Diagnosing Hybrid Systems: a Bayesian Model Selection Approach

Sheila A. McIlraith
Knowledge Systems Laboratory
Stanford University
Stanford CA 94025
Phone: 650-723-7932, Fax: 650-725-5850
sam@ksl.stanford.edu

Abstract

In this paper we examine the problem of monitoring and diagnosing noisy complex dynamical systems that are modeled as hybrid systems – models of continuous behavior, interleaved by discrete transitions. In particular, we examine continuous systems with embedded supervisory controllers that experience abrupt, partial or full failure of component devices. Building on our previous work in this area (MBCG99; MBCG00), our specific focus in this paper is on the mathematical formulation of the hybrid monitoring and diagnosis task as a Bayesian model tracking and selection problem, and provision of a suitable tracking algorithm. The nonlinear dynamics of many hybrid systems present challenges to probabilistic tracking. Further, probabilistic tracking of a system for the purposes of diagnosis is problematic because the models of the system corresponding to failure modes are numerous and generally very unlikely. To focus tracking on these unlikely models and to reduce the number of potential models under consideration, we exploit logic-based techniques for qualitative model-based diagnosis to conjecture a limited initial set of consistent candidate models. In this paper we discuss alternative tracking techniques that are relevant to different classes of hybrid systems, focusing specifically on a method for tracking multiple models of nonlinear behavior simultaneously using factored sampling and conditional density propagation. To illustrate and motivate the approach described in this paper we examine the problem of monitoring and diagnosing NASA's Sprint AERCam, a small spherical robotic camera unit with 12 thrusters that enable both linear and rotational motion.

Introduction

We have been conducting an ongoing project to investigate how to diagnose hybrid systems – complex dynamical systems whose behavior is modeled as a hybrid system. Following the description in (MBCG99; MBCG00), hybrid models comprise both discrete and continuous behavior. They are typically represented as a sequence of piecewise continuous behaviors interleaved with discrete transitions (e.g., (Bra95)). Each period of continuous behavior represents a so-called *mode* of the system. For example, in the case of NASA's Sprint AERCam, a spherical airborne robot camera unit, modes might include *translate_X-axis*, *rotate_X-axis*, *translate_Y-axis*, etc. (AG98). In the case of an Airbus fly-by-wire system, modes might include *take-off*, *landing*, *climbing*, and *cruise*. Mode transitions generally result in changes to the set of equations governing the continuous behavior of the system, as well as to the state vector that initializes that behavior in the new mode. Discrete transitions that dictate such mode switching are modeled by finite state automata, temporal logics, switching functions, or some other transition system, while continuous behavior within a mode is modeled by, e.g., ordinary differential equations (ODEs), difference equations, or differential and algebraic equations (DAEs). For the purposes of this paper, we restrict our attention to discrete-time estimation for the class of systems whose hybrid models contain no autonomous jumps. I.e., all nominal transitions between system modes are induced by a controller action; none are induced by the system state and mode (Bra95).

In (MBCG99) we presented the hybrid diagnosis problem:

> *Given a hybrid model of system behavior, a history of executed controller actions, a history of observations, including observations of aberrant behavior relative to the model, isolate the fault that is the cause for the aberrant behavior.*

Our task was to perform diagnosis online in conjunction with the continued operation of the system. Hence, we divided our diagnosis task into two stages, initial conjecturing of candidate diagnoses and subsequent refinement and tracking to select the most likely diagnoses. We cast the diagnosis problem as the problem of finding a model and associated parameter values that best fit the data. In that paper we focused on the problem of dealing with the multitude of potential models of the system by exploiting qualitative diagnosis techniques to generate a set of candidate qualitative diagnoses, and we described two parameter estimation techniques to deal with estimating the parameters associated with the model, particularly when erroneous behavior manifested itself some period of time after the initial occurrence of a fault. (See (MBCG00; MBCG99) for details.) We did not discuss the specific problem of tracking multiple candidate models, nor did we discuss how to compare them.

In this paper, we formulate the hybrid monitoring and diagnosis task as a Bayesian model tracking and selection problem (e.g., (Mac91)). In particular, we wish to estimate the state (model) of the system at successive time instants, given a history of observations. The system *diagnosis* is de-

scribed by the value of a specific subset of the state variables – namely those that designate whether components are normal or abnormal, and what their associated parameter values are. We estimate state by tracking the posterior distribution of the state, given the observations.

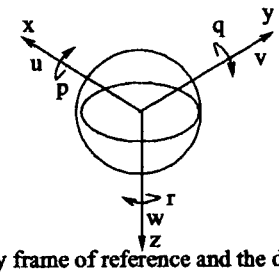
Probabilistic tracking of complex hybrid systems for diagnosis purposes presents a number of interesting challenges. Kalman filtering techniques, traditionally used for tracking linear dynamical systems with Gaussian noise, assume a Gaussian density which is unimodal, making a Kalman filter (Kal60) inadequate for simultaneously tracking alternative candidate models. Multiple Kalman filters, one for each candidate model, can sometimes be used to track multiple candidate models of linear dynamical systems with Gaussian noise (e.g., (Fra90)). More importantly, hybrid systems often have complex nonlinear, nonGaussian and potentially nondeterministic behavior. The nonlinearities come from both the mode switching (faulty or normal modes of behavior), and from the nonlinear dynamics within a mode. The latter has been addressed in some cases by using local linear (Taylor series) approximations of the nonlinear continuous dynamics, such as is done with Extended Kalman Filters (e.g., (BF88)) or Iterated Extended Kalman Filters (e.g., (Jaz70)).

In this paper, following research on bootstrap filters, particle filters and the condensation algorithm (e.g., (GSS93; IB98)), we use a factored sampling technique to sample and represent our multimodal posterior distribution of the state (models) given the observations. Such a technique enables us to track multiple models of nonlinear systems simultaneously. Unfortunately, sampling techniques for probabilistic tracking focus on the most likely models within the distribution, whereas most fault models have low probability, initially. To overcome this bias, we show how to integrate the qualitative diagnosis techniques described in (MBCG00; MBCG99) into the temporal prior of our Bayesian formulation to focus sampling on models that are indicated by our qualitative candidate diagnoses.

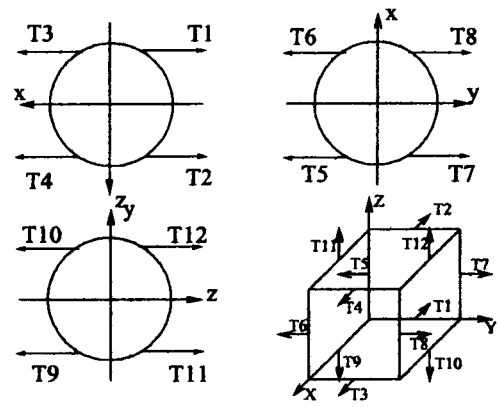
In the next section, we provide a brief description of NASA's Sprint AERCam, which we have used as a motivating example and which we will use to illustrate certain concepts in this paper. In the section that follows the description of the AERCam, we present a formal characterization of the class of hybrid systems we study and the diagnosis problem they present. Next, we describe our Bayesian formulation of the problem and the algorithm we use for computing and propagating posterior distributions. In the final section, we summarize, discuss our continuing research in this area, and reference some related work.

The AERCam

We are using NASA's Sprint AERCam and a simulation of system dynamics and the controller written in Hybrid CC (HCC) (AG98) as a testbed for this work. To make this paper somewhat self-contained, we condense and repeat the description provided in (MBCG99). The AERCam is simpler than many of the complex systems we intend to diagnose, but it serves well in illustrating the concepts developed



The body frame of reference and the directions of velocities (u, v, w) are the components of the translation velocity. (p, q, r) are components of the angular velocity.



Three views of the AERCam, showing the thrusters, and showing all the thrusters together in the cube circumscribing the AERCam.

Figure 1: The AERCam axes and thrusters

here, and has provided an excellent testbed for our preliminary work. We describe the dynamic model of the AERCam system briefly, a more detailed description of the model and simulation appear in (AG98).

The AERCam is a small spherical robotic camera unit, with 12 thrusters that allow both linear and rotational motion (Fig. 1). For the purposes of this model, we assume the sphere is uniform, and the fuel that powers the movement is in the center of the sphere. The fuel depletes as the thrusters fire.

The dynamics of the AERCam are described in the AERCam body frame of reference. The translation velocity of this frame with respect to the shuttle inertial frame of reference is 0. However, its orientation is the same as the orientation of the AERCam, thus its orientation with respect to the shuttle reference frame changes as the AERCam rotates (i.e., it is not an inertial frame). The twelve thrusters are aligned so that there are four along each major axis in the AERCam body frame. For modeling purposes, we assume the positions of the thrusters are on the centers of the edges of a cube circumscribing the AERCam. Thus, for example, thrusters T_1, T_2, T_3, T_4 are parallel to the x-axis and are used for translation along the x-axis or rotation around

the y-axis. I.e., firing thrusters T_1 and T_2 results in translation along the positive x-axis, and firing thrusters T_1 and T_4 results in a negative rotation around the y-axis. AERCam operations are simplified by limiting them to either translation or rotation. Thrusters are either on or off, therefore, the control actions are discrete. In a normal mode of operation, only two thrusters are on at any time.

AERCam dynamics

A simplified model of the AERCam dynamics based on Newtonian laws is derived using an inertial frame of reference fixed to the space shuttle. The AERCam position in this frame is defined as the triple (x, y, z). Let $\vec{\mathbf{V}}$ be the velocity in the AERCam body frame, with its vector components given by (u, v, w). The frame rotates with respect to the inertial reference frame with velocity $\omega = (p, q, r)$, the angular velocity of the AERCam. The rotating body frame implies an additional Coriolis force acting upon the AERCam. We assume uniform rotational velocity since in the normal mode of operation, the AERCam does not translate and rotate at the same time (Arn78, pg. 130). Similar equations can be derived for the rotational dynamics (AG98).

$$d(m\,\vec{\mathbf{V}})/dt = \vec{\mathbf{F}} - 2m(\vec{\mathbf{V}} \times \vec{\omega}) \qquad \text{Newton's Law}$$
$$\vec{\mathbf{V}}\, dm/dt + m d(\vec{\mathbf{V}})/dt = \vec{\mathbf{F}} - 2m(\vec{\omega} \times \vec{\mathbf{V}})$$

The resultant equation for each coordinate:

$$du/dt = F_x/m - 2(qw - vr) - (u/m) * dm/dt$$
$$dv/dt = F_y/m - 2(ru - pw) - (v/m) * dm/dt$$
$$dw/dt = F_z/m - 2(pv - qu) - (w/m) * dm/dt,$$

where the force F on each axis, is a function of the percentage degradation of the thrusters that are exerting force in that direction as specified in Figure 1. Under normal operating conditions, the thrusters operate at 100%.

We use these models to predict the position of the AERCam at time $t+1$, given the position at time t. We add noise to each of the models above. In this case the noise is white Gaussian noise with a mean of zero and a standard deviation σ. As noted above, these models are implemented in HCC and are used to compute the likelihood described in the next section.

Position Control Mode of the AERCam

In the position control mode, the AERCam is directed to go to a specified position and point the camera in a particular direction. Assume the AERCam is at position A and directed to go to position B. In the first phase, the AERCam rotates to get one set of thrusters pointed towards B. These are then fired, and the AERCam cruises towards B. Upon reaching a position close to B, it fires thrusters to converge to B, and then rotates to point the camera in the desired direction.

To facilitate the illustration of the diagnosis problem, we use a simple trapezoidal controller, which we explain in two dimensions. Suppose the task is to travel along the x-axis for some distance, then along the y-axis. Such manoeuvres are needed for navigating in the space shuttle. In order to do this, the AERCam fires its x thrusters for some time. Upon reaching the desired velocity, these are switched off. When the AERCam has reached a position close to the desired x position, the reverse thrusters are switched on, and the AERCam is brought to a halt — the velocity graph is a trapezium. The process is analogous for the y direction.

Problem Formulation

In this section we describe our formulation of the hybrid diagnosis problem. Once again, the hybrid systems we examine are discrete-time hybrid systems. Observations and state estimation are made at regular intervals $1, 2, \ldots, t, t+1, \ldots$. Further, we assume that our systems contain no autonomous jumps. I.e., all nominal transitions between system modes are induced by a controller action, none are induced by the system state and mode (Bra95). Autonomous jumps are common in hybrid models where a mode with complex nonlinear behavior has been simplified by creating multiple modes of less complex behavior, with state-induced autonomous jumps connecting them. Building on the concepts in (MBCG00):

Definition 1 (Hybrid System) A hybrid system is a 5-tuple $\langle \mathcal{M}, X, \Sigma, V, f \rangle$:

- $\mu \in \mathcal{M}$ is the discrete state or mode of the system, where $\mathcal{M}$ is a finite collection of variables. μ_t is the system mode at time t.
- $x \in X \subseteq R^n$ is the continuous state vector of the system. x_t is the continuous state at time t.
- $\sigma \in \Sigma$, is the discrete input, where Σ is a finite collection of actions. I.e., the controller actions that transition the system between modes.
- $v \in V \subseteq R^v$ is the continuous input.
- f is the system dynamics function that maps the mode, the continuous state, and the input into the mode and continuous state at the next discrete time point. $(\mu_{t+1}, x_{t+1}) = f(\mu_t, x_t, \sigma_t, v_t, w_t)$, where $w_t \in R^m$ is zero-mean white noise of known pdf, and $f : \mathcal{M} \times X \times \Sigma \times R^v \times R^m \rightarrow \mathcal{M} \times X$. f is often expressed as a collection of functions, e.g., functions that describe the continuous behavior within a specific mode, and a function that describes the discrete transitions between modes, based on discrete input.
- $obs \in R^p$ is the observation vector of the system. obs_t is the observation vector at time t. obs_t is related to the continuous state vector x_t by the function $obs_t = h(x_t, \nu_t)$ where $\nu_t \in R^r$ is zero-mean white noise of known pdf, and $h : X \times R^r \rightarrow R^p$.

Definition 2 (System State) The state of a hybrid system at time t, (μ_t, x_t) comprises the discrete mode of the system and the continuous state at t.

To define the hybrid diagnosis problem, we augment Definition 1 as follows.

Definition 3 (Diagnosable Hybrid System) A diagnosable hybrid system, $\langle \mathcal{M}, X, \Sigma, V, f, COMPS \rangle$ is a hybrid system comprised of m potentially malfunctioning components $COMPS = (c_1, \ldots, c_m)$ where

- For each $\mu \in \mathcal{M}$, μ includes a designation of whether each $c_i \in COMPS$ is operating normally, or abnormally, i.e., $[\neg]ab(c_i)$.
- For each μ, continuous state vector x includes a set of distinguished parameters θ associated with that mode.
- We assume that transitions to fault modes are achieved by exogenous actions. Hence, $\Sigma = \Sigma_c \cup \Sigma_e$, where
  - Σ_c is a finite set of controller actions, and
  - Σ_e is a finite set of exogenous actions.

We introduce the following additional notation,

- $\mathcal{O}$, designates the observation history, the sequence of time-indexed observations. $\mathcal{O}_l$ designates the observation history to time l.
- μ_F denotes a faulty mode, i.e., a mode for which at least one $c_i \in COMPS$ is $ab(c_i)$ in μ_F. θ_F denotes the parameters associated with μ_F.

In the case of the AERCam example, the potentially malfunctioning components are the 12 thrusters, and a mode μ includes the behavior mode (e.g., translate-x, translate-y, rotate-x, etc.) and $[\neg]ab(T_i), i = 1, \ldots, 12$, for each thruster. The continuous state vector includes the x, y, z position of the AERCam, velocity and acceleration. The parameter values, θ associated with each μ are the percentage degradation of each of the thrusters. As we will see later on, we make a Markov assumption with respect to computing the temporal dynamics of our system. Hence all relevant state must be included explicitly in the state variables.

Definition 4 (Model) A model of a diagnosable hybrid systems is a time-indexed mode sequence and associated parameter values $([\mu_1, \ldots, \mu_m], [\theta_1, \ldots, \theta_m])$. The model to time l is denoted $(\vec{\mu}, \vec{\theta})$ and the model at time l is denoted $s_l = (\mu_l, \theta_l)$. The model is a distinguished subset of the entire system state.

In this paper we make several simplifying assumptions regarding our diagnosis task. In particular, we make a single-time fault assumption. We assume that our systems do not experience multiple sequential faults. Further, we assume that faults are abrupt, resulting in partial or full degradation of component behavior. We cast the hybrid diagnosis task as the problem of finding the most likely model for the observation history, $P(s_l \mid \mathcal{O})$, i.e, the mode and parameter values (μ_l, θ_l) that best fit the observations over time. To do this, we appeal to a Bayesian formulation of the problem.

Bayesian Formulation

To monitor and diagnose a hybrid system, we must compute the posterior probability distribution over models at time l, given the observation history. Recall, using Bayes' rule that the *posterior* is proportional to the *likelihood* times the *prior*. I.e.,

$$p(\text{model} \mid \text{observations}) \propto p(\text{observations} \mid \text{model})\, p(\text{model}).$$

Our objective is to find the posterior probability distribution over models at time l, s_l, given the observation history up to time t, $\mathcal{O}_l$. I.e, we wish to compute $p(s_l \mid \mathcal{O}_l)$.

To compute the temporal dynamics of our system, we make a Markov assumption, i.e.,

$$p(s_l \mid s_{l-1}, \ldots, s_0) = p(s_l \mid s_{l-1})$$

Further, we assume that at each time point, there is a small probability of an exogenous action, leading to a transition to a failure mode. Finally, we assume that given the current model s_l, the current observations obs_l and previous observation history $\mathcal{O}_{l-1}$ are independent.

Hence, in order to track our hybrid system, we can compute the *posterior* distribution of the model at time l given the observation history which, according to Bayes' rule and our assumptions above, is proportional to the *likelihood* of the observation at time l given the model at time l ($p(obs_l \mid s_l)$) and the *temporal prior*, the prediction of the current model, given the observation history up to $l-1$,($p(s_l \mid \mathcal{O}_{l-1})$. I.e.,

$$p(s_l \mid \mathcal{O}_l) = kp(obs_l \mid s_l)\, p(s_l \mid \mathcal{O}_{l-1}),$$

where k ensures that the distribution integrates to one.

The likelihood of the observations given the state is easily evaluated for the AERCam following the model described in the previous section. The temporal prior, i.e., the probability of the current model given the observation history to $l-1$ depends on the posterior over models at the previous time point, $p(s_{l-1} \mid \mathcal{O}_{l-1})$ and the *temporal dynamics*, $p(s_l \mid s_{l-1})$. I.e.,

$$p(s_l \mid \mathcal{O}_{l-1}) = \int_{s_{l-1}} p(s_l \mid s_{l-1})\, p(s_{l-1} \mid \mathcal{O}_{l-1}) ds_{l-1}$$

The temporal prior expresses the probability of a particular model given the observation history up to that point. In the case of a fault diagnosis, the likelihood of a fault model will initially be very low. If we are tracking using a finite number of parallel filters, or using a factored sampling method as suggested in the next section, this may mean that we will initially not track these fault models, or alternately that we track many low probability models which is computationally expensive. In order to focus the temporal prior more quickly and accurately on the appropriate diagnostic models, we make use of qualitative diagnosis techniques.

In (MBCG00; MBCG99), we proposed to use qualitative diagnosis techniques to generate qualitative candidate diagnoses – candidate mode and parameter values that were consistent with observations $\mathcal{O}$ in some window of time.

Definition 5 (D-tuple (MBCG00)) A D-tuple is a 4-tuple $\langle C, \mu_F, t_F, \theta_F \rangle$, where μ_F is a fault mode, t_F is the time the fault mode commenced, θ_F is the parameter values associated with the fault mode behavior, and C is the set of failed (abnormal) components in μ_F.

Definition 6 (Candidate Qualitative Diagnosis (MBCG00)) Given a diagnosable hybrid system with model $(\vec{\mu}, \vec{\theta})$, input history $\mathcal{I}^1$, and observation history, $\mathcal{O}$, D-tuple $\langle C, \mu_F, t_F, \theta_F \rangle$ is a candidate qualitative diagnosis iff there exists a range of parameter values $\theta_F = [\theta_l, \theta_u]$, and time range $t_F = [t_l, t_u]$ such that the occurrence of fault mode μ_F with parameter values θ_F in time range t_F is consistent with $\mathcal{O}$, $\mathcal{I}$ and $(\vec{\mu}, \vec{\theta})$.

We do not repeat the diagnosis algorithms here, but refer the reader to (MBCG00; MBCG99) for details. These generated diagnoses are used to propose a set of different models to be tracked by the system. The candidate models are generated by exploiting previous work on qualitative diagnosis of continuous systems (e.g., (MB99)), adapting the authors' causal propagation algorithms to deal with the discrete state variables and mode transitions of the hybrid systems. To incorporate this so-called oracle into our Bayesian formulation, we use it to bias or focus the temporal prior. This will in turn more heavily weight the posterior for the corresponding fault models, s_l. In the case of particle filtering, the technique we propose in the next section to compute the posterior, this focusing of the temporal prior will help the algorithm sample from the appropriate part of the distribution. To incorporate this qualitative diagnosis "oracle" we may alter our view of the posterior we are computing as follows.

$$\begin{aligned} p(s_l \mid \mathcal{O}_l, \text{oracle}) &\propto p(obs_l \mid s_l, \text{oracle}) \\ &\qquad p(s_l \mid \mathcal{O}_{l-1}, \text{oracle}) \\ &\propto p(obs_l \mid s_l)\, p(s_l \mid \mathcal{O}_{l-1}, \text{oracle}) \end{aligned}$$

where $p(s_l \mid \mathcal{O}_{l-1}, \text{oracle})$ is equal to $p(s_l \mid \mathcal{O}_{l-1})$ above, when the observations are consistent with the current model, and otherwise $p(s_l \mid \mathcal{O}_{l-1}, \text{oracle})$ is simply the normalized probability of the faulty models, given the observations. To ensure the speed of the oracle, and because of the lack of reliable numbers for such calculations, the probabilities generated by the oracle are normalized prior probabilities of different faults given the observations, as defined by the system builder.

Once the posterior is computed, different models can be compared by estimating the expected value of different models, normalizing and comparing. For example, we may sum the likelihoods for all samples having like $[\neg]ab(c_i)$ designations, and compare these to determine which components are likely malfunctioning.

Computing the Posterior

In the previous section we presented the problem of tracking and diagnosing hybrid systems using a Bayesian formulation. As noted in the introduction, there are many algorithms for probabilistic tracking of dynamical systems, though most are not tailored to simultaneously tracking multiple candidate models nor to dealing with nonlinear dynamics. Our posterior distribution $p(s_l \mid \mathcal{O}_l)$ will be a multi-dimensional, multi-modal distribution, reflecting the multiple competing diagnostic models. There is no closed-form (parametric) representation for this distribution, as there is, for example, for a unimodal Gaussian. Consequently, to compute this posterior, we appeal to factored sampling techniques to provide an approximation of the distribution, and project this distribution forward through time according to its dynamics, using the Condensation algorithm (IB98), derivative of the bootstrap algorithm (GSS93) and commonly referred to as a particle filter.

[1] Previously referred to as the action history.

More specifically, the posterior distribution $p(s_l \mid \mathcal{O}_l)$, is represented as a set of N weighted samples $\{\mathbf{s}^{(1)}, \ldots, \mathbf{s}^{(N)}\}$, with associated weights $\{\pi^{(1)}, \ldots, \pi^{(N)}\}$. Intuitively, the larger the N, the better the approximation, but the more costly the computation. Hence we would like to sample the distribution as sparsely as possible, while maximizing our coverage of our distribution, and thus weighting samples more heavily in those parts of the distribution that have greater volume.

At each time step, the basic algorithm comprises three steps: select, predict, and update.

Select: We start with the posterior from the previous time step, $p(s_{l-1} \mid \mathcal{O}_{l-1})$, represented as the factored sample $(\mathbf{s}^{(i)}_{l-1}, \pi^{(i)}_{l-1})$, $i = 1, \ldots, N$. Sample N times with replacement with probability $\pi^{(i)}_{l-1}$, the sample $\{\mathbf{s}^{(i)}_{l-1}\}$, producing the samples $\{\mathbf{s}'^{(i)}_l\}$. Note that samples with high weights may be chosen multiple times.

Predict: For each new sample $\mathbf{s}'^{(i)}_l$, propagate the sample forward according to the dynamics of the system to produce new samples $\{\mathbf{s}^{(i)}_l\}$. In the case of our AERCam, these are the dynamics described in the previous section, together with zero-mean Gaussian white noise. This new set of samples approximates a fair random sample for the effective prior $p(s_l \mid \mathcal{O}_{l-1})$. What remains to compute is the weights.

Update: Compute the weights, $\pi^{(i)}_l = p(obs_l \mid s_l = \mathbf{s}^{(i)}_l)$. From the observations obs_l, evaluate the likelihood of each sample, and normalize the likelihoods of the samples so they sum to 1. I.e.,

$$\pi^{(i)}_l = \frac{p(obs_l \mid \mathbf{s}^{(i)}_l)}{\sum_{n=1}^{N} p(obs_l \mid \mathbf{s}^{(n)}_l)}$$

The above algorithm does not reflect our qualitative diagnosis oracle. In order to suitably focus the temporal prior, we use a linear combination of the samples from the computed temporal prior, and samples from the oracle. This technique was inspired by (BF99), and could also be achieved using importance sampling.

The sample approximation to the distribution, $p(s_l \mid \mathcal{O}_l)$ can be used to compute the expected value for some moment f of the density, for example a mean of some state variable, i.e.,

$$E[f(s_l) \mid \mathcal{O}_l] = \sum_{i=1}^{N} \pi^{(i)}_l f(\mathbf{s}^{(i)}_l)$$

In this way, we can compare the sum of the likelihoods for each distinct model.

Summary and Related Work

In this paper we expanded the hybrid diagnosis framework described in (MBCG99; MBCG00) to present a mathematical formulation and computational techniques for generating diagnoses of hybrid systems in terms of Bayesian tracking and model comparison. We characterized the evaluation of our models (system mode and associated parameter values)

as the computation of the posterior distribution of models given a history of observations. Exploiting a Markov assumption, we showed that this could be computed in terms of the likelihood of the observations at time l, given the model at time l, times a prior. Exploiting the work described in (MBCG99; MBCG00) for generating qualitative diagnoses of hybrid systems, we treated our qualitative monitoring and diagnosis system as an oracle. If the observations were consistent with the current model, then the qualitative monitoring and diagnosis system had no effect on the computation of the posterior. However, if the observations were inconsistent then the oracle would generate a set of candidate diagnoses that would be used to adjust the prior to focus the likelihood computation on that part of the model space that was indicated by the qualitative monitoring and diagnosis engine.

Since hybrid systems are generally nonlinear, and hence the distribution of the posterior multimodal and non-Gaussian, we represented the posterior distribution as discrete samples and exploited factored sampling techniques, used in particle filtering and in the Condensation algorithm, to propagate conditional probability densities over time.

We are still in the early stages of experimenting with these techniques, but preliminary results look promising. Condensation has proven effective for some near realtime visual tracking tasks (e.g., (IB98)), but we anticipate that more complex hybrid systems with large state spaces and partial observability will require further computation and larger amounts of memory that will compromise realtime computation, just as they do, for example, with POMDPs. Such systems will require new variants of many of the techniques we currently employ in model-based diagnosis including exploiting problem decomposition, compact representations of state spaces, abstractions of problems, and approximation of inference. In summary, Bayesian tracking and model comparison and factored sampling techniques for dynamical systems provide a sound mathematical formalism and promising tools for monitoring and diagnosing complex dynamical systems.

The problem of monitoring and diagnosing hybrid systems has received little attention to date, although there is much related work. Within the AI community, there has been a great deal of research on diagnosing static systems (e.g., (HCD92)), while much less on diagnosing discrete dynamical systems (e.g., (CT94; McI98; WN96; BLPZ99)), qualitative diagnosis of continuous systems (e.g.,. (MB99)), and tracking (e.g., (RK99)). Most recently, (LPKB00), have developed related techniques for monitoring and diagnosing *Conditional Linear Gaussian* hybrid systems using a Dynamic Bayes Nets to compactly represent the conditional probability distribution, and proposing algorithms for hypothesis reduction and smoothing. Within the FDI community, the largest proportion of research has focused on diagnosing continuous systems (e.g., (Ger98; Fra90)). These approaches have often used observer schemes and/or Kalman filters to track continuous system behavior. Diagnosis of discrete-event systems has also been studied within the FDI community (e.g, (SSLST96; Lun99)). Nevertheless, our work and the concurrent work of (LPKB00) has been the first to propose a Bayesian tracking approach to diagnosing hybrid systems. Our use of factored sampling techniques and particle filtering drawn from the statistics and computer vision communities, presents a significant contribution to a challenging problem.

Acknowledgements

This work was funded in part by NASA grant NAG2-1337. We would like to thank David Fleet for useful discussion relating to this work.

References

L. Alenius and V. Gupta. Modeling an AERCam: A case study in modeling with concurrent constraint languages. In *Proceedings of the CP'97 Workshop on Modeling and Computation in the Concurrent Constraint Languages*, 1998.

V. I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer Verlag, 1978.

P. Baroni, G. Lamperti, P. Pogliano and M. Zanella Diagnosis of large active systems *Artificial Intelligence*, 110(1):135–183, 1999.

Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*, Academic Press, Boston, 1988.

M. Branicky. *Studies in Hybrid Systems: Modeling, Analysis, and Control*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1995.

M.J. Black and D.J. Fleet. Probabilistic detection and tracking of motion discontinuities. In *IEEE International Conference on Computer Vision*, pp. 551-558, 1999.

M. Cordier and S. Thiébaux. Event-based diagnosis for evolutive systems. Technical Report 819, IRISA, Cedex, France, 1994.

P.M. Frank. Fault diagnosis in dynamic systems using analytic and knowledge-based redundancy: a survey and some new results. *Automatica*, vol. 26, pp. 459-474, 1990.

A. Gelb *Applied Optimal Estimation*. MIT press, Boston, 1974.

J.J. Gertler. *Fault Detection and Diagnosis in Engineering Systems*. Marcel Dekker, New York, 1988.

N. Gordon, D. Salmond, A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation In *IEE Proceedings on Radar, Sonar and Navigation*, Vol 140, No 2, 107-113, 1993.

W. Hamscher, L. Console and J. de Kleer *Readings in Model-based Diagnosis*. Morgan Kaufmann, 1992.

M. Isard and A. Blake. CONDENSATION – conditional density propagation for visual tracking . In *Int. J. Computer Vision*, 29(1):5–28, 1998.

A. Jazwinski. *Stochastic Processess and Filtering Theory*. Academic Press. New York, 1970.

R.E. Kalman A new approach to linear filtering and predition problems. In *Journal of Basic Engineering*, 82:35-45, 1960.

U. Lerner, R. Parr, D. Koller, G. Biswas Bayesian Fault Detection and Diagnosis in Dynamic Systems. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI'2000)*, To appear, 2000.

J. Lunze. A timed discrete-event abstraction of continuous-variable systems. *Intl. Jour. of Control*, vol. 72, no. 13, pp. 1147-1164, 1999.

MacKay, D. J. C. Bayesian interpolation. *Neural Computation* 4(3):415–447, 1991.

S. McIlraith. Explanatory diagnosis: Conjecturing actions to explain observations. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, pp. 167–177, 1998.

S. McIlraith, G. Biswas, D. Clancy and V. Gupta. Hybrid Systems Diagnosis. In *Proceedings of Hybrid Systems: Computation and Control Workshop*, pp. 282–295, 2000.

S. McIlraith, G. Biswas, D. Clancy and V. Gupta. Towards Diagnosing Hybrid Systems. In *Proceedings of the Tenth International Workshop on Principles of Diagnosis (DX'99)*, pp. 193-203, 1999.

P. Mosterman and G. Biswas. Diagnosis of continuous valued systems in transient operating regions. *IEEE Transactions on Systems, Man, and Cybernetics*, 1999. vol. 29, no. 6, pp. 554-565, 1999.

P. Mosterman and G. Biswas. Building hybrid observers for complex dynamic systems using model abstractions. In *International Workshop on Hybrid Systems: Computation and Control*, Nijmegen, Netherlands, March 1999.

B. Rinner and B. Kuipers. Monitoring piecewise continuous behavior by refining trackers and models. In *Hybrid Systems and AI: Modeling, Analysis and Control of Discrete + Continuous Systems*, AAAI Technical Report SS-99-05, pp. 164–169, 1999.

M. Sampath, R. Sengupta, S. Lafortune, K. Sinnamohideen and D. Teneketzis. Failure diagnosis using discrete-event models. *IEEE Trans. on Control Systems Technology*, vol. 4, no. 2, pp. 105-124, 1996.

B. Williams and P.P. Nayak. A model-based approach to reactive self-configuring systems. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, pages 971–978, 1996.