Knowledge-Based Estimation Methods

D.D. Sworder
Dept. of ECE
Univ. of California, San Diego
La Jolla, CA 92093-0407

P.F. Singer and D. Doria
Hughes Aircraft Co.
El Segundo, CA 90245

R.G. Hutchins
ORINCON Corp.
San Diego, CA 92121

Abstract

At its most basic level, an estimation algorithm is a causal mapping from a spatiotemporal observation to an approximation of a primary process. To gain insight into the idiosyncracies of knowledge-based estimators, and to contrast them with those of their more established brethren, this paper considers the performance of an image-enhanced estimator in a specific scenario; an agile target moving in the plane with intermittent maneuvers is tracked from a fixed location.

1. Introduction

At its most basic level, an estimation algorithm is a causal mapping from a spatiotemporal observation (the measurement) to an approximation of a primary process (the signal or state depending on the context). Estimators have become increasingly sophisticated as more versatile online processors have become available. If the signal has a suitable structure, improved estimation and prediction is achieved through model-based synthesis procedures. These approaches, as their name suggests, use a comprehensive model to delineate both the signal dynamics and the precise relationship between the signal and its measurement. Through the model, the estimator is tuned to subtle patterns in the measurement, thus enabling good performance to be achieved in the presence of significant measurement ambiguity.

Perhaps the most widely studied model-based algorithm is the Kalman filter and its lineal variants. In a tracking problem, the dynamic features of the target would be represented by a state process \( \{x_t\} \) having as components; positions, velocities and perhaps accelerations in an appropriate reference frame. The observation process \( \{y_t\} \) would be derived from a radar or a similar device that provides a measurement—albeit a noisy one—of the center-of-reflection location of the target; e.g., range and bearing. To illustrate these ideas in the context of a specific example, consider the problem of tracking the motion of an agile object in the X-Y plane. A simple model for an evasive target moving at essentially constant speed is;

\[
\begin{align*}
    \dot{x} &= 0.01x, \\
    \dot{y} &= 0.01y, \\
    \dot{v_x} &= 0.001 - \omega_x, \\
    \dot{v_y} &= 0.002 - \omega_y.
\end{align*}
\]

where \( \{x_t,y_t\} \) are position coordinates, and \( \{v_x,v_y\} \) are associated velocities. The target is subject to two types of acceleration; a wide band omnidirectional acceleration described by the Brownian motion \( \{w_x,w_y\} \) of intensity \( \sigma \), and a maneuver acceleration represented by the turn rate process \( \{\omega_x,\omega_y\} \). The former fits well within the conventional linear-Gauss-Markov (LGM) framework, but the latter is troublesome in two respects. First, because of their temporal structure, the sample paths of the maneuver accelerations are better described by a discontinuous rather than a continuous process, and are clearly non gaussian. Secondly, the turn enters into the target dynamics as a state multiplier, and consequently, renders the dynamic equation nonlinear in the augmented state space.

In many applications, the measured relations between the indicated variables are nonlinear. With a smooth nonlinearity, it is possible to linearize the measurement equation about the estimated state. If the linear (or linearized) equations are adequate, there is a well known solution to the mean-square inference problem; the (extended) Kalman filter (EKF). Let the measurement equation be given by

\[
\begin{align*}
    \dot{y}_t &= D_x \Delta t + d_n,
\end{align*}
\]

and denote the information pattern generated by the sensor measurements by \( \{Y_t\} \). The best estimate of the state \( \hat{x}_t = E(x|Y_t) \) satisfies an equation of the form:

\[
\begin{align*}
    \dot{\hat{x}}_t &= A_{\hat{x}} \Delta t + P_{\hat{x}} D_R^{-1} d_{\Delta t},
    \quad \text{with } d_{\Delta t} = D_x \Delta t + d_n,
\end{align*}
\]

the error covariance matrix:

\[
\begin{align*}
    \quad \quad (d/dt) P_{\hat{x}} &= \hat{x} A_{\hat{x}} + P_{\hat{x}} A^{T} + W - P_{\hat{x}} D_R^{-1} D_P_{\Delta t},
    \quad \text{subject to appropriate initial conditions.}
\end{align*}
\]

The error covariance is contingent upon the intensity of the exogenous processes in both state and observation; e.g., as \( \sigma \) increases, the increment in \( P_{\hat{x}} \) increases proportionately. This has an intuitive justification. As the state process becomes
more volatile, $P_{\text{obs}}$ increases, and through this intermediary, the EKF becomes "faster" and more responsive to state changes. Of course, as a corollary to this, the same filter will amplify the measurement noise [$n_i$].

Recently, alternative tracking architectures have become feasible as novel sensors have been developed. Video and FLIR sensors create a sequence of pictures of a scene containing the target. No longer is the observation restricted to a point-equivalent target, but instead, features of spatial extent can be extracted from the image. The raw image data is received as a sequence of matrices of grey levels, generated in packets or frames at a fixed frame rate $\lambda$ frames/sec. An image processor recasts the data, and extracts relevant features from it. The processor acts as a pattern classifier, placing each image in one of a prespecified set of topical bins. The information obtained from the image is often complementary to that obtained from a point-location sensor; e.g., target type and orientation can be measured from shape data rather than being inferred from target motion. However, the errors inherent in the image link are quite distinct from those found in the conventional observation links. The relevant errors are misclassifications as occur when the target is placed in feature bin $i$ despite the fact that the correct choice is bin $j$. Misclassifications depend upon the fidelity of the image and the sophistication of the processing, and these in turn depend upon depend on range and geometry, the angular bin size, the sensitivity of the image elements and the processing algorithm, etc.

Figure 1: Image-enhanced tracking architecture

The maneuvering target problem has received considerable attention in the literature where it is usually assumed that estimation and prediction is accomplished using bearing and perhaps range data. The simplest estimators use an EKF derived by replacing the turn rate process with additive white noise, with additive colored noise, or by augmenting $W$ with pseudonoise. It will be assumed here that image augmentation is exploited. Figure 1 shows a tracking architecture proposed in several references which is useful for integrating the information derived from an imager with that from conventional sensors. The lower path shows the usual mapping from center-of-reflection observations to an estimate of target location. The upper path is image-specific, and uses observations of target shape to augment the point-location data in the lower path.

To gain insight into the idiosyncracies of image-enhanced estimators, and to contrast them with those of their more established brethren, this paper considers the performance of an image-enhanced estimator in a specific scenario; an agile target moving in the plane with intermittent maneuvers is tracked from a fixed location. This application is fundamentally nonlinear and nongaussian, and illustrates both the promise and the limitations associated with image augmentation. The next section introduces some of the issues that arise in image generation and interpretation in a tactical environment.

2. Knowledge-based Image interpretation

To implement upper path in Figure 1, the orientation of a target in $R^3$ must be inferred from the image. A 2-D projection provides the only data upon which this inference can be made. Unfortunately, distortions in both the visible and the infrared bands can cause changes in the apparent shape and internal structure of the target. For example, smoke and contrails in the visible band and plumes in the infrared band can cause the ostensible contour of an aircraft to differ from that of a prestored model of the target. Shadows or internal reflections of sunlight can cause local edges in the interior of a visible image. Internal heating or external heat sources can likewise change the internal distribution of thermal signatures. Especially for ground vehicles, the likelihood of partial occlusion is significant in some situations. Pattern recognition approaches that involve global features of the object (e.g., moment invariants or Fourier descriptors) degrade when confronted with even local changes in perceived shape.

To achieve robust performance in the face of these disturbances, a model based approach to target matching that utilizes local features of the observed target signature has been adopted. This approach employs the Generalized Hough Transform (GHT). The GHT algorithm encodes the target contour into a prestored table (the R-table) with distance and the angle to a selected reference point used as indices. For each sample image, the R-table features of a measured contour are computed. The sorting process proceeds by matching computed R-features with prestored points associated with each view. The data-to-model matching calculation yields an estimate of the model reference point position and the rotation. The algorithm accomplishes the match by "voting" for a particular value of the position and orientation of the target. The votes are accumulated as a three dimensional histogram in what is called P-space. Measured range is used to scale the match.
The votes resulting from ambiguous data tend to spread out in an amorphous region of P-space. The correct data-to-model matchings tend to be more concentrated. From the histogram the position and orientation of the target reference point can be estimated.

To evaluate the feasibility of estimating the orientation of a moving target, an experiment on actual data set was taken from the IMPRINT (Image Processing for the Identification of Non-Cooperative Targets) data base. Hughes Aircraft Company under the sponsorship of the Naval Air Development Center in Warminster, Pa., collected video imagery of military aircraft and evaluated the usefulness of a variety of pattern recognition techniques for target identification. For a subset of the IMPRINT imagery, target orientation coordinates were recorded from an on-board inertial navigation system (INS). These were used to evaluate the errors in the estimated target orientation obtained with the GHT. Figure 2 shows the sequence of imagery used in this experiment, in which an F-14 is undergoing a rolling maneuver.

The "truth" model was obtained using plastic scale models of the anticipated targets mounted on a mechanism that included computer controlled gimbals such that the aircraft was able to be rotated about its yaw and roll axes. The quantization of both the yaw and roll axes of the model imagery was 30 degrees in this experiment. Using the best fit to the model in yaw and roll the orientation of the longitudinal axis of the aircraft can be estimated along with aircraft type. To obtain an estimate of the accuracy of the classification procedure, the angle between the unit vectors obtained from the GHT and from the INS ground truth was calculated for each video frame. It was found that the mean difference in the angles was 16.6 degrees with a standard deviation of 8.7 degrees.

3. Image-Based Models

To integrate the image information into an estimation algorithm like that of the EKF, the encounter model must be completed. There are two accelerations influencing the target path. Because the turn rate tends to be nearly constant over intervals with sudden changes at unpredictable times, a useful model is created by partitioning the range of turn rates into discrete levels. Let \( \alpha_t \) be the maneuver indicator process; \( \alpha_t = 0 \), if \( \alpha_t = a \), at time \( t \). The target is described by a family of \( K \) stochastic differential equations indexed by the maneuver acceleration. The most common way to quantify the dynamics of the maneuver process is to suppose the successive maneuver modes are represented by a Markov chain with transition rate matrix \( Q \).

Before analyzing a tracker with the full sensor suite, consider an architecture which has no image path; i.e., suppose the tracker avails itself of just the lower path in Figure 1 using a range-bearing measurement. These measurements will be taken at a fixed rate (10 samples/sec.) with independent errors. A 4-dimensional filter is sufficient to generate the state estimate (position and velocity). The filter gain depends parametrically on the intensity of the exogenous disturbance, and the filter will be denoted by EKF. A slight increase in model complexity gives a more realistic motion equation. Rather than neglecting the turn rate, it can be replaced with a Gaussian surrogate. The turn rate shaping filter is a two parameter family, with \( \tau \) the time constant, and \( R(0) = E((\alpha_t)^2) \). The resulting estimator will be called EKFs.

In the dual-path tracking architecture shown in Figure 1, the sensor suite contains an imager in addition to the conventional range-bearing sensors. As pointed out earlier, the imager is a feature matching block, and its errors are not well described by the additive Gaussian noise. To create a more realistic model of the orientation classifier, suppose that the imager collects data at a rate of \( \lambda \) frames/sec. and places each target image into one of \( L \) equally spaced orientation bins. The output of the processor can be written as an \( L \)-dimensional counting process \( \{N_i\} \), the ith component of which is the number of times the target has been placed in bin \( i \) on the interval \([0,t]\). This sequence of counts (or symbols) can be interpreted by a temporal processor to give the relative likelihoods of the various turn rate hypotheses. This development is carried out in [1]. The quality of the imager is determined by the \( L \times L \)-discernibility matrix \( P = [p_{ij}] \) where \( p_{ij} \) is the probability that bin \( i \) will be selected by the processor if bin \( j \) contains the true target orientation at time of image creation; i.e., (see [2])

The fidelity of image interpretation is a function of three things: the frame rate \( \lambda \), the ability of the image to
correctly classify a single image (P), and the tempo of the encounter (q). In [3], an equation for the conditional probabilities of the turn rate was derived, and used to modify an EKF. This filter will be called EKF_p. To illustrate the contrasts in performance, consider a specific vignette with a target initially located at \( (x_0,y_0) = (6.4, 1.0) \text{KM} \), and initial velocity \( (v_{x0},v_{y0}) = (5.0,-13.3) \text{ meters/sec} \). Moving at constant velocity for \( t \in [0, 30) \text{ sec.} \), a 0.5 g turn is made for \( t \in [30,38) \), after which the target returns to constant velocity motion. The omnidirectional accelerations will be assumed to be slight; \( (aw_1^2,aw_2^2) = (0.01 \text{ dt (m/sec}^2))^2 \). Figure 3 shows the target path (truth) without the wide band acceleration. Tracking will be accomplished with a 10 samples/sec. range-bearing sensor located at the origin with Gaussian errors of standard deviation 0.5 meters and 0.25” respectively, independent in time and type. If an imager is available, it will be collocated, and make measurements at the same rate, placing the target in angular bins of width 22.5”. It will be supposed that the lateral accelerations in the maneuver model are given by \( (aw_1 = 20^\circ/\text{sec. turn right}, aw_2 = \text{no turn}, aw_3 = 20^\circ/\text{sec. turn left}) \), and that the chain \( (q) \) is symmetric about the coasting mode; i.e., the probabilistic character of \( q = 1 \) and 3 are identical. The elements of the q-matrix are determined jointly by the mean sojourn times in each acceleration mode, and the likelihood of alternative modal transitions.

Both the motion model and the observation equation are nonlinear. Furthermore, if the vignette is partitioned into pre-, post-, and maneuver coincident segments, the relative performance of the trackers differs as a function of the portion of the path that is being reviewed. A good indication of performance can, however, be determined by simulation. Figure 3 shows the position response of selected estimators. To display tracker biases, a 20-trial mean sample path is shown rather than a (noisier) single sample path. All of the filters perform well in the premaneuver phase, but both EKF and EKF_p have difficulty following a turn; the latter making a pirouette at apogee. Only EKF_p performs satisfactorily in the postmaneuver initiation phase of the engagement.

Figure 3: Tracker performance in the plane.

Figure 4: Velocity estimates for different trackers.

Figure 4 shows sample-mean velocity profiles, and the nonimaging trackers are not good. The reason for the clearly degraded velocity estimates is the fact that velocity is a slack variable in the estimation algorithm. It is included in the model, and the EKFs estimate its value concurrently with estimates of the location variables, but there is no direct velocity measurement. The filter tends to assign observation residuals to slack variables to a much greater degree than one might expect. This is true to some extent during the nonmaneuvering phase, \( (aw_1 = 0) \), where the model underlying the EKF is valid, but the effect is magnified during a maneuver. In this phase of the scenario, there is conspicuous misidentification of the velocity by both EKF and EKF_p, with neither recognizing the correct sense of rotation; the pirouette yields a velocity profile that is the mirror image of the actual path. This poor performance in velocity is magnified by prediction algorithms, the simplest of which extrapolate to the predicted position along the direction of the velocity estimate. The small velocity errors in EKF_p portend much improved prediction fidelity.

4. Performance Sensitivities

As noted earlier, the fidelity of the upper link in Figure 1 depends on the number of bins, the processor fidelity and the frame rate. Although \( L \) and \( \lambda \) are fixed by design, \( P \) changes during an encounter. The discernibility matrix is clearly dependant on range because target extent changes as a function of the square of range. A target which sub-
tends few pixels offers little spatial structure to the imager, and the imager will tend to classify orientation arbitrarily; i.e., \( q \) will tend to be uniformly distributed.

As the target approaches the imager, and as more target pixels become available, silhouette projection errors come to the fore. The target silhouette is compatible with two distinct orientation angles, thus creating an equivocal situation for the processor. When the number of pixels is large enough, this projection ambiguity can be resolved from an analysis of the internal features of the target. For the encounter discussed in the previous section, such errors had little influence on performance. This was due to two special aspects of the target path. The initial estimate of orientation was correct, and the kinematics were sufficiently simple to prevent subsequent confusion regarding the projection of the image. Further, the geometry of the path was such as to avoid having the target change its turn while broadside to the sensor. Such situations create maximum projection effect.

![Figure 5: Estimation error as a function of discernibility.](image)

Nearest neighbor errors are important in certain circumstances, but were not studied in detail here. At long ranges, uniform errors are dominant, while at short ranges, projection errors become so. The region in P-space most closely associated with nearest neighbor errors is in an intermediate zone. Note however that nearest neighbor errors are very important since a move to an adjacent bin is the initial signature of a maneuver. False alarm rates are sensitive to the probability of landing in the neighboring bin.

Figure 5 shows the mean 5-second prediction errors on this scenario for different levels of uniform error. For a sample of size ten, the degradation of performance is apparent, but even when the probability of error is 70% the predictor performs well. Above this level, performance falls off quickly, and the hybrid filter/predictor reduces to the EKF (also shown) when the uniform error reaches 100%.

5. Conclusions

New sensors require novel architectures to capture their potential for performance improvement. It has been shown that the response of an image-enhanced tracking loop is superior to that attainable from conventional trackers without imaging capability. From this study, it is evident that an image enhanced tracker can achieve good quiescent performance without a severe penalty during turn. This exchange is difficult to make with conventional algorithms because they tend to balance improvements in one part of the encounter with degradation elsewhere.

Acknowledgements

This research was partially supported by a grant from the Hughes Aircraft Co., and by the MICRO Program of the State of California under Project No. 91-156.

References

