Proceedings of the 42nd IEEE
Conference on Decision and Control
Maui, Hawaii USA, December 2003

Operator Theoretic Methods for Robust Active Vision Problems

Octavia Camps Cecilia Mazzaro Brian Murphy Mario Sznaier
Department of Electrical Engineering
The Pennsylvania State University
University Park, PA 16802

Abstract— **A key issue that must be addressed when deploying active vision systems in unstructured, potentially hostile environments is their fragility. As we show in this paper, robustness issues can be addressed by appealing to a common systems theoretic substrate to recast these problems into a tractable convex optimization form. These ideas are illustrated with experimental results from multiframe tracking, visual servoing and activity recognition.**

I. Introduction and Motivation

Recent hardware developments have rendered active vision a viable option for a very diverse spectrum of applications ranging from MEMS manufacture [6] to assisting individuals with disabilities [15], and Intelligent Vehicle Highway Systems [12]. Indeed, computer vision and control are already linked through many successful proof–of–concept systems[1].

However, active vision techniques have been applied outside controlled environments in relatively few instances. To a large extent, this can be traced to the *fragility* of the resulting systems when confronted with unstructured environments. *Thus, the essential challenge in designing and deploying active vision systems is to reconcile their potential fragility with the precise information requirement to accomplish the tasks outlined above,* and to do so within the constraints imposed by the need for real time operation in uncertain environments. In this paper we show that the *fragility* arising in many seemingly dissimilar active vision problems can be addressed by appealing to a common systems theoretic substrate to reduce the problem to analyzing the existence of a bounded ℓ_2 to ℓ_2 operator that satisfies certain interpolation conditions. While the details are somewhat different in each case, this allows for exploiting convex analysis and integral quadratic constraints methods to recast the problems into a LMI optimization form that can be efficiently solved using commercially available tools.

[1]See for instance `http://robustsystems.ee.psu.edu`

II. Interpolation Problems in Active Vision

In this section we show that many computer vision problems such as robustly tracking an object in a sequence of frames, robust visual servoing and recognizing human activities are equivalent to analyzing the existence of a bounded ℓ_2 to ℓ_2 operator that satisfies certain interpolation conditions.

A. Multiframe tracking

A requirement common to most active vision applications is the *ability to track* objects in a sequence of frames. 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 [4]. However, this process is far from trivial in a cluttered environment.

Figure 1 shows the results of using a Mean Shift based tracking (white crosses). Although this algorithm is designed to improve tracking robustness by exploiting color information [5], it begins to track poorly in frame 18, and by frame 20 it has completely lost the target due to a combination of clutter and moderate occlusion. As we show next, this difficulty can be solved by modelling the motion of the target as the output of an ARMA model and identifying the relevant dynamics.

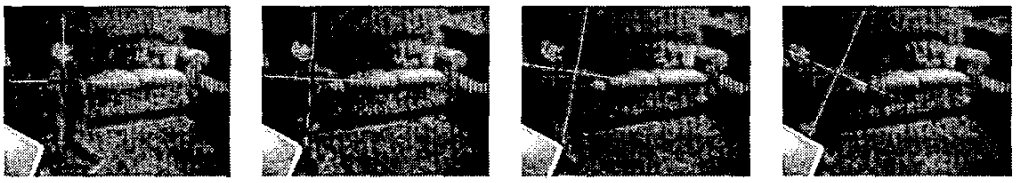
1) Multiframe tracking as an interpolation problem: Assume that the present position of a given target feature is given by:

$$y(z) = \mathcal{F}(z)e(z) + \eta(z) \tag{1}$$

where $\mathbf{e} = \begin{pmatrix} e_k & e_{k-1} & \dots & e_{k-m} \end{pmatrix}$ represents a stochastic input, y_k denotes the available measurement of the feature, corrupted by noise $\eta_k \in \mathcal{N}$, and 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:

(a) 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.

0-7803-7924-1/03/$17.00 ©2003 IEEE



Frame 8 Frame 14 Frame 19 Frame 21

Fig. 1. Robust identification based tracking (black cross) versus Mean Shift (white cross)

(b) The operator $\mathcal{F}$ admits a finite expansion of the form $\mathcal{F} = \overbrace{\sum_{j=1}^{n} p_j \mathcal{F}^j}^{\mathcal{F}_p} + \mathcal{F}_{np}$. 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 target[2]. 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 [7, 3].

(c) The residual operator $\mathcal{F}_{np} \in \mathcal{BH}_{\infty,\rho}(K)$ for some known $\rho \leq 1$. That is, a bound on how fast the approximation error of the finite expansion $\mathcal{F}_p = \sum_{j=1}^{n} p_j \mathcal{F}^j$ diverges is available.

In this context, the next location of the target feature y_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 $F(z) \in \mathcal{S} \doteq \{F(z)\colon F = F_p + F_{np}\}$ such that $y - \eta = \mathcal{F}e$, precisely the class of interpolation problem addressed in [10]. 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 in the variables $\mathbf{p}, \mathbf{h}$ and K is feasible:

$$\mathsf{M}_R(\mathrm{h}) = \begin{bmatrix} \mathsf{R}_\rho^2 & \mathsf{T}_h^T \\ \mathsf{T}_h & K^2 \mathsf{R}_\rho^{-2} \end{bmatrix} \geq 0 \qquad (2)$$

$$\mathbf{y} - \mathsf{T}_u \mathsf{P} \mathbf{p} - \mathsf{T}_u \mathbf{h} \in \mathcal{N} \qquad (3)$$

where T_x denotes the Toeplitz matrix associated with a given sequence $\mathbf{x} = [x_1, \ldots, x_n]$, $\mathsf{R}_\rho \doteq \mathrm{diag}\,[1\ \rho\ \cdots\ \rho^N]$, $\mathsf{P} \doteq [f^1\ f^2\ \cdots\ f^{N_p}]$,where f^i is a column vector containing the first n Markov parameters of the i-th transfer function $F^i(z)$, $\mathbf{h}$ contains the first n Markov parameters of $F_{np}(z)$ and K is an upper bound of the ℓ_2 induced norm of the non–parametric part of the operator, F_{np}.

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:

$$\mathcal{N} \doteq \{\eta \in \Re^N\colon \mathsf{L}(\eta) = \mathsf{L}_0 + \sum_{k=1}^{N} \mathsf{L}_k \eta_{k-1} \geq 0\} \qquad (4)$$

where L_i are given real–valued symmetric matrices. Then equations (2)-(4) reduce to a set of LMIs in the variables h, η and K^2. This allows for finding the minimum value of K^2 such that the LMIs (2)-(4) 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 $\mathcal{F}_{np}$ does not provide a good description of the motion of the feature, 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 (4). 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 (2) is such that its response to the input e interpolates, within the experimental noise level η_k, the given location of the feature f_k, $k = 0, 2, \ldots, N-1$. 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:

$$\mathcal{T}(\mathrm{y}) \doteq \{F \in \mathcal{S}\colon \{y_k - (F * e)_k\}_{k=0}^{N-1} \in \mathcal{N}\} \qquad (5)$$

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

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

$F_{id} \in \mathcal{T}(\mathbf{y})$. 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 set[3], it follows that, given the first N measurements y_i, $i = 0, \ldots, N-1$ a bound on the worst case prediction error over the horizon $[0, M-1]$, $M > N$, is given by:

$$\|\hat{\mathbf{f}} - \mathbf{f}\|_{\ell_\infty[0,M-1]} \leq \sup_{\mathbf{y}} d[\mathcal{T}(\mathbf{y})] = \mathcal{D}(\mathcal{I}) \quad (6)$$

where $d(.)$ 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 [11]) that:

$$\mathcal{D}(\mathcal{I}) \leq 2 \sup_{F \in \mathcal{S}(0)} \|F\|_{\ell_\infty[0,M-1]} \quad (7)$$

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

2) A Simple Example: Consider again the problem of predicting the location of the centroid of the child shown in Figure 1, from past measurements of its coordinates, (x_k, y_k), corrupted by uncorrelated noise, η. For the sake of briefness we report below only the results for the x coordinate, since those for y are similar.

The following *a priori* information was used:

1) $\mathcal{N} = \{\eta \in \ell^\infty, \|\eta\|_{\ell_\infty} \leq 5.5\}$.
2) $\mathcal{E} = \delta(0)$, i.e. motion of the target was modelled as the impulse response of the unknown operator F.
3) The parametric part of the model $F_p \in \text{span}(\mathbf{G})$, $\mathbf{G}(z) \doteq [\frac{z^2}{z^2-2z+1}, \frac{z}{z^2-2z+1}]^T$.
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 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 $\mathbf{p} = [127.7763 \; -135.0723]^T$. 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 1 where the black crosses indicate the position of the centroid predicted by our model. The numerical values of the error, computed as the difference between the predicted and actual values are given in Table 2. 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.

Sample	14	16	18	20
Mean–Shift	35.93	45.63	57.53	64.80
Id–based	6.14	13.03	15.72	26.04
Worst case bound	15	19	23	27

Fig. 2. Id error as a function of time.

Finally, the last row in Table 2 shows the error bounds as a function of the frame number k, computed by solving a LP problem in $\mathbf{p}$ and $\mathbf{h}$. 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 $k = 20$.

B. Visual Servoing

Perhaps the more direct connection between control and active vision occurs in visual servoing problems, where vision is directly used as a sensor in a closed loop configuration. Consider for example the problem of smooth tracking of a non–cooperative target, illustrated in the block diagram shown in Figure 3. Here the goal is to internally stabilize the plant and to track target motions, y_{target}, using as measurements images possibly corrupted by noise, while zooming in and out of features of interest.

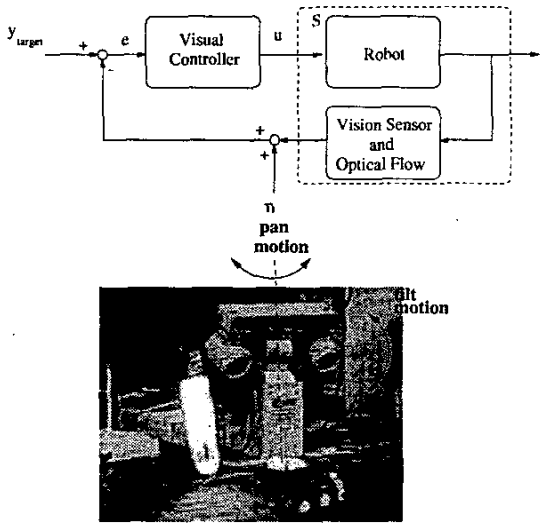


Fig. 3. (top) Block diagram of a visual tracking system, (bottom) The experimental setup.

[3]As long as the *a priori* information is indeed correct.

Designing a controller for this application requires, as a first step, finding a model for the block labeled S in Figure 3(top) that maps the command input to the head, in encoder units, to the position of the target (in pixels) [14]. This map depends on the time–varying focal length f of the lenses, unknown *a priori*, but measurable in real time.

1) Identifying the dynamics of the plant: Physical considerations, corroborated by experimental data, [14], suggest that the combined dynamics of the plant/image processing can be represented by a model of the form:

$$\begin{aligned} y &= (G_p + G_{np})u + \eta \\ G_{np} \in \mathcal{BH}_{\infty,\rho}(K),\ G_p &= \sum_{i=1}^{N_p} p_i \mathcal{F}_u(G_i, \Upsilon) \end{aligned} \tag{8}$$

where y is the measured position of the target corrupted by noise $\eta \in \mathcal{N}$. In this model p_i are unknown scalars, $G_i(z)$ are known, stable transfer matrices and $\Upsilon = \mathrm{diag}\,(v_1 I_{r_1}, \ldots, v_s I_{r_s})$ is a set of time–varying parameters, in this case the focal lenght, that are unknown *a priori*, but can be measured in real time. In this context, identifying the dynamics of the plant can be precisely stated as:

Problem 1. *Given experimental data* $(\mathbf{y}, \mathbf{u}, \Upsilon)$ *and a set description of the measurement noise* $\eta \in \mathcal{N}$, *find* p *and* G_{np} *so that the resulting Linear Time Varying operator* $G_o = \sum_{i=1}^{N_p} p_i \mathcal{F}_u(G_i, \Upsilon) + G_{np}$ *satisfies:*

$$\mathbf{y} - \mathsf{T}^N_{G_o} u \in \mathcal{N}$$

that is, the operator interpolates the (finite) experimental data within the experimental noise levels.

This problem belongs to the class of interpolation problems addressed in [13] and thus can be reduced to an LMI optimization problem.

2) Experimental Validation: Next, we illustrate the theory above by using these tools to design a robust LPV controller for the setup shown in Figure 3 (bottom). In this case experimental data indicates that the parametric component of the LPV model $\mathcal{F}_u(G_p, \Upsilon)$ can be modelled using just one transfer function, i.e. $p_1 \mathcal{F}_u(G_1, \Upsilon)$, and that its dependence with the time varying parameter v_1 can be considered to be affine. Regarding the non–parametric component G_{np}, based on the time–constant obtained with experiments involving only the mechanical components of the system, we determined a value of $\rho = 1.5$ for the *a priori* stability margin.

The experimental information considered consists of $N = 35$ samples of the time response of the real system y to a unit step input u while the time–varying parameter v_1 was allowed to vary between 0% and 80% of the maximum value of the zoom during the experiment. By repeatedly measuring the location of the centroid of the target in the absence of input, the experimental noise measurement was determined to be bounded by $\epsilon = 4$ pixels, i.e. $\mathcal{N} = \{\boldsymbol{\eta} \in \Re^N : |\eta_k| \leq \epsilon\}$. This experimental error is mainly due to fluctuating conditions such as ambient light.

The resulting LMI optimization problem was solved using Matlab's LMI toolbox, leading to the values $K = 0.0444$ and $p_1 = 0.9743$. The complete identified model has the following structure:

$$\begin{aligned} \mathbf{x}_{k+1} &= \begin{bmatrix} \mathsf{A}_p & 0 \\ 0 & \mathsf{A}_{np} \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} \mathsf{B}_{1p} & \mathsf{B}_{2p} \\ 0 & \mathsf{B}_{np} \end{bmatrix} \begin{bmatrix} r_k \\ u_k \end{bmatrix} \\ \begin{bmatrix} s_k \\ z_k \end{bmatrix} &= \begin{bmatrix} \mathsf{C}_{1p} & 0 \\ p_1 \mathsf{C}_{2p} & \mathsf{C}_{np} \end{bmatrix} \mathbf{x}_k + \\ &\quad \begin{bmatrix} \mathsf{D}_{11p} & \mathsf{D}_{12p} \\ p_1 \mathsf{D}_{21p} & p_1 \mathsf{D}_{22p} + \mathsf{D}_{np} \end{bmatrix} \begin{bmatrix} r_k \\ u_k \end{bmatrix}, \end{aligned} \tag{9}$$

Here $\{\mathsf{A}_{np}, \mathsf{B}_{np}, \mathsf{C}_{np}, \mathsf{D}_{np}\}$ are the state space matrices of G_{np}, the non–parametric component of the model G, and $\{\mathsf{A}_p, \mathsf{B}_p, \mathsf{C}_p, \mathsf{D}_p\}$ the state space matrices of G_1, the *a priori* parametric information that enters the term $\mathcal{F}_u(G_1, \Upsilon)$[4]. Finally, through a model (in)validation step, it was determined that the (multiplicative) model uncertainty associated with this description satisfies $\Delta_{mult} \in \mathcal{BH}_\infty(0.26)$.

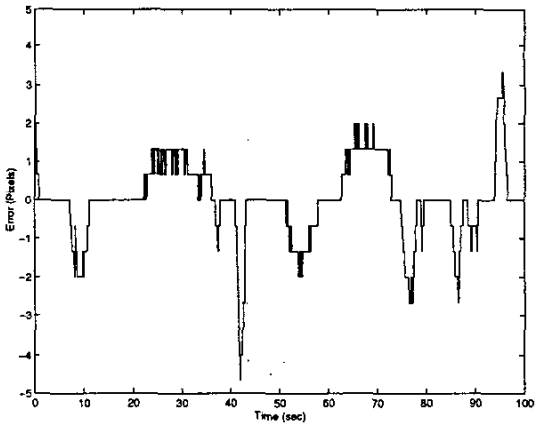


Fig. 4. Tracking error while zooming in and out.

In order to further validate the proposed approach, the plant description (9) and the uncertainty description $\Delta_{mult} \in \mathcal{BH}_\infty(0.26)$ were combined with the

[4]Numerical values for these matrices, as well as the controller, omitted for space reasons, can be obtained by contacting the authors.

technique used in [1] to design an LFT scheduled $\mathcal{H}_\infty$ controller. Figure 4 shows the results of experiments where a person is tracked while zooming in and out of his features. As illustrated there, the LPV controller was able to achieve good tracking performance in spite of the substantial change in the dynamics of the plant due to the change in f.

C. Activity Recognition:

Consider the problem of distinguishing between humans walking, running or walking up stairs, using as data time sequences of four joint angles (shoulder, elbow, hip and knee). A typical plot of these sequences is shown in Figure 5[5].

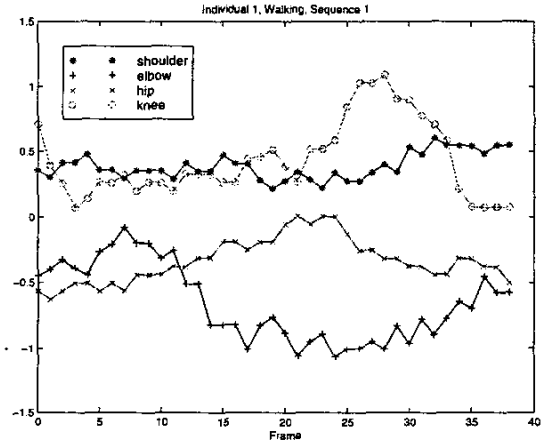


Fig. 5. Joint Angles of a walking individual.

In principle, activity recognition can be accomplished by assuming that these time–series are realizations of second–order stationary stochastic processes, and comparing the distances between the underlying models, identified for instance using subspace identification methods [3]. However, as pointed out in [8] this approach may be fragile, leading to activity miss–classification. To avoid this difficulty, we propose to addresses the problem of human gait recognition using model (in)validation motivated techniques.

1) Activity Recognition as a Model (In)Validation problem: In the sequel we will assume that a a sequence $\mathbf{y}_k^S$ of measurements of the angles of the shoulder, elbow, hip and knee joints of a person walking, running or walking a staircase, can be modelled as the response of an LTI system S to a signal $e \in \mathcal{B}\ell_2$ [6]:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathsf{A}\mathbf{x}_k + \mathsf{K}\mathbf{e}_k, \quad \hat{\mathbf{y}}_k^S = \mathsf{C}\mathbf{x}_k + \mathbf{e}_k \\ \hat{\mathbf{y}}_k^S &\doteq \mathbf{y}_k^S - \mathcal{E}(\mathbf{y}_k^S), \quad \mathcal{E}(\mathbf{y}_k^S) = \mu \ \forall k, \end{aligned} \tag{10}$$

where $\mathcal{E}$ denotes expected value. The triplet $(\mathsf{A}, \mathsf{K}, \mathsf{C})$ can be obtained from the experimental data by using subspace identification methods.

Consider now the problem of assigning an *unknown* sequence $\hat{y}$ to a model S from a set of candidate models $\{S_i\}$, each representative of a particular gait type. To this effect we will assume that *all* sequences $\mathbf{y}_i$ corresponding to the i^{th} class of activities can be generated as the output of the model S_i corrupted by structured dynamic uncertainty, that accounts for unmodelled dynamics and modelling errors. This amounts to establishing the existence of an admissible input $e_i \in \mathcal{U} = \{e\colon \sum_{k=1}^N \mathbf{e}^T\mathbf{e} \le 1\}$ and an ℓ_2 bounded operator Δ such that:

$$\hat{y} = [(I + \Delta)S] * e, \tag{11}$$

This is precisely a model (in)validation problem of the form:

Problem 2. *Given a nominal model for a given gait type S as in (10), sets $\mathcal{U}$ and $\mathbf{\Delta}$ of possible inputs and uncertainty blocks:*

$$\mathcal{U} = \{\mathbf{e}\colon \sum_{k=1}^{N} \mathbf{e}^T\mathbf{e} \le \epsilon^2\}, \quad \mathbf{\Delta} = \{\Delta\colon \|\Delta\|_\infty \le \delta < 1\}$$

and the experimental sequence $\hat{y}$, determine whether or not there exists at least one pair $(\mathbf{e}, \Delta) \in \mathcal{U} \times \mathbf{\Delta}$ so that (11) holds.

Once the process has been repeated for each representative model S_i, the unknown activity can be assigned to the class associated with the lowest uncertainty value.

2) Reducing the problem to an LMI optimization: Next we show that Problem 2 can be recast as an LMI feasibility problem, by invoking Carathéodory-Fejér interpolation theory.

Theorem 1. *Problem 2 has an affirmative answer if and only if there exists a finite sequence $e = \{\mathbf{e}_0, \mathbf{e}_1, \cdots, \mathbf{e}_n\}$ so that the following set of LMIs hold:*

$$\begin{aligned} \mathsf{A}_1(e) &\doteq \begin{bmatrix} \mathsf{X}(e) & (\mathsf{T}_S^n \mathsf{T}_e^n)^T \\ \mathsf{T}_S^n \mathsf{T}_e^n & -\alpha(\delta)\mathsf{I} \end{bmatrix} \le 0 \\ \mathsf{A}_2(e) &\doteq \begin{bmatrix} \epsilon^2 & \mathsf{Y}^T(e) \\ \mathsf{Y}(e) & \mathsf{I} \end{bmatrix} \ge 0 \end{aligned} \tag{12}$$

[5]This data has been provided by Professor Stefano Soatto, C.S. Dept., UCLA.

[6]This provides a deterministic, set membership approximation to a stochastic signal.

where

$$\alpha(\delta) = (1-\delta^2)^{-1}$$
$$\mathsf{X}(e) \doteq (\mathsf{T}^n_{\hat{y}})^T\mathsf{T}^n_{\hat{y}} - (\mathsf{T}^n_{\hat{y}})^T\mathsf{T}^n_S\mathsf{T}^n_e - (\mathsf{T}^n_S\mathsf{T}^n_e)^T\mathsf{T}^n_{\hat{y}}$$
$$\mathsf{Y}(e) \doteq \begin{bmatrix}\mathbf{e}_0^T & \mathbf{e}_1^T & \cdots & \mathbf{e}_n^T\end{bmatrix}^T,$$

and $\mathsf{T}^n_{\hat{y}}$, T^n_e and T^n_S are the Toeplitz matrices associated with the sequences $\hat{y}, e$ and the impulse response of S, respectively.

Proof: Feasibility of (11) is equivalent to the existence of a pair (e, Δ) such that $\Delta(S * e) = \hat{y} - S * e$. From Carathéodory-FejérTheorem [2] it can be easily shown that such a Δ exists if and only if

$$(\mathsf{T}^n_{\hat{y}})^T\mathsf{T}^n_{\hat{y}} - (\mathsf{T}^n_{\hat{y}})^T\mathsf{T}^n_S\mathsf{T}^n_e - (\mathsf{T}^n_S\mathsf{T}^n_e)^T\mathsf{T}^n_{\hat{y}} - (\delta^2-1)(\mathsf{T}^n_S\mathsf{T}^n_e)^T\mathsf{T}^n_S\mathsf{T}^n_e \leq 0.$$

The first LMI in (12) follows now using Schur complements. The second LMI is a simple restatement of $\sum_{k=1}^N \mathbf{e}^T\mathbf{e} \leq 1$.

Note that since $\alpha(\delta) \doteq (1-\delta^2)^{-1}$ for $\delta \in (0,1)$ is a convex function of δ, it is possible to optimize over the size of the uncertainty required to explain the data, by solving the following problem:

$$\min \alpha \quad \text{st: } \mathsf{A}_1(e,\ \alpha) \leq 0,\ \mathsf{A}_2(e) \geq 0,\ \alpha > 1$$

3) Example: Human Gait Recognition:

The experimental data consists of 30 vector sequences, taken from 5 different persons, labelled A, B, C, D and E, walking (sequences 1–10), running (11–20) or walking a staircase (21–30).

The models consist of a 4 input/4 output system S_i of the form (10), and its associated sequence y_i, mean $\mu_i = \mathcal{E}(\mathbf{y}_{i_k})$ and an upper bound on the input energy ϵ_i, computed as the input energy required for model S_i to generate y_i, i.e. $\epsilon_i \doteq \|e\|_{\ell_2[0,N]n} : e = S_i^{-1} * y_i$. Given a gait type and a set of models $\mathcal{S}$, define the nominal model $S \in \mathcal{S}$ as the one that is closest to each other element in its class, in the sense of minimizing the norm of the (multiplicative) uncertainty required to map the two models under consideration, i.e

$$S = \arg \min_{\hat{S}_i, \hat{S}_j \in \mathcal{S}} \left\{ \|(\hat{S}_i - \hat{S}_j)\hat{S}_j^{-1}\|_\infty \right\}, \qquad (13)$$

where $\hat{S}_i \doteq \epsilon_i S_i$ [7]. Proceeding as described above yields the following three nominal models, denoted as S_{walk} for walking, S_{run} for running and S_{stair} for walking a staircase:

$$S_{walk} \doteq S_{10}, \quad S_{run} \doteq S_{20}, \quad S_{stair} \doteq S_{30}. \quad (14)$$

[7]The scaling is required to make models comparable in the context of Problem 2.

Sequence	S_{walk}	S_{run}	S_{stair}
y_1–y_9	0.0–0.2†	0.6–0.7	0.24–0.9
y_{11}–y_{19}	0.9–1	0–0.35†	0.5-1.0
y_{21}–y_{25}	0.7–0.96	0.41–0.6	0.05–0.39†
y_{26}	0.6828†	0.7127	0.8827
y_{27}	0.5553	0.5818	0.4682†
y_{28}	0.2650	0.6801	0.1699†
y_{29}	0.0391†	0.6102	0.1470

Fig. 6. Gait Recognition Results

Thus, sequences $\{y_{10}, y_{20}, y_{30}\}$ are the training data for the problem.

The results. Table 6 shows the results of applying Theorem 1, using 20 sample points per sequence to the remaining sequences. In all cases, the first column contains the experimental sequences to be recognized; the second, third and fourth columns display the minimum value of $\|\Delta\|_\infty$, the uncertainty block required for the nominal models S_{walk}, S_{run} and S_{stair} to reproduce the given data. A given unknown sequence is assigned to the activity type corresponding to the smallest $\|\Delta\|_\infty$ (indicated by a †). As shown there, the proposed method can successfully recognize 25 out of the 27 sequences under consideration; it only miss–classifies sequences y_{26} and y_{29}, –walking up a staircase– as walking sequences. The failure could be attributed to the length of the experiment used for recognition purposes, or simply to faulty sequences, specially because the proposed method is able to correctly recognize sequences $\{y_{25}, y_{27}\}$ and y_{28} from A and C respectively.

III. Conclusions

In the past few years active 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 many cases of practical interest this fragility can be addressed by using interpolation and LMI tools to recast these problems into a tractable optimization form. Additional examples and clips showing the effectiveness of this approach can be found at `http://robustsystems.ee.psu.edu`.

It is also worth mentioning that there are important cases where the techniques developed in this paper do not provide a complete solution, since the resulting problem is not convex in all the variables in-

volved. An example is gait recognition when the experimental data is corrupted by measurement noise. Research is currently underway seeking to overcome this difficulty by combining the approach pursued in this paper with risk–adjusted (in)validation methods [9].

ACKNOWLEDGEMENTS

Support from NSF under grants ECS–9907051 and IIS–0117387, and AFOSR under grant F49620–00–1–0020 is gratefully acknowledged.

REFERENCES

[1] P. Apkarian and P. Gahinet. A convex characterization of gain–scheduled $\mathcal{H}_\infty$ controllers. *IEEE Trans. Aut. Control*, 40(5):853–864, 1995.

[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 CVPR*, Kauai, Hawaii, USA, December 2001.

[4] A. Blake and M. Isard. Condensation - condensation density propagation for visual tracking. *Int. J. Comp. Vision*, 29(1):5–28, 1998.

[5] 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.

[6] J. Feddema. Microassembly of micro–electromechanical systems (MEMS) using visual servoing. In *Block Island Workshop on Vision and Control*, 1997.

[7] D. J. Fleet, M. J. Black, Y. Yacoob, and A. D. Jepson. Design and use of linear models for image motion analyis. *Int. J. of Comp. Vision*, 36(3):171–194, February/March 2000.

[8] C. Mazzaro, M. Sznaier, O. Camps, S. Soatto, and A Bissacco. A model (in)validation approach to gait recognition. In *1rst International Symposium on 3D Data Processing, Visualization and Transmission*, pages 700–703, Padova, Italy, 2002.

[9] M.C. Mazzaro, M. Sznaier, and C. Lagoa. A risk–adjusted approach to model (in)validation. In *2003 ACC*, Denver, Co, USA, June 2003.

[10] P.A. Parrilo, R. Sánchez Peña, and M. Sznaier. A parametric extension of mixed time/frequency robust identification. *IEEE Trans. Aut. Control*, 44(2):364–369, February 1999.

[11] R. Sánchez Peña and M. Sznaier. *Robust Systems Theory and Applications*. Wiley & Sons, Inc., 1998.

[12] C. E. Smith, C. A. Richards, S. A. Brandt, and N. P. Papanikolopoulos. Visual tracking for intelligent vehicle–highway systems. *IEEE Trans. Vehicular Tech.*, 45(4):744–759, 1996.

[13] M. Sznaier and C. Mazzaro. An LMI approach to control oriented identification and model (in)validation of LPV systems. *IEEE Transactions on Automatic Control*, 48(9), 2003.

[14] M. Sznaier, B. Murphy, and O. Camps. An LPV approach to synthesizing robust active vision systems. In *2000 IEEE Conf. Dec. Control*, pages 2545–2550, December 2000.

[15] J. K. Tsotsos, G. Verghese, S. Dickinson, M. Jenkin, A. Jepson, E. Milios, F. Nuflo, S. Stevenson, M. Black, D. Metaxas, S. Culhane, Y. Ye, and R. Mannn. PLAYBOT: A visually-guided robot for physically disabled children. *Image and Vision Computing*, 16(4): 275–292, April 1998.