

A Caratheodory–Fejer Approach to Robust Multiframe Tracking

Octavia Camps

Hwasup Lim

Cecilia Mazzaro

Mario Sznaiar

Department of Electrical Engineering

The Pennsylvania State University

University Park, PA 16802

Abstract

A requirement common to most dynamic vision applications is the ability to track objects in a sequence of frames. This problem has been extensively studied in the past few years, leading to several techniques, such as Unscented Particle Filter based trackers, that exploit a combination of the (assumed) target dynamics, empirically learned noise distributions and past position observations. While successful in many scenarios, these trackers remain fragile to occlusion and model uncertainty in the target dynamics. As we show in this paper, these difficulties can be addressed by modeling the dynamics of the target as an unknown operator that satisfies certain interpolation conditions. Results from interpolation theory can then be used to find this operator by solving a convex optimization problem. As illustrated with several examples, combining this operator with Kalman and UPF techniques leads to both robustness improvement and computational complexity reduction.

1. Introduction

A requirement common to most dynamic vision applications is the *ability to track* objects in a sequence of frames. This problem has been extensively studied in the past few years, leading to several techniques. Some of these techniques can track unknown objects [9, 11, 14, 17, 18, 20, 31], while others require prior knowledge of the target [4, 15, 23, 32, 34]. Orwell et. al. [25] and Collins et. al. [12] use color to track objects with multiple cameras. Hager and Toyama [19] track primitive features within small regions of interest (ROI) that are warped and matched against canonical configurations. Reid and Murray [28] use affine structure to track clusters of corners. Calabi *et al.* [9] use differential invariant signature curves to track objects. Blake and Isard [5] use active contours and geometrical constraints to model the likelihood of their deformations.

Correspondences between individual frames are usually integrated over time to improve robustness by exploiting the dynamical properties of the target. Kalman filter–based trackers use a model of the target dynamics and the probability distribution of the process and measurement noise

to produce estimates of the future positions of the target based on (noisy) measurements of its past locations. Condensation trackers and unscented Kalman Filters [6, 24, 21] generalize Kalman–filter based ones by allowing more general (multimodal, nonlinear) models. In this case, analytical propagation is not longer possible and numerical methods must be used instead.

Most trackers assume a simple dynamic model such as a system moving with constant velocity. While successful in many scenarios, this approach suffers from the fact that the tracker must rely on the assumed model of the target dynamics to produce estimates of its future positions, introducing a potential source of fragility. A mismatch between this model and the actual dynamics will lead to incorrect predictions¹. This lack of robustness is shown in Figure 1 illustrating the effects of clutter. As shown there, both a regular Kalman filter based tracker and an Unscented Particle Filter (UPF), lose the target in frame 95, due to occlusion.

It should be noted that more precise dynamic models have been tried in the particular case of human motion tracking. For example, models based on biomechanics have been successfully used to produce computer-based animations [8] and track humans [33]. However, biomechanical models are usually very complex and difficult to estimate from visual data alone. An alternative approach was proposed in [27] where models are learned from a training corpus of observed state space trajectories. A drawback of this approach is that the states of the system must be specified a priori, which requires a priori knowledge of the order of the model.

In addition, due to their stochastic nature, Kalman and UPF based approaches can provide neither a region guaranteed to contain the target nor falsify the information about the distribution. That is, if this information is rendered obsolete, for instance by the target entering a region where the clutter is substantially different from the one used in training, a probabilistic approach would typically not be able to establish with certainty that the present scenario is no longer compatible with the assumptions.

¹This is the well known divergence phenomenon, see for instance [1], page 133.

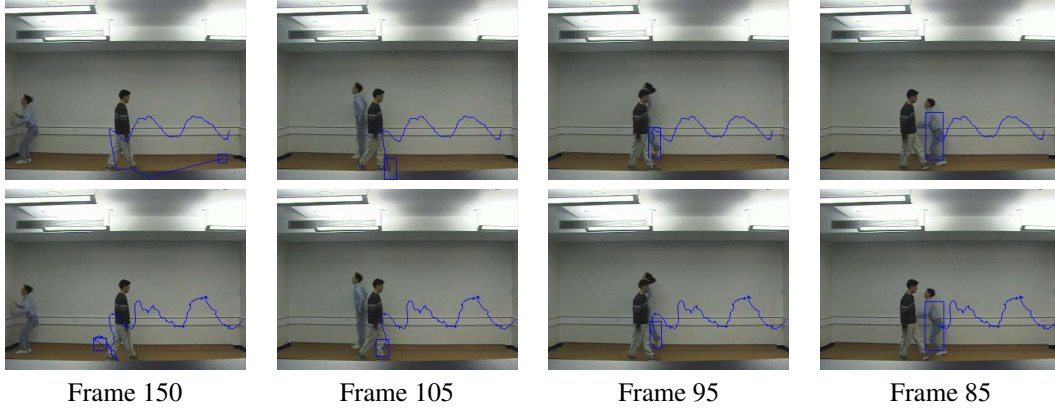


Figure 1: Kalman (top) and Unscented Particle Filter (bottom) based tracking in the presence of occlusion

In this paper we show that all of the above issues can be addressed by reducing the tracking problem to that of establishing the existence of an ℓ_2 to ℓ_2 operator that satisfies certain interpolation conditions. This allows for exploiting convex analysis and integral quadratic constraints methods recently developed (mostly in the control community) to recast the problems into a Linear Matrix Inequality (LMI) optimization form that can be efficiently solved using commercially available tools.

The benefits of using this new framework are multiple. Firstly, it allows to approach the tracking problem from an input – output point of view and thus it does not require prior knowledge of a state space realization of the system, or even its order. Secondly, it provides mechanisms to invalidate a priori assumptions about the dynamics of the target and the noise characterization. Thirdly, it provides worst-case estimates of the identification error that can be used both to determine for how long the predictions of the target position are valid and, in the context of *robust* filters such as mixed $\mathcal{H}_2/\mathcal{H}_\infty$ [29], to improve tracking robustness. Finally, as we illustrate with several examples, the proposed method can be combined with existing particle and unscented Kalman filtering techniques leading to algorithms capable of robustly tracking targets in the presence of severe occlusion. When compared to existing techniques this combination allows for significantly improving robustness, while at the same time reducing the computational complexity of the resulting algorithm.

2. Preliminaries

2.1. Notation

Below we summarize the notation used in this paper:

\mathbf{x}	real-valued (unless otherwise stated) column vector.
x_k	k^{th} element of a vector \mathbf{x} .
$\ \mathbf{x}\ _p$	p -norm of a vector: $\ \mathbf{x}\ _p \doteq (\sum_{k=1}^m x_k ^p)^{\frac{1}{p}}$, $p \in [1, \infty)$, $\ \mathbf{x}\ _\infty \doteq \max_{k=1, \dots, m} x_k $.
A^T	conjugate transpose of matrix A .
$A_{i,j}$	(i, j) element of A .
$(A)_i$	i^{th} row of A .
$\bar{\sigma}(A)$	maximum singular value of the matrix A .
$A > 0$	$A = A^T$ is positive definite, i.e. $\mathbf{x}^T A \mathbf{x} > 0 \forall \mathbf{x} \in \mathbf{C}^n, \mathbf{x} \neq \mathbf{0}$.
$\mathcal{B}\mathcal{X}(\gamma)$	open γ -ball in a normed space \mathcal{X} : $\mathcal{B}\mathcal{X}(\gamma) = \{x \in \mathcal{X} : \ x\ _{\mathcal{X}} < \gamma\}$
$\overline{\mathcal{B}\mathcal{X}(\gamma)}$	closure of $\mathcal{B}\mathcal{X}(\gamma)$
$\mathcal{B}\mathcal{X}(\overline{\mathcal{B}\mathcal{X}})$	open (closed) unit ball in \mathcal{X}
(\mathcal{X}, m)	metric space of elements in \mathcal{X} equipped with the metric $m(x_1, x_2)$.
$d(\mathcal{A})$	diameter of $\mathcal{A} \subseteq \mathcal{X}$: $d(\mathcal{A}) \doteq \sup_{x, a \in \mathcal{A}} m(x, a)$.
ℓ_p^m	extended Banach space of vector valued real sequences equipped with the norm: $\ x\ _p \doteq (\sum_{i=0}^{\infty} \ x_i\ _p^p)^{\frac{1}{p}}$ $p \in [1, \infty)$ $\ x\ _\infty \doteq \sup_i \ x_i\ _\infty$.
\mathcal{L}_∞	Lebesgue space of complex-valued matrix functions essentially bounded on the unit circle, equipped with the norm: $\ G\ _\infty \doteq \text{ess sup}_{ z =1} \bar{\sigma}(G(z))$.

\mathcal{H}_∞	subspace of functions in \mathcal{L}_∞ with bounded analytic continuation inside the unit disk, equipped with the norm: $\ G\ _\infty \doteq \text{ess sup}_{ z <1} \bar{\sigma}(G(z))$.
$\mathcal{H}_{\infty,\rho}$	space of transfer matrices analytic in $ z \leq \rho$, equipped with the norm $\ G\ _{\infty,\rho} \doteq \sup_{ z <\rho} \bar{\sigma}(G(z))$.
$X(z)$	Z -transform of a right-sided real sequence $\{x\}$: $X(z) = \sum_{i=0}^{\infty} x_i z^i$.

2.2. Definitions and Additional Notation

From an input–output viewpoint any operator of interest H will be represented by its convolution kernel $\{h_{i,j}\}$ or by an infinite lower triangular matrix \mathbf{T}_H mapping (scalar) sequences in ℓ_2 :

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} h_{0,0} & 0 & 0 & \dots \\ h_{1,0} & h_{1,1} & 0 & \dots \\ h_{2,0} & h_{2,1} & h_{2,2} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \end{bmatrix}. \quad (1)$$

When dealing with input–output sequences on the horizon $[0, n-1]$, we will use the finite upper left submatrix of $n \times n$, \mathbf{T}_H^n , obtained from the infinite matrix above.

In the sequel, we will also represent finite dimensional Linear Time Invariant (LTI) operators by using either a minimal state–space realization:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{A}\mathbf{x}_k + \mathbf{B}u_k \\ y_k &= \mathbf{C}\mathbf{x}_k + \mathbf{D}u_k. \end{aligned} \quad (2)$$

or a (rational) complex–valued transfer function:

$$H(z) \doteq \sum_{k=0}^{\infty} h_k z^k,$$

Definition 1. A set $\mathcal{A} \subset \mathcal{X}$ is called symmetric if there exists an element $c \in \mathcal{X}$ such that for any $a \in \mathcal{X}$ for which $c + a \in \mathcal{A}$ then $c - a \in \mathcal{A}$. The element c is called the symmetry point of set \mathcal{A} (note that it may not belong to \mathcal{A}).

2.3 Background results on interpolation

The following results will be used in the paper to establish the existence of operators with the appropriate features.

Lemma 1 (Carathéodory-Fejér). Given a matrix valued sequence $\{\mathbf{L}_i\}_{i=0}^{n-1}$, there exists a causal, discrete-time, LTI operator $L(z) \in \mathcal{BH}_\infty$ such that

$$L(z) = \mathbf{L}_0 + \mathbf{L}_1 z + \mathbf{L}_2 z^2 + \dots + \mathbf{L}_{n-1} z^{n-1} + \dots \quad (3)$$

if and only if

$$(\mathbf{T}_L^n)^T \mathbf{T}_L^n \leq \mathbf{I} \quad (4)$$

where \mathbf{I} denotes the identity matrix of compatible dimension.

Proof. See for instance [16], Chapter 1 \square

In the sequel we will consider operator families of the form \mathcal{S} :

$$\mathcal{S} \doteq \{G(z) = H(z) + P(z)\}. \quad (5)$$

where $H(z) \in \mathcal{BH}_\infty(K)$ and $P(z)$ represent the nonparametric and parametric components of the operator, respectively. We will further assume that $P(z)$ belongs to the following class \mathcal{P} of affine operators:

$$\mathcal{P} \doteq \{P(z) = \mathbf{p}^T \mathbf{G}_p(z), \mathbf{p} \in \mathcal{R}^{N_p}\}, \quad (6)$$

where the N_p components $\mathbf{G}_{p_i}(z)$ of vector $\mathbf{G}_p(z)$ are known, linearly independent, rational transfer functions. The following result gives a necessary and sufficient condition for two finite vector sequences to be related by an operator in the family \mathcal{S} .

Lemma 2. Given K , and two vector sequences (\mathbf{u}, \mathbf{y}) , there exists an operator $S \in \mathcal{S}$ such that $\mathbf{y} = \mathbf{S}\mathbf{u}$ if and only if there exists a vector \mathbf{h} satisfying:

$$\begin{aligned} M(\mathbf{h}) \doteq \begin{bmatrix} \mathbf{I} & (\mathbf{T}_h^N)^T \\ \mathbf{T}_h^N & \frac{\mathbf{I}}{K^2} \end{bmatrix} \geq 0 \\ \mathbf{y} = \mathbf{T}_u \mathbf{P} \mathbf{p} - \mathbf{T}_u \mathbf{h} \end{aligned} \quad (7)$$

where $(\mathbf{P})_k \doteq [g_k^1 \ g_k^2 \ \dots \ g_k^{N_p}]$, where g_k^i denotes the k -th Markov parameter of the i -th transfer function $G_{p_i}(z)$ and h_k is the k -th Markov parameter of the nonparametric component $H(z)$. Moreover, in this case all such operators S can be parameterized in terms of a free parameter $Q(z) \in \mathcal{BH}_\infty$. In particular, the choice $Q(z) = 0$ leads to the “central” model $S_{\text{central}}(z) = H_o(z) + \mathbf{p}^T \mathbf{G}_p(z)$ where an explicit state–space realization of $H_o(z)$ is given by:

$$H_o(z) = \mathbf{C}_H (z\mathbf{I} - \mathbf{A}_H)^{-1} \mathbf{B}_H + \mathbf{D}_H$$

$$\mathbf{A}_H = \left\{ \mathbf{A} - [\mathbf{C}_-^T \mathbf{C}_- + (\mathbf{A}^T - \mathbf{I})^{-1} \mathbf{C}_-^T \mathbf{C}_- (\mathbf{A} - \mathbf{I})] \right\}^{-1}$$

$$\mathbf{B}_H = [\mathbf{C}_-^T \mathbf{C}_- (\mathbf{A}^T - \mathbf{A} - \mathbf{I}) - (\mathbf{A}^T - \mathbf{I}) \mathbf{A}]^{-1} \mathbf{C}_-^T$$

$$\mathbf{C}_H = \mathbf{K} \mathbf{C}_+ -$$

$$\mathbf{K} \mathbf{C}_+ \left\{ \mathbf{A} - [\mathbf{C}_-^T \mathbf{C}_- + (\mathbf{A}^T - \mathbf{I})^{-1} \mathbf{C}_-^T \mathbf{C}_- (\mathbf{A} - \mathbf{I})] \right\}^{-1}$$

$$\mathbf{D}_H = \mathbf{K} \mathbf{C}_+ \left\{ [\mathbf{C}_-^T \mathbf{C}_- + (\mathbf{A}^T - \mathbf{I}) \mathbf{A}] - \mathbf{C}_-^T \mathbf{C}_- (\mathbf{A} - \mathbf{I}) \right\}^{-1} \mathbf{C}_-^T, \quad (8)$$

with

$$\mathbf{A} = \begin{bmatrix} 0 & \mathbf{I}_{N \times N} \\ 0 & 0 \end{bmatrix}, \quad \mathbf{C}_- = \overbrace{[\mathbf{1} \ 0 \ \dots \ 0]}^{N+1}, \quad \mathbf{C}_+ = \frac{\mathbf{h}^T}{K}. \quad (9)$$

Proof. See Theorem 18.5.2 in [2] and [26]. \square

Corollary 1. [26] Consider the problem of identifying an operator $S \in \mathcal{S}$ from measurements of its output y to a known input $u \in \ell_2[0, N]$, corrupted by additive bounded noise η in a given set \mathcal{N} :

$$y_k = (S * u)_k + \eta_k, \quad k = 0, 1, \dots, N \quad (10)$$

Then there exist $S \in \mathcal{S}$ that satisfies (10) if and only there exists a pair of vectors (\mathbf{h}, \mathbf{p}) such that $M(\mathbf{h}) > 0$ and $\mathbf{y} - \mathbf{T}_u \mathbf{P} \mathbf{p} - \mathbf{T}_u \mathbf{h} \in \mathcal{N}$. In that case, one such operator is given $S_{\text{central}} = \mathbf{p}^T \mathbf{G}_p + H_o$, where H_o has the state-space realization (8).

3. Multiframe Tracking as an Interpolation Problem.

In this section we show that the problem of robustly tracking an object in a sequence of frames is equivalent to finding an ℓ_2 to ℓ_2 operator that satisfies certain interpolation conditions. This approach allows for appealing to IQCs [22], convex analysis and interpolation concepts to recast these problems into a tractable LMI optimization form.

As mentioned in the introduction, in principle, the location of the target can be predicted using a combination of its (assumed) dynamics, empirically learned noise distributions and past position observations [6]. However, as shown in Figure 1 this process is far from trivial in a cluttered environment. As shown there both, Kalman-based and Unscented Particle Filter based trackers, begin to track poorly in frame 95, and by frame 105 have completely lost the target due to a combination of occlusion and the use of dynamics that do not exactly match the exact dynamics of the target. As we show next, these difficulties can be solved by modelling the motion of the target as the output of an ARMA model and using the results in section 2.3 to identify the relevant dynamics.

Specifically, assume that the present position of a given target feature, f_k is related to its past N values by²:

$$\begin{aligned} f_k &= \mathcal{A}f + \mathcal{B}e \\ y_k &= f_k + \eta_k \end{aligned} \quad (11)$$

where $\mathbf{f} = (f_{k-1} \dots f_{k-N})^T$ contains the past observations of the feature, $\mathbf{e} = (e_k \ e_{k-1} \dots e_{k-m})$ represents a stochastic input, y_k denotes the available measurement of the feature, corrupted by noise η_k , and where \mathcal{A} and \mathcal{B} are suitable LTI operators. Alternatively, (by taking z -transforms in the equation above), one can use the description:

$$y(z) = \mathcal{F}(z)e(z) + \eta(z) \quad (12)$$

²For simplicity, we consider a single feature, but the equations generalize trivially to the multiple-feature case.

where the operator \mathcal{F} is not necessarily ℓ_2 stable. In the sequel, we will assume that the following *a priori* information is available:

1. A set membership description of the measurement and process noise: $\eta_k \in \mathcal{N}$ and $e_k \in \mathcal{E}$. These sets can be used to impose correlation constraints.
2. The operator \mathcal{F} admits a finite expansion of the form $\mathcal{F} = \underbrace{\sum_{j=1}^n p_j \mathcal{F}^j}_{\mathcal{F}_p} + \mathcal{F}_{np}$. Here \mathcal{F}^j are known, given, not necessarily ℓ_2 stable operators that contain all the information available about possible modes of motion of the target³. An example of this situation is tracking moving persons where the \mathcal{F}^j can be obtained off-line by training with a representative set of motions [15, 3].
3. The residual operator $\mathcal{F}_{np} \in \mathcal{BH}_{\infty, \rho}(K)$ for some known $\rho \leq 1$. That is, a worst case estimate is available of how fast the approximation error of the finite expansion $\mathcal{F}_p = \sum_{j=1}^n p_j \mathcal{F}^j$ diverges.

In this context, the next location of the target feature f_k can be predicted by first identifying the relevant dynamics \mathcal{F} and then using it to propagate its past N values. In turn, identifying the dynamics entails finding an operator $\mathcal{F}(z) \in \mathcal{S} \doteq \{\mathcal{F}(z): \mathcal{F} = \mathcal{F}_p + \mathcal{F}_{np}\}$ such that $y - \eta = \mathcal{F}e$, precisely the class of interpolation problem addressed in Corollary 1. By noticing that $H(z) \in \mathcal{BH}_{\infty, \rho} \iff H(\frac{z}{\rho}) \in \mathcal{BH}_{\infty}$, it follows that such an operator exists if and only if the following set of equations is feasible:

$$\begin{aligned} \mathbf{M}_R(\mathbf{h}) &= \begin{bmatrix} \mathbf{R}_\rho^2 & \mathbf{T}_h^T \\ \mathbf{T}_h & K^2 \mathbf{R}_\rho^{-2} \end{bmatrix} \geq 0 \\ \mathbf{y} - \mathbf{T}_u \mathbf{P} \mathbf{p} - \mathbf{T}_u \mathbf{h} &\in \mathcal{N} \end{aligned} \quad (13)$$

where $\mathbf{R}_\rho \doteq \text{diag}[1 \ \rho \ \dots \ \rho^N]$.

In addition to providing an estimate of the next position of the target, this approach also has the following advantages:

1. **Model (in)validation:** Assume that the set \mathcal{N} is described by a set of LMIs of the form:

$$\mathcal{N} \doteq \{\boldsymbol{\eta} \in \mathbb{R}^N : \mathbf{L}(\boldsymbol{\eta}) = \mathbf{L}_0 + \sum_{k=1}^N \mathbf{L}_k \eta_{k-1} \geq 0\} \quad (14)$$

where \mathbf{L}_i are given real-valued symmetric matrices. This noise set is a generalization of the set $\{\boldsymbol{\eta} \in \mathbb{R}^N : |\eta_k| \leq \epsilon\}$ usually considered [10], that allows for

³If this information is not available the problem reduces to purely non-parametric identification by setting $\mathcal{F}^j \equiv 0$.

taking into consideration correlated noise. Then equations (13)-(14) reduce to a set of LMIs in the variables \mathbf{h}, η and K^2 . This allows for finding the minimum value of K^2 (recall that K is an upper bound of the ℓ_2 induced norm of the non-parametric part of the operator, F_{np}) such that the LMIs (13)-(14) are feasible. In turn, this value can be used as a “sanity check” to assess the quality of the approximation. A large value of K indicates that the non-parametric portion of the model \mathcal{F}_p does not provide a good description of the motion of the feature, hence the need for a large non-parametric part, indicating that it may be necessary to re-identify the set $\{\mathcal{F}^i\}$. Infeasibility of the LMIs indicates that the experimental data is not compatible with the *a priori* assumptions, possibly indicating either (i) a new target activity not described by elements of the set $\{\mathcal{F}^i\}$ or (ii) the target entering a region where the noise and clutter models are no longer compatible with the description (14). Either case points to the need for re-assessing the *a priori* information.

2. **Worst-case estimates of the prediction error.** By construction, the operator found from the solution to the LMIs (13) is such that its response to the input \mathbf{e} interpolates, within the experimental noise level η_k , the given location of the feature f_k , $k = 0, 2, \dots, N - 1$. However, when used to predict the *future* location of the feature, it is of interest to obtain bounds on the worst case prediction error. This can be accomplished as follows: Given a sequence $\{y_k\}_{k=0}^{N-1}$ of measurements of the location f_k of the feature, define the consistency set as:

$$\mathcal{T}(\mathbf{y}) \doteq \{F \in \mathcal{S} : \{y_k - (F * e)_k\}_{k=0}^{N-1} \in \mathcal{N}\} \quad (15)$$

i.e., the set of all models consistent with both the *a priori* information and the experimental data. Note that the proposed method is interpolatory, that is, it always generates a candidate operator $F_{id} \in \mathcal{T}(\mathbf{y})$. Thus, since the “true” operator F_o that maps the input \mathbf{e} to the feature locations \mathbf{f} must also belong to the consistency set⁴, it follows that, given the first N measurements y_i , $i = 0, \dots, N - 1$ a bound on the worst case prediction error over the horizon $[0, M - 1]$, $M > N$, is given by:

$$\begin{aligned} & \|\hat{\mathbf{f}} - \mathbf{f}\|_{\ell_\infty[0, M-1]} \\ & \leq \sup_{F_1, F_2 \in \mathcal{T}(\mathbf{y})} \|F_1 * e - F_2 * e\|_{\ell_\infty[0, M-1]} \\ & = d[\mathcal{T}(\mathbf{y})] \\ & \leq \sup_{\mathbf{y}} d[\mathcal{T}(\mathbf{y})] = \mathcal{D}(\mathcal{I}) \end{aligned} \quad (16)$$

⁴As long as the *a priori* information is indeed correct.

where $d(\cdot)$ and $\mathcal{D}(\mathcal{I})$ denote the diameter of the set $\mathcal{T}(\mathbf{y})$, in the $\ell_\infty[0, M - 1]$ metric and the diameter of information, respectively. Moreover, since the *a priori* sets $(\mathcal{S}, \mathcal{N})$ are convex and symmetric, with points of symmetry $F_s = 0$ and $\eta_s = 0$ respectively, it can be shown (see for instance Lemma 10.3 in [30]) that:

$$\mathcal{D}(\mathcal{I}) \leq 2 \sup_{F \in \mathcal{S}(0)} \|F\|_{\ell_\infty[0, M-1]} \quad (17)$$

where $\mathcal{S}(0)$ indicate the set of operators compatible with the zero outcome: $y_k = 0$, $k = 0, 1, \dots, N - 1$. As we will illustrate in the sequel with a simple example, computing this bound reduces to a convex optimization problem.

4. Illustrative Examples

In this section we illustrate the potential of the proposed method with several examples. In the first one we consider, for the sake of simplicity, static tracking, and indicate how to use the proposed approach to both predict future locations of the target and obtain *worst case* bounds on the prediction error. In the second example we show how to combine the proposed approach with existing Kalman and UPF techniques to improve robustness.

4.1. Inter-Frame Tracking and Prediction

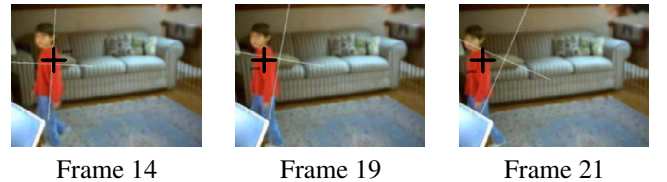


Figure 2: Robust identification based tracking (black cross) versus Mean Shift (white cross)

In this example we consider the problem of predicting the location of the centroid of the child shown in Figure 2, from past measurements of its coordinates, (x_k, y_k) , corrupted by uncorrelated noise, η . For the sake of brevity we report below only the results for the x coordinate, since those for y are similar.

The following *a priori* information was used:

1. $\mathcal{N} = \{\eta \in \ell^\infty, \|\eta\|_{\ell^\infty} \leq 5.5\}$ ⁵.
2. $\mathcal{E} = \delta(0)$, i.e. motion of the target was modelled as the impulse response of the unknown operator F ⁶.

⁵This value was quantified from fluctuations in the data taken when the person was at rest.

⁶This is equivalent to lumping together the dynamics of the plant and the input signal.

3. The parametric part of the model $F_p \in \text{span}(\mathbf{G})$, $\mathbf{G}(z) \doteq \left[\frac{z^2}{z^2-2z+1}, \frac{z}{z^2-2z+1} \right]^T$.
4. The reminder, nonparametric component, which explains the unmodelled dynamics satisfies $F_{np} \in \mathcal{BH}_{\infty, \rho}(K)$, with $\rho = 0.99$.

The experimental, *a posteriori* data consisted of the first $N = 12$ frames of the sequence. The resulting LMI problem was solved using MATLAB's LMI Toolbox, leading to $K_{opt} = 1.35e^{-12}$ and $\mathbf{p} = [127.7763 \quad -135.0723]^T$. Note that the very low value of K indicates that indeed the parametric part F_p provides an accurate model of the dynamics of the target.

The advantage of this approach is illustrated in Figure 2, comparing the position of the centroid predicted by our model (black crosses) against the results of using a Mean Shift based tracking (white crosses) implemented in Intel's Open Source Computer Vision Library [7]. Although Mean Shift is designed to improve tracking robustness by exploiting color information [13], it begins to track poorly in frame 19, and by frame 21 it has completely lost the target due to a combination of clutter and moderate occlusion. The corresponding numerical values of the error, computed as the difference between the predicted and actual values⁷ are given in Table 3. As shown there, the identified model is able to predict the location of the target, far beyond the point where the Mean Shift tracker has failed.

Finally, notice that in this case computing the worst case prediction error bound (17) reduces to a Linear Programming problem in $\mathbf{p} = [p_1 \cdots p_{N_p}]^T$ and $\mathbf{h} = [h_0 \cdots h_N]^T$. The last row in Table 3 shows the error bounds as a function of k . As expected these values increase with time, since no new data is being used beyond $k = 12$. However, they became comparable with the width of the target (30 pixels) only beyond $k = 20$.

4.2. Improving Robustness of Kalman and UPF Trackers

In this example we illustrate how the proposed approach can be used to improve the robustness of trackers, such as Kalman and UPF, that rely on a combination of past measurements and the dynamics of the target to estimate its future location. Proceeding as in the previous example, we used a combination of *a priori* information:

(a) 5% noise level

(b) $\mathcal{F}_p \in \text{span} \left[\frac{1}{z-1}, \frac{z}{z-1}, \frac{z}{(z-1)^2}, \frac{z^2}{(z-1)^2}, \frac{z^2 - \cos \omega z}{z^2 - 2 \cos \omega z + 1}, \frac{\sin \omega z^2}{z^2 - 2 \cos \omega z + 1} \right]$

⁷These values were obtained using off-line image processing.

and measurements from the first 32 frames, where the target is not occluded, to estimate its dynamics. This dynamics was then used in conjunction with a Kalman filter, leading to the results shown in Figure 4. Contrary to the case illustrated in Figure 1, the algorithm is now able to track the target past the occlusion. Similar results are shown at <http://robustsystems.ee.psu.edu>. A salient feature of these results is the fact that the combination Caratheodory–Fejer/Kalman Filter achieved virtually the same performance as CF/UPF and substantially outperforms the UPF alone. Thus the proposed approach can both improve robustness and alleviate the computational complexity of the problem.

It is also worth mentioning that consistent experience suggests that, for a given target and gait type, it is not necessary to re-identify the dynamics of the target for each sequence. For instance, the results at the bottom of Figure 4 were obtained using the dynamics identified using the top sequence.

5 Conclusions

In the past few years dynamic vision techniques have proved to be a viable option for a large number of applications, ranging from surveillance and manufacturing to assisting individuals with disabilities. Arguably, at this point one of the critical factors limiting widespread use of these techniques is the potential fragility of the resulting systems. In this paper we show that in the case of multiframe tracking this fragility can be addressed by using interpolation and LMI tools recently developed in the control community to recast these problems into a tractable optimization form. The advantages of this approach, and in particular its potential to result in robust tracking algorithms when combined with existing UPF techniques was illustrated with several experimental results. Research is currently underway seeking to combine the proposed technique with mixed $\mathcal{H}_2/\mathcal{H}_\infty$ filters in an attempt to better capture the properties of the problem.

It is also worth mentioning that there are important cases where this approach does not provide a complete solution, since the resulting problem is not convex on all the variables involved, leading to an NP hard optimization problem. An example of this situation is the case where both the dynamics of the plant and its input must be identified, and the experimental data is corrupted by measurement noise. We are currently exploring the possibility of overcoming this difficulty by combining the approach pursued in this paper with risk-adjusted (in)validation methods.

Sample	13	14	15	16	17	18	19	20
Mean-Shift	25.90	35.93	41.32	45.63	54.65	57.53	65.05	64.80
Id-based	8.87	6.14	10.04	13.03	10.31	15.72	19.50	26.04
Worst case bound	13.00	15	17	19	21	23	25	27

Figure 3: Id error as a function of k . Target width is 30 pixels.

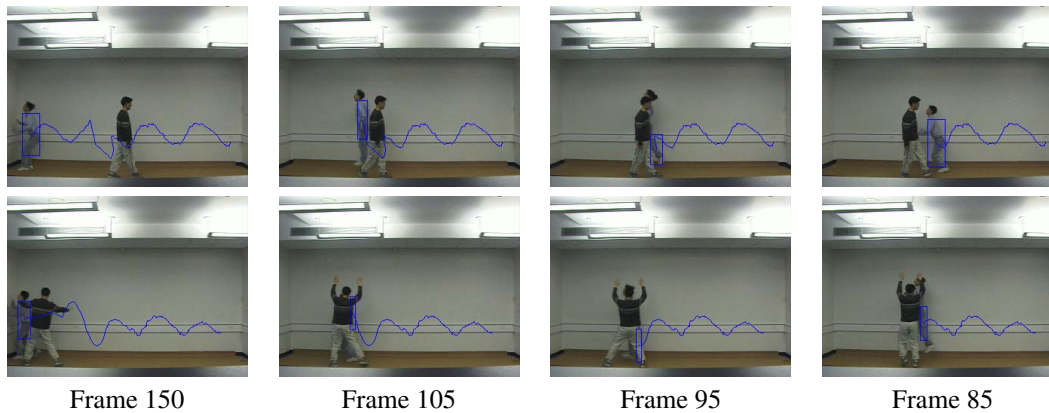


Figure 4: Combination Kalman and CF based tracking in the presence of occlusion

References

- [1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, New Jersey, 1979.
- [2] J. Ball, I. Gohberg, and L. Rodman. *Interpolation of Rational Matrix Functions, Operator Theory: Advances and Applications*, volume 45. Birkhäuser, Basel, 1990.
- [3] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto. Recognition of human gaits. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, USA, December 2001.
- [4] M. J. Black and A. D. Jepson. Eigenttracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal of Computer Vision*, 26(1):63–84, 1998.
- [5] A. Blake and M. Isard. *Active Contours*. Springer Verlag, Berlin, 1998.
- [6] A. Blake and M. Isard. Condensation - condensation density propagation for visual tracking. *Int. J. Comp. Vision*, 29(1):5–28, 1998.
- [7] G. R. Bradski and V. Pisarevsky. Intel's computer vision library: applications in calibration, stereo, segmentation, tracking, gesture, face and object recognition. In *IEEE CVPR*, volume II, pages 796–797, 2000.
- [8] C. Bregler. Learning and recognizing human dynamics in video sequences. In *cvpr*, pages 568–574, San Juan, Puerto Rico, June 1997.
- [9] E. Calabi, P. J. Olver, C. Shakiban, A. Tannenbaum, and S. Haker. Differential and numerically invariant signature curves applied to object recognition. *International Journal of Computer Vision*, 26(2):107–135, 1998.
- [10] J. Chen, C. Nett, and M. Fan. Worst-case system identification in \mathcal{H}_∞ : Validation of *a Priori* information, essentially optimal algorithms and error bounds. *IEEE Transactions on Automatic Control*, 40(7):1260–1265, July 1995.
- [11] L. D. Cohen. On active contour models and balloons. *Computer Vision Graphics and Image Processing: Image Understanding*, 53(2):211–218, 1991.
- [12] R. Collins, O. Amidi, and T. Kanade. An active camera system for acquiring multi-view video. In *Int. Conf. on Image Processing*, volume I, pages 517–520, 2002.
- [13] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *IEEE CVPR*, pages 142–149, June 2000.
- [14] D. Coombs and C. Brown. Real-time binocular smooth pursuit. *International Journal of Computer Vision*, 11(2):147–164, 1993.

- [15] D. J. Fleet, M. J. Black, Y. Yacoob, and A. D. Jepson. Design and use of linear models for image motion analysis. *International Journal of Computer Vision*, 36(3): 171–194, February/March 2000.
- [16] C. Foias and A. E. Frazho. *The commutant lifting approach to interpolation problems*, *Operator theory: Advances and Applications*, volume 44. Birkhäuser, 1990.
- [17] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee. Using adaptive tracking to classify and monitor activities in a site. In *IEEE Computer Vision and Pattern Recognition*, pages 22–29, 1998.
- [18] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, 1997.
- [19] G. Hager and K. Toyama. A new method for the non-linear transformation of means and covariances in filters and estimators. *IEEE Trans. on Automatic Control*, 45(3):477–482, 2000.
- [20] M. Irani and P. Anandan. Unified approach to moving object detection in 2d and 3d scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(6): 577–589, June 1998.
- [21] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte. XVI-SION: A portable substrate for real-time vision applications. *Computer Vision Graphics and Image Processing: Image Understanding*, 69(1):23–37, 1998.
- [22] A. Megretski and S. Treil. Power distribution inequalities in optimization and robustness of uncertain systems. *Journal of Mathematical Systems, Estimation and Control*, 1993.
- [23] S. K. Nayar, H. Murase, and S. A. Nene. Learning, positioning, and tracking visual appearance. In *IEEE International Conference on Robotics and Automation*, pages 3237–3246, May 1994.
- [24] B. North, A. Blake, M. Isard, and J. Rittscher. Learning and classification of complex dynamics. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(9):1016–1034, September 2000.
- [25] J. Orwell, P. Remagnino, and G. A. Jones. Multi-camera color tracking. In *2nd IEEE Int. Workshop on Visual Surveillance*, Fort Collins, CO, June 1999.
- [26] P.A. Parrilo, R. Sánchez Peña, and M. Sznaiier. A parametric extension of mixed time/frequency robust identification. *IEEE Trans. Aut. Control*, 44(2):364–369, February 1999.
- [27] V. Pavlovic, J. M. Rehg, T.J. Cham, and K. P. Murphy. A dynamic bayesian network approach to figure tracking using learned dynamic models. In *iccv*, pages 94–101, Corfu, Greece, Sep. 1999.
- [28] I. D. Reid and W. Murray. Active tracking of foveated feature clusters using affine structure. *International Journal of Computer Vision*, 18(1):41–60, April 1996.
- [29] H. Rotstein, M. Sznaiier, and M. Idan. $\mathcal{H}_2/\mathcal{H}_\infty$ filtering: Theory and an aerospace application. *International Journal of Robust and Non-Linear Control, Special Issue on \mathcal{H}_∞ and Robust Estimation*, 6(4): 347–366, 1995.
- [30] R. Sánchez Peña and M. Sznaiier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.
- [31] J. Shi and C. Tomasi. Good features to track. In *IEEE Computer Vision and Pattern Recognition*, pages 593–600, 1994.
- [32] C. R. Wen, A. Azarbayejani, T. Darrell, and A. P. Pentland. Pfunder: real-time tracking of the human body. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):780–785, July 1997.
- [33] C. R. Wren and A.P. Pentland. Dynamic models of human motion. In *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 22–27, Nara, Japan, Apr. 1998.
- [34] Y. Yacoob and L. S. Davis. Learned models for estimation of rigid and articulated human motion from stationary or moving camera. *International Journal of Computer Vision*, 36(1):5–30, 2000.