Available online at www.sciencedirect.com

ScienceDirect

IFAC-PapersOnLine 49-32 (2016) 095–100

Extraction and Deployment of Human Guidance Policies

Andrew Feit and Bérénice Mettler

Department of Aerospace Engineering and Mechanics, University of Minnesota, Minneapolis, MN 55455 (e-mail: feit0003@umn.edu and mettler@umn.edu)

Abstract:
Robust and adaptive human motion performance depends on learning, planning, and deploying primitive elements of behavior. Previous work has shown how human motion behavior can be partitioned at subgoal points, and primitive elements extracted as trajectory segments between subgoals. An aggregate set of trajectory segments are described by a spatial cost function and guidance policy. In this paper, Gaussian process regression is used to approximate cost and policy functions extracted from human-generated trajectories. Patterns are identifying in the policy function to further decompose guidance behavior into a sequence of motion primitives. A maneuver automaton model is introduced, simplifying the guidance task over a larger spatial domain. The maneuver automaton and approximated policy functions are then used to generate new trajectories, replicating original human behavior examples.

Keywords: Motion guidance, Perception, Planning, Autonomous guidance, Transfer learning, Motion automaton

1. INTRODUCTION

1.1 Motivation

Humans and animals routinely generate a wide range of adaptive and robust motion behavior in cluttered and dynamic environments. Their behavior far exceeds the capabilities of autonomous systems, despite also having sensory and computational limitations. Understanding the processes that allow agents to learn and generate this behavior would be a significant advancement for autonomous vehicle performance and human-machine system interaction.

From a control theory perspective, guidance behavior can be defined by a value function and guidance policy. Kong and Mettler (2013) and Feit and Mettler (2015) propose that human guidance behavior is also described by value and policy functions. Lee et al. (1976) and Mettler et al. (2014) suggest that humans generate motion by deploying sensory-motor interaction patterns. Interaction patterns extend the motion-primitive automaton concept introduced by Frazzoli et al. (2005) by introducing guidance primitives. Guidance primitives are discrete agent-environment interaction elements. An agent can deploy a series of guidance primitives chosen from a library of learned elements. Each element optimizes closed-loop perception and action behavior performance.

The present work focuses on identifying, representing, and deploying human motion guidance policies. First, we approximate cost and guidance policy functions that describe an ensemble of human example behavior. Second, we use the identified guidance policy to generate new unconstrained and constrained trajectories to validate the use of subgoals and motion primitives as elements of human behavior. Modeling human motion policy that optimize a value function is a necessary precursor to modeling perception-action guidance primitives in future work.

2. RELATED WORK

2.1 Invariants in Human Behavior

Previous work investigates the guidance process by identifying patterns in human behavior. Simon (1972) introduced satisficing, which encompasses approaches that simplify or reduce the problem domain, in exchange for reducing solution optimality. In addition, Simon (1990) observes patterns in agent-environment interaction that simplify the guidance task. Specifically, Simon notes that elements of human motion behavior can be described by the maintenance of invariants between perceptual and kinematic quantities. Finally, Lee et al. (1976) suggests that biological motion is composed of primitive elements, which are described by a τ parameter, the instantaneous time of a motion gap closure. Motion profiles are generated by maintaining a constant $\dot{\tau}$.

Human Guidance Models Mettler and Kong (2012) integrates these concepts by introducing a hierarchical model for human guidance. This model suggests that human behavior consists of planning, perceptual guidance, and tracking. In this model, planning involves choosing a sequence of subgoal states. Perceptual guidance involves deploying primitive elements of motion and perception to reach each subgoal. Tracking involves regulating system-environment dynamics to implement a motion primitive

2405-8963 © 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.
Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2016.12.196

element. Kong and Mettler (2013) and Mettler et al. (2014) show that guidance primitives consist of sensory-motor interaction patterns that achieve desired motion performance and risk. Prior work on guidance (Feit and Mettler (2016)) explored closed-loop modes between perceptions and actions provided by interaction patterns.

Motion Primitive Maneuver Automaton Frazzoli et al. (1999) and Frazzoli et al. (2002) introduce the maneuver automaton (MA), which is a finite-state approximation of system dynamics. A MA generates complex trajectories by constructing a sequence of motion primitive elements. Motion primitives are chosen from a library of known behaviors consisting of, for example, trim and maneuver elements. Mettler et al. (2002) evaluates this approach in the application of aerobatic rotorcraft control.

To identify motion primitives in human behavior, Li and Mettler (2015) investigates the dynamic clustering of surgical motion. This provides an approach to classifying observed behavior into specific motion primitive groups. MA approaches require dynamic programming to choose an optimal sequence of motion primitives to complete a task in a constrained environment. Feit et al. (2015) proposes a human-inspired approach to constrained optimal control based on choosing a series of subgoal states. Subgoal candidates are defined by a set of necessary conditions, based on constraint geometry and a known guidance policy.

Control Theory and Human Guidance In control theory, a value function $V(x)$ specifies the cost accrued by following a policy, $\pi(x)$ beginning at state $x = [x_p, x_v]$, partitioned into configuration and velocity states, x_p and x_v. Kong and Mettler (2011) observes that configuration (x_p) and velocity (x_v) dynamics can be partitioned as $\dot{x}_p = x_v$ and $\dot{x}_v = f(x_v, u)$. In this case, the guidance policy defines the optimal velocity at a given configuration, $x_v^* = \pi(x_p)$. The value function is then defined over the configuration space, $V^*(x_p)$, as the spatial cost-to-go. Kong and Mettler (2013) models the spatial cost-to-go (CTG) and guidance policy (also termed the velocity-vector-field (VVF)) for an ensemble of third-person control tasks involving guiding a model-helicopter through an obstacle field. Feit and Mettler (2015) performs a similar experiment using a first-person computer simulated environment. Both experiments show that the ensemble of resulting trajectories can be partitioned based on subgoal properties defined by Kong and Mettler (2013). The resulting sets of trajectory segments between subgoals are modeled by consistent CTG and VVF functions, using time-to-go as the cost function.

Guidance Policy Representation Kong and Mettler (2013) and Feit and Mettler (2015) observe that a library of human motion primitives are described by a spatial CTG and VVF function, which together form a control policy. Russell et al. (2003) summarizes a variety of approaches to learning or identifying cost and policy functions from a set of example behavior data. A utility function can be directly approximated when the value function is known along each example trajectory. A data regression method estimates $\hat{V}^*(x_p) \approx V^*(x_p)$, based on a set of observed example data $D = [\mathbf{x_p}, \mathbf{w}]$. This approach approximates observed behavior, with observed cost $\mathbf{w}$.

Reinforcement learning (RL) can be used to determine an optimal guidance policy from a set of (possibly sub-optimal) example data. Temporal difference (TD) learning identifies the optimal cost function if rewards are known for each state transition. This approach iterates a Bellman-like update equation to converge to the optimal cost function. Q-learning also uses an iterative equation, but determines the expected value of each action at each system state, determining both the CTG and optimal policy.

The above approaches to utility function learning require a known reward function of state transitions, which is often not available for human behavior. Inverse reinforcement learning seeks to determine a utility function, given a set of example behavior that is assumed to be optimal. Ng et al. (2000) summarizes algorithms available for this process.

Function Approximation A regression method is required to represent estimated CTG and guidance policy functions over the spatial domain. This method should be parameter-free, and not be restricted to any specific function type. In addition, the method should provide accuracy of the function value at evaluated points, to elucidate the role of uncertainty in human decision making. Gaussian process (GP) regression is employed to meet these requirements, as described in Ebden (2008). Note that GP regression has been extended for use in a TD learning framework in Engel et al. (2003). The GPStuff toolbox for Matlab is used in this work to implement function approximation (Vanhatalo et al. (2013)).

2.2 Summary

The rest of the paper is organized as follows. Sec. 3 defines the guidance task, and introduces an experimental framework used to observe human guidance behavior. Second, Sec. 4 describes the decomposition of observed behavior into primitive elements, which are then described by a policy function and a maneuver automaton model. Next, Sec. 5 shows the application of this model to generate new autonomous trajectories based on the observed example behavior. Finally, Sec. 6 contains concluding remarks.

3. APPROACH

Modeling guidance behavior using a policy function links human behavior with control theory. The guidance policy investigation begins with defining the guidance task.

3.1 Guidance Task

This work focuses on the perceptual guidance process within the hierarchical model of human control, introduced by Mettler and Kong (2012) and Mettler et al. (2014). The perceptual guidance process generates motion between subgoal states based on perception of the environment. In a motion guidance task, an agent must determine a control sequence $u(t)$ that drives a dynamic system, $\dot{x} = f(x, u)$, from a start state $x(t_0)$ to a goal state $x(t_g)$, while satisfying constraints $c(x, u) > 0$. A solution control sequence depends on agent-environment dynamics $f(\cdot, \cdot)$ and environment structure $c(\cdot)$. Warren (2006) expresses the interaction between perception and guidance by coupling