannot be described by the model. This is especially noticeable in the steering angle velocity trajectory, where the model can only reproduce an

average value. The reason is that these high-frequency signals are not part of the cost function. The aforementioned jumps in the model trajectory of the steering torque u, which can be seen in Fig. 4 around $t = 10$, $t = 30$ and near the end, are due to the discontinuities in the reference trajectory. Reason for this is that the approach determines the optimal trajectories w.r.t. the cost function and does not prohibit abrupt changes in the values at the same time. In order to mitigate this effect, the system dynamics could be extended with a submodel which describes human arm dynamics.

To summarize, it can be stated that the identification algorithm is able to describe individual behavior. Different behavior leads to different cost function parameters which generate trajectories which approximate the specific motion trajectories adequately.

V. Conclusion

In this paper, we presented an approach for human behavior identification by means of Inverse Reinforcement Learning. The model of the human is given by the optimal control approach where human behavior when performing a control task is produced by optimizing a cost function. We proposed an approach based on the principle of maximum entropy in order to determine cost function parameters. The performed experiment with 5 participants showed that inverse optimization methods can be used to identify individual parameters of a cost function underlying human behavior. It could also be observed that the behavior of different subjects varies in spite of performing the same specified task. The differences in their behavior leads to different cost function parameters. Furthermore, the approach allows the identification of different skill groups. The experiment confirmed that the optimal control approach is able to model specific human behavior which is essential for shared control systems. In future work, we will analyze different structures of the cost function for identification. Additionally, we will perform similar studies concerning identification of human behavior in a control task in the presence of an automatic controller.

References

[1] M. Flad, J. Otten, S. Schwab, and S. Hohmann, "Necessary and sufficient conditions for the design of cooperative shared control," in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 2014, pp. 1253–1259.

[2] D. Abbink, M. Mulder, and E. Boer, "Haptic shared control: Smoothly shifting control authority?" *Cognition, Technology & Work*, vol. 14, no. 1, pp. 19–28, 2012.

[3] D. Richards and A. Stedmon, "To delegate or not to delegate: A review of control frameworks for autonomous cars," *Applied Ergonomics*, vol. 53, pp. 383–388, 2016.

[4] N. Strand, J. Nilsson, I. Karlsson, and L. Nilsson, "Semi-automated versus highly automated driving in critical situations caused by automation failures," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 27, pp. 218–228, 2014.

[5] L. Bainbridge, "Ironies of automation," *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

[6] M. Mulder, D. Abbink, and E. Boer, "Sharing control with haptics: Seamless driver support from manual to automatic control," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 54 (5), pp. 786–798, 2012.

[7] D. Abbink, D. Cleij, M. Mulder, and M. van Paassen, "The importance of including knowledge of neuromuscular behaviour in haptic shared control," in *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 3350–3355.

[8] S. Scott, "Optimal feedback control and the neural basis of volitional motor control," *Nature Reviews Neuroscience*, vol. 5, no. 7, pp. 532–546, 2004.

[9] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *The journal of Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[10] Y. Uno, M. Kawato, and R. Suzuki, "Formation and control of optimal trajectory in human multijoint arm movement," *Biological cybernetics*, vol. 61, no. 2, pp. 89–101, 1989.

[11] B. Berret, E. Chiovetto, F. Nori, and T. Pozzo, "Evidence for composite cost functions in arm movement planning: An inverse optimal control approach," *PLoS Computational Biology*, vol. 7, no. 10, 2011.

[12] W. Li, E. Todorov, and D. Liu, "Inverse optimality design for biological movement systems," *Proceedings of the 18th IFAC World Congress*, pp. 9662–9667, 2011.

[13] M. C. Priess, R. Conway, J. Choi, J. M. J. Popovich, and C. Radcliffe, "Solutions to the inverse lqr problem with application to biological systems analysis," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 2, pp. 770–777, 2015.

[14] A. Ramadan, J. Choi, and C. Radcliffe, "Inferring human subject motor control intent using inverse mpc," *American Control Conference (ACC)*, pp. 5791–5796, 2016.

[15] H. El-Hussieny, A. A. Abouelsoud, S. F. M. Assal, and S. M. Megahed, "Adaptive learning of human motor behavior: An evolving inverse optimal control approach," *Engineering Applications of Artificial Intelligence*, vol. 50, pp. 115–124, 2016.

[16] K. Mombaur, A. Truong, and J. Laumond, "From human to humanoid locomotionan inverse optimal control approach," *Autonomous Robots*, vol. 28, no. 3, pp. 369–383, 2010.

[17] P. Englert and M. Toussaint, "Inverse KKT learning cost functions of manipulation tasks from demonstrations," in *Proceedings of the International Symposium of Robotics Research*, 2015.

[18] M. Johnson, N. Aghasadeghi, and T. Bretl, "Inverse optimal control for deterministic continuous-time nonlinear systems," in *IEEE 52nd Annual Conference on Decision and Control (CDC)*, 2013, pp. 2906–2913.

[19] A.-S. Puydupin-Jamin, M. Johnson, and T. Bretl, "A convex approach to inverse optimal control and its application to modeling human locomotion," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2012, pp. 531–536.

[20] A. M. Panchea and N. Ramdani, "Towards solving inverse optimal control in a bounded-error framework," in *American Control Conference (ACC), 2015*, 2015, pp. 4910–4915.

[21] E. Todorov, "Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system," *Neural computation*, vol. 17, no. 5, pp. 1084–1108, 2005.

[22] B. Ziebart, A. Maas, J. Bagnell, and A. Dey, "Maximum entropy inverse reinforcement learning," in *Proceedings of the 23rd National Conference on Artificial Intelligence (AAAI),*, 2008, pp. 1433–1438.

[23] E. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, pp. 620–630, 1957.

[24] S. Levine and V. Koltun, "Continuous inverse optimal control with locally optimal examples," in *Proceedings of the 29th International Conference on Machine Learning (ICML)*, 2012, pp. 41–48.

[25] S. Levine, "Motor skill learning with local trajectory methods," Ph.D. dissertation, Standford University, 2014.

[26] J. Epperson, *An introduction to numerical methods and analysis*, second edition ed. Hoboken, New Jersey: John Wiley & Sons, Inc, 2013.

[27] M. D'Andola, B. Cesqui, A. Portone, L. Fernandez, F. Lacquaniti, and A. d'Avella, "Spatiotemporal characteristics of muscle patterns for ball catching," *Frontiers in Computational Neuroscience*, vol. 7, no. 107, 2013.

[28] C. Seegelke, C. Hughes, A. Knoblauch, and T. Schack, "The influence of reducing intermediate target constraints on grasp posture planning during a three-segment object manipulation task," *Experimental Brain Research*, vol. 233, no. 2, pp. 529–538, 2015.