The Role of Control in Computer Vision and Image Processing

Octavia Camps* **Mario Sznaier*** **(Trans.: Hideaki Ishii)**

*Electrical and Computer Engineering, Northeastern University, Boston, MA 02115
*E-mail: {camps,msznaier}@ece.neu.edu

Keywords: Multiframe Tracking, Dynamic Motion Segmentation, Texture Analysis, Texture Inpaiting, Robust Identification, Model (In)Validation, Caratheodory Fejer Interpolation

概 要

Dynamic vision and imaging systems can substantially improve our quality of life. However, key issues that must be addressed in order to deploy these systems are their potential fragility and the need to process vast amounts of information in real time. As we show in this paper, these issues can be addressed by appealing to a common systems theoretic substrate that allows for recasting a wide range of problems into a tractable convex optimization form. These ideas are illustrated with several applications including multiframe tracking, motion segmentation, texture analysis/synthesis and video reconstruction and inpainting.

1. Introduction

Dynamic vision and imaging – the confluence of dynamics, computer vision, image processing and control – is uniquely positioned to enhance the quality of life for large segments of the general public. Aware sensors endowed with tracking and scene analysis capabilities can prevent crime and reduce time response to emergency scenes. Enhanced imaging methods can substantially reduce the amount of radiation required in medical procedures. Moreover, the investment required to accomplish these goals is relatively modest, since a large number of imaging sensors are already deployed and networked. The challenge now is to develop a theoretical framework that allows for *robustly* processing this vast amount of information, within the constraints imposed by the need for real time operation in dynamic, partially stochastic scenarios. The goal of this paper is to illustrate the central role that dynamic models and their associated predictions can play in developing a comprehensive, computationally tractable robust dynamic vision and imaging framework. Establishing a connection with a rich set of robust systems theory tools allows for recasting a wide spectrum of problems arising in this context – robustly tracking an object in a sequence of frames, modeling appearance changes, recovering structure from motion, and classifying textured images – into a tractable, finite dimensional convex optimization.

†This work was supported in part by NSF under grants ECS–9907051 and IIS–0117387, and AFOSR under grant F49620–00–1–0020.

2. Notation

$\mathcal{H}_{\infty,\rho}$ space of functions analytic in $|z| \leq \rho$, equipped with the norm $\|G\|_{\infty,\rho} \doteq \sup_{|z|<\rho} \overline{\sigma}(G(z))$, where $\overline{\sigma}(.)$ denotes maximum singular value.

$\mathcal{BH}_\infty(K)$ open K–ball in $\mathcal{H}_\infty$

3. Interpolation Problems in Dynamic Vision

In this section we show that many dynamic vision problems such as can be reduced to a convex optimization problem, through the use of well established system–theoretic tools.

3.1 Multiframe Tracking

A requirement common to most dynamic vision applications is the *ability to track* objects in a sequence of frames. Current ap-

proaches integrate correspondences between individual frames over time, through a combination of some assumed simple target dynamics (e.g. constant velocity), empirically learned noise distributions and past position observations [5, 9]. However, while successful in many scenarios, these approaches still remain vulnerable to model uncertainty, occlusion and appearance changes, as illustrated in Figure 1.

As shown next, this difficulty can be solved by modeling the motion of the target as the output of a dynamical system, to be identified from the available data. To this effect, start by modeling y_k, the position of a given target feature as:

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

where $\mathbf{e}$ and $\eta_k \in \mathcal{N}$ represent a suitable input and measurement noise, respectively. Further, we will assume that the following *a priori* information is available:

(a) Set membership descriptions $\eta_k \in \mathcal{N}$ and $e_k \in \mathcal{E}$. These can be used to provide deterministic models of the stochastic signals e, η.

(b) $\mathcal{F}$ admits an expansion of the form $\mathcal{F} = \overbrace{\sum_{j=1}^{N_p} p_j \mathcal{F}^j}^{\mathcal{F}_p} + \mathcal{F}_{np}$. Here $\mathcal{F}^j$ are known, given, not necessarily stable operators that contain all the information available about possible modes of motion of the target.

(c) $\mathcal{F}_{np} \in \mathcal{BH}_{\infty,\rho}(K)$ for some known $\rho \leq 1$, e.g. a bound on the divergence rate of the approximation error of the expansion $\mathcal{F}_p$ to $\mathcal{F}$ 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 values. In turn, identifying the dynamics entails finding an operator $\mathcal{F}(z) \in \mathcal{S} \doteq \{\mathcal{F}(z) \colon \mathcal{F} = \mathcal{F}_p + \mathcal{F}_{np}\}$ such that $y - \eta = \mathcal{F}e$, precisely the class of interpolation problem addressed in [10]. As shown there, such an operator exists if and only if the following set of equations in $\mathbf{p}, \mathbf{h}$ and K is feasible:

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

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

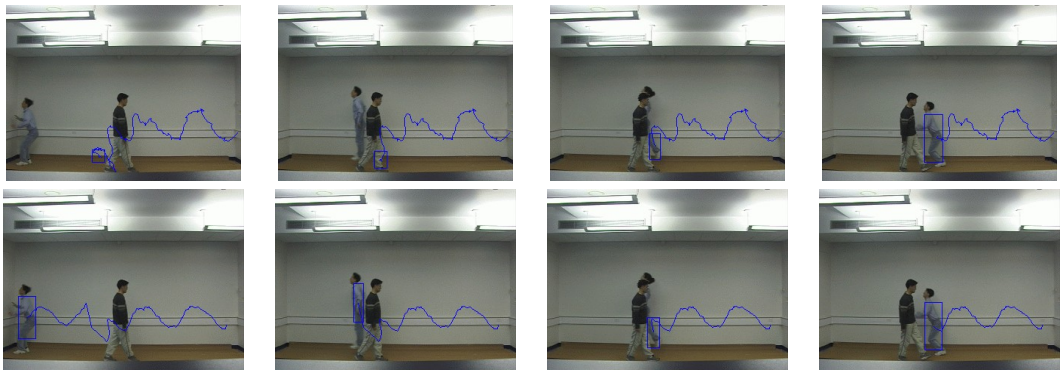
where T_x denotes the Toeplitz matrix associated with a sequence $\mathbf{x} = [x_1, \ldots, x_n]$, $\mathsf{R}_\rho \doteq \mathrm{diag}\,[1\ \rho\ \cdots\ \rho^n]$, $\mathsf{P} \doteq [f^1\ f^2\ \cdots\ f^{N_p}]$,where f^i is a vector containing the first n Markov parameters of the transfer function $\mathcal{F}^i(z)$ and $\mathbf{h}$ contains the first n Markov parameters of $\mathcal{F}_{np}(z)$

A Simple Tracking Example: Consider again the problem illustrated in Figure 1. The experimental information consists of centroid position measurements from the first 20 frames, where the target is not occluded. The *a priori* information, estimated from the non–occluded portion of the trajectory is:

1. 5% noise level
2. $\mathcal{E} = \delta(0)$, i.e. motion of the target was modelled as the impulse response of the unknown operator $F^\dagger$.
3. $\mathcal{F}_p \in \mathrm{span}[\frac{1}{z-1},\ \frac{z}{z-a},\ \frac{z}{(z-1)^2},\ \frac{z^2}{(z-1)^2},\ \frac{z^2-\cos\omega z}{z^2-2\cos\omega z+1},\ \frac{\sin\omega z^2}{z^2-2\cos\omega z+1}]$ where $a \in \{0.9, 1, 1.2, 1.3, 2\}$ and $\omega \in \{0.2, 0.45\}$
4. $F_{np} \in \mathcal{BH}_{\infty,\rho}(K)$, with $\rho = 0.99$

As shown in Figure 1, a Kalman filter tracker that uses the identified dynamics is now able to track the target past the occlusion. It is worth emphasizing that this combination significantly outperforms a tracker based solely on an unscented particle filter [5]. Hence, exploiting dynamical information through the use of control–motivated tools, leads to *both* robustness improvement and substantial computational complexity reduction. In addition, as discussed in [1], the framework described above furnishes *deterministic, worst–case bounds* on the prediction error that can be used to disambiguate among targets with neighboring tracks.

[†]This is equivalent to lumping together the dynamics of the plant and the input signal.



Frame 150 Frame 105 Frame 95 Frame 85

Figure 1 Tracking in the presence of occlusion. Top: Unscented Particle Filter based tracker. Bottom: Combination Identified Dynamics/Kalman Filter.

3.2 Dynamic Appearance Modeling.

Arguably, one of the hardest challenges in tracking is to overcome changes to its appearance. In principle, this difficulty can be solved by using *dynamic* appearance models obtained using the same robust identification approaches employed to identify the motion dynamics [6]. However, moving beyond a few simple descriptors requires addressing the issues of high computational costs, due to the poor scaling properties of LMI based identification algorithms.

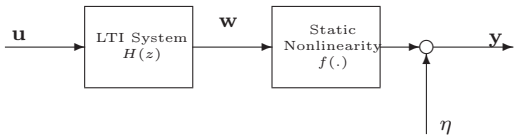


Figure 2 Wiener System Structure

This challenge can be addressed through the use of nonlinear dimensionality reduction techniques to map the data to a lower dimensional manifold where the identification/tracking is performed. Since the projection onto the lower dimensional manifold can be modeled as a static nonlinearity, this approach leads naturally to a Wiener system structure of the form illustrated in Figure 2, consisting of the interconnection of a LTI system $H(z)$ and a memoryless nonlinearity $f(.)$. Next, we illustrate the effectiveness of this approach using the problem of human motion modeling and tracking. The experimental data, partially shown in Figure 3(a) consists of the first 20 frames of a human walking, each having 1728 pixels. Thus, modeling pixel evolution become infeasible even when using just a few frames. On the other hand using the risk–adjusted approach proposed in [7] and the following *a priori* information

1.- $\omega \in R^3$ (since it represents the coordinates of the centroid of the target).

2.- The static nonlinearity $\mathbf{f}(.)$ has the form†: $\mathbf{f}(\mathbf{x}) = \mathbf{B}\mathbf{\Psi}(\mathbf{x})$ where $\mathbf{B} \in R^{1726\times 6}$ is an unknown matrix and the bases $\mathbf{\Psi}(\mathbf{x})$: $R^3 \rightarrow R^6$ are given by:

$$\mathbf{\Psi}(\mathbf{x}) = [\exp(-0.8\|\mathbf{x}-\mathbf{t_1}\|_2^2), \exp(-0.8\|\mathbf{x}-\mathbf{t_2}\|_2^2), 1, \mathbf{x}^T]^T$$

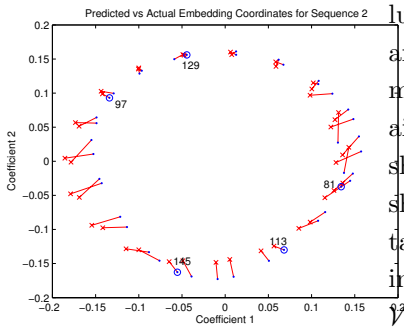
where

$$\mathbf{t_1} = \begin{bmatrix} 0.6833 & -0.4521 & -0.0033 \end{bmatrix}$$
$$\mathbf{t_2} = \begin{bmatrix} -0.7552 & 0.4997 & 0.0036 \end{bmatrix}$$

led to model with a fifth order linear portion that interpolates the data within 10%. In addition, Figure 3(b) shows close agreement between the temporal evolution of the points on the manifold and the positions predicted using the linear dynamic model. This substantiates the conjecture posed in [8], that human motion tracking can be decoupled into: (a) a linear tracking problem in a low dimensional manifold, accounting for the *dynamics* of the motion, and (b) a nonlinear, static mapping that accounts for the changes in appearance of the target.

†This hypothesis is motivated by the bases proposed in [3] to model human silhouettes.

(a)



(b)

Figure 3 Learning appearance using a Wiener system. (a) Top: Walking sequence (from CMU MoBo database), Bottom: impulse response of the identified Wiener system. (b) Evolution on a 2D projection of the 3D manifold: predicted (cross) and actual (dot).

3.3 Structure Recovery from Dynamics:

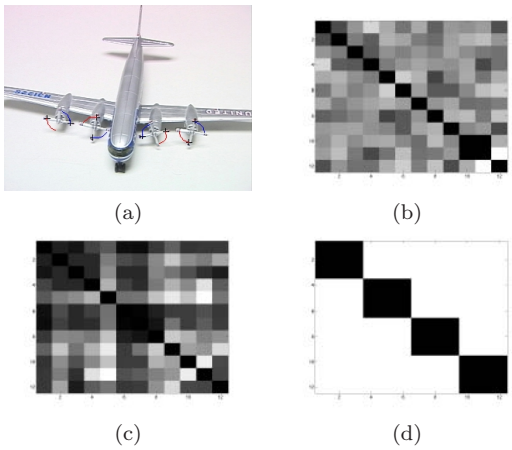
When tracking an unknown number N_o of moving objects, it is of interest to identify (i) the number of objects, (ii) the individual dynamics and, (iii) assign points in the image to each. To illustrate the issues involved, start by considering P features from a single rigid object, tracked over F frames with image coordinates $\{(u_t^p, v_t^p)\}$, $p = 1, \ldots, P$, $t = 1, \ldots, F$. Define the measurement matrix $W_{1:F}$, by:

$$\mathcal{W}_{1:F} = \begin{bmatrix} u_t^p - u_t \\ v_t^p - v_t \end{bmatrix} \in R^{2P \times F} \qquad (4)$$

where (u_t, v_t) denote coordinates of the centroid of the features. Under the assumptions of affine projection it can be shown [14] that $\mathcal{W}_{1:F}$ has at most rank 3 and can be decomposed into a rotation $R_{1:F}$ and a "structure" matrix S

$$\mathcal{W}_{1:F} = \begin{bmatrix} R_{1:F}^u \\ R_{1:F}^v \end{bmatrix} S = R_{1:F} S \qquad (5)$$

In the case of multiple objects, the number of objects and the corresponding geometry can be obtained by factoring $\mathcal{W}$ into rank 3 submatrices. This basic idea lies at the core of factorization based approaches (see for instance [16, 15]), leading to computationally efficient solutions. However, these approaches cannot disambiguate objects that partially share motion modes, such as the same–wing propellers of the airplane shown in Figure 4(a). It can be easily shown that in this case rank$(\mathcal{W}) = 6$. Hence, as shown in Figure 4 (b)–(c), any motion segmentation approach based solely on finding linearly independent subspaces of the column space of $\mathcal{W}$ will fail, since it cannot distinguish this case from the case of two independently moving propellers. Intuitively, the main difficulty here is that any approach based on properties of $\mathcal{W}$ that are invariant under column permutations, *take into account only geometrical constraints, but not dynamical ones.*



(a) (b) (c) (d)

Figure 4 (a) Propeller tracks. (b) Costeira-Kanade motion segmentation. (c) Zelnik-Manor-Irani motion segmentation using six eigenvectors. (d) Dynamics based segmentation.

As we show next, robustness can be substantially improved by exploiting the fact that points on the same rigid share more modes of motion than points on different objects. Specifically, begin by associating to the j^{th} object, its centroid $\mathbf{O}^{(j)}$ and an affine basis $b^{(j)}$, centered at $\mathbf{O}^{(j)}$, defined by three no coplanar vectors $\mathbf{V}_i^{(j)}$. Finally, denote by $o^{(j)}(k)$, $v_i^{(j)}(k)$ the coordinates of the image of $\mathbf{O}^{(j)}(k)$ and the projections of $\mathbf{V}_i^{(j)}(k)$ onto the image plane, respectively. Given any point $\mathbf{P}_i^{(j)}$ belonging to the j^{th} object, the coordinates at time k of its image $\mathbf{p}^{(i)}(k)$ are given by:

$$\mathbf{p}_i^{(j)}(k) = \mathbf{o}^{(j)}(k) + \alpha_i^{(j)} \mathbf{v}_1^{(j)}(k) + \beta_i^{(j)} \mathbf{v}_2^{(j)} k + \gamma_i^{(j)} \mathbf{v}_3^{(j)}(k)$$

where $\alpha_i^{(j)}, \beta_i^{(j)}$ and $\gamma_i^{(j)}$ are the *affine invariant* coordinates of $\mathbf{P}_i^{(j)}$ with respect to the basis $b^{(j)}$. Note that, for any two points $\mathbf{P}_r^{(j)}, \mathbf{P}_s^{(j)}$ in the same object, the dynamics of $\mathbf{o}^{(j)}$ are *unobservable* from $\delta_{r,s}(k) \doteq \mathbf{p}_r^{(j)}(k) - \mathbf{p}_s^{(j)}(k)$. Thus, the underlying subsystem is rank deficient when compared to a subsystem describing difference between points on different objects. Roughly speaking, the relative motion of points in a given object, carries no information about the motion of other objects. It follows that points can be clustered in objects according to the complexity of the model required to explain their relative motion. In turn, the order of this model can be estimated by simply computing the rank of the Hankel matrix constructed from the pairwise differences $\delta_{rs}(k)$, leading to a simple segmentation algorithm, computationally no more expensive than a sequence of SVDs. The effectiveness of this approach is illustrated in Figure 4(d), showing that it correctly identified the presence of four independently moving objects.

4. Textured Image Processing

Texture has been the subject of research in image processing for over three decades, with applications ranging from medical diagnosis to entertainment to human computer interfaces. During the past few years, significant advances have been made in addressing multiple aspects of the problem, ranging from inpainting and synthesis to classification. However, at present, each sub-problem is addressed using a specific set of tailored tools [4]. Next we briefly illustrate how the use of system theoretic tools can lead to a unified framework capable of exploiting the synergism between different aspects of the problem to improve robustness and reduce the computational burden.

4.1 Texture Modeling and Synthesis

Compact models of textured images can be obtained by treating the intensity values $\mathcal{I}(k,l)$ at the (k,l) pixel of the image as the as the output of a *two-dimensional*, discrete linear shift-invariant system driven by white noise, reducing the problem to an identification one: obtaining a model G from image data, possibly corrupted by noise. Note that this requires considering two–dimensional, *non–causal* systems, since the intensity value at a pixel is likely to depend on the values of all pixels in its neighborhood, not just on those preceding it in some ordering of the image pixels. This difficulty can be circumvented by considering a given $n \times m$ image as one period of an infinite 2D signal with period (n,m). Thus, at any given location (i,j) in the image, the intensity values $\mathcal{I}(r,s)$ at other pixels are available also at position $(r-qn, s-qm)$, and the integer q can always be chosen so that $r-qn < i, s-qm < j$. From this observation, it follows that the unknown system G admits a state space representation of form:

$$\begin{aligned} x'(i,j) &= Ax(i,j) + Bu(i,j) \\ \mathcal{I}(i,j) &= Cx(i,j) + Du(i,j) \end{aligned} \tag{6}$$

where

$$x'(i,j) = \begin{bmatrix} x^v(i+1,j) \\ x^h(i,j+1) \end{bmatrix}, x(i,j) = \begin{bmatrix} x^v(i,j) \\ x^h(i,j) \end{bmatrix}$$

$$A = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix}, B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

subject to an additional constraint of the form

$$\begin{matrix} g(i+N,j) = g(i,j) \\ g(i,j+M) = g(i,j) \end{matrix} \quad \text{for some finite } N, M > 0$$

where $g(.,.)$ denotes the impulse response of G. With these assumptions, the problem becomes one of identifying a state–space realization from experimental data, subject to a periodicity constraint, precisely the type of problems solved in [2]. The potential of this approach is illustrated in Fig. 5, where it was used to expand partial images by first identifying the underlying model and then simply computing its impulse response.

4.2 Texture Classification

In this section we show how the models obtained above can be used for texture classification. Proceeding as in [13], we will recast

Figure 5 Using 2-D Models to Expand Images

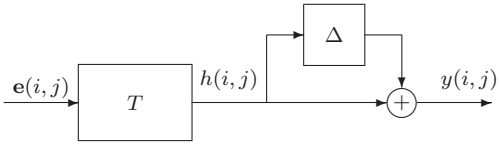
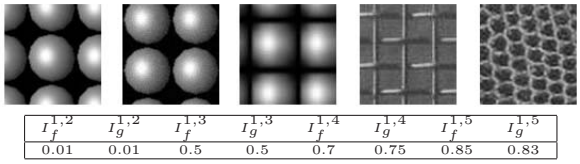


Figure 6 Texture Recognition Set-up

the problem into a robust semi-blind model (in)validation form. To this effect, we will postulate that all images corresponding to realizations of a given texture $\mathcal{T}$ can be obtained as the output of a 2-D operator S to an unknown input signal e with unit spectral density, applied in $(-\infty, 0] \times (-\infty, 0]$. This leads to the set-up shown in Figure 6, where $T(z_1, z_2)$ represents a nominal model of a particular texture, $h(i, j)$ and $y(i, j)$ denote the intensity value of the ideal and actual images, respectively, and where the (unknown) operator $\Delta(z_1, z_2)$ describes the mismatch between these two images.

In this context, given a set of texture families, each represented by a model T_i, an unknown specimen can be classified by (i) performing a sequence of invalidation models to find the lowest uncertainty value $\|\Delta_i\|$ required to explain the specimen in terms of the model T_i, and (ii) assigning the unknown texture to the family corresponding to smallest uncertainty norm. By identifying first a (separable) model of the nominal texture, the corresponding 2-D model invalidation problem can be reduced to two decoupled 1-D semi–blind validation problems that can be solved using the LMI–based technique developed in [13].

Figure 7 shows the results of applying the technique outlined above to classify several images. Here $I_f^{1,j}$ and $I_g^{1,j}$ denote the results obtained when comparing the decompositions corresponding to the first image against the models obtained from the j^{th} texture. As shown there, this approach correctly indicates that the first three images belong to the same family†.



$I_f^{1,2}$	$I_g^{1,2}$	$I_f^{1,3}$	$I_g^{1,3}$	$I_f^{1,4}$	$I_g^{1,4}$	$I_f^{1,5}$	$I_g^{1,5}$
0.01	0.01	0.5	0.5	0.7	0.75	0.85	0.83

Figure 7 Top: Sample Textures. Bottom: Optimal γ

4.3 Video Inpainting as a Rank Minimization Problem

Video inpainting, that is the process of seamlessly restoring or altering portions of a video clip, has been the subject of considerable attention in the past few years (see for instance [11] and references therein), but the problem is far from solved. Existing algorithms are limited in the types of sequences that can handle and have relatively high computational complexity. Next, we briefly outline how the use of system theoretic ideas can lead to simple, computationally efficient algorithms that exploit (global) spatio–temporal information. The main idea is to (i) find a set of descriptors that encapsulate the information necessary to reconstruct a frame, (ii) find an optimal estimate of the value of these descriptors for the missing/corrupted frames, and (iii) use the estimated values to reconstruct the frames. In turn, the optimal descriptor estimates can be efficiently obtained postulating that the correct values of the missing descriptors are such that the resulting inpainted sequence is described by the simplest possible (eg. lowest order) dynamical model†. Since the order of the underlying model can be estimated from the Hankel matrix of the data, this idea leads to a rank minimization problem, which in turn can be relaxed to an LMI optimization, resulting in

†The higher values of $I_f^{1,3}$ and $I_v^{1,3}$ are due to the use of a lower quality image for the third texture.

†It can be analytically shown that this is indeed the case for periodic sequences, but empirical results show that this hypothesis works well also for non–periodic textures.

the following algorithm:

1.- Given the observed values of the descriptors f^o, form the following (Hankel) matrix:

$$H_f \doteq \begin{bmatrix} f_1 & f_2 & \cdots & f_{n/2} \\ f_2 & f_3 & \cdots & f_{n/2+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n/2} & f_{n/2+1} & \cdots & f_{n-1} \end{bmatrix} \quad (7)$$

Here f denotes either the observed data f^o_k, if the k frame is present, or the unknown value f^m_k, if the frame needs to be inpainted, and n denotes the total number of frames.

2.- Estimate the values f^m which are maximally consistent with f^o by solving the following Linear Matrix Inequality (LMI) optimization problem,

$$\begin{array}{lll} \text{minimize w.r.t} & f^m & Tr(Y)+Tr(Z) \\ \text{subject to} & & \begin{bmatrix} Y & H_f \\ (H_f)^T & Z \end{bmatrix} \geq 0 \end{array}$$

where $Y^T = Y \in \mathcal{R}^{n\times n}$, $Z^T = Z \in \mathcal{R}^{n\times n}$ and $H_f \in \mathcal{R}^{n\times n}$.

The potential of this approach is illustrated in Fig. 8, where it was used to restore the occluded person. In this particular example, the positions $f_k = (x^i_k, y^i_k)$ of the 6 feature points indicated in the figure were chosen as descriptors. The video has 36 frames, and occlusion occurs in frames 17 through 19. Using the algorithm outlined above implemented in MATLAB to inpaint the missing descriptors required approximately 20 seconds on a P-III 1.2G PC.

5. Conclusions

Dynamic vision and imaging is arguably one of the few areas where both further advances and widespread field deployment are being held up not by the lack of a supporting infrastructure, but the lack of *supporting theory*. In this paper paper we illustrated the central role that systems theory can play in developing a comprehensive framework leading to provably robust dynamic vision and imaging systems. In turn, these fields can provide a rich environment both draw inspiration from and to test new developments in systems theory. For instance, the applications addressed in this paper point out, among others, to the need for further research into low complexity nonlinear identification methods, the development of worst-case identification methods for switched systems that are not necessarily stable (to allow for parsing video sequences into different activities), and to extend currently available 1-D identification methods to the 2-D case.

Acknowledgements

The authors are indebted to Drs B. Franzen, T. Ding, H. Lim, C. Mazzaro and W. Ma, and to V. Morariu and R. Lublinerman for their contributions to the research reported here.

References

1) O. Camps and M. Sznaier. System Theoretic Methods in Computer Vision and Image Processing. In *2007 European Control Conference*, Kos, Greece.

2) T. Ding, M. Sznaier, and O. Camps. Robust identification of 2-d periodic systems with applications to texture synthesis and classification. In *45th CDC*, San Diego, CA, 2006. pp. 3678-3683.

3) Ahmed Elgammal and Chan-Su Lee. Inferring 3d body pose from silhouettes using activity manifold learning. In *Proc. 2004 CVPR*, pages 681–688.

4) D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Prentice Hall, 2003.

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

6) Hwasup Lim, Octavia I. Camps, and Mario Sznaier. A caratheodory-fejer approach to appearance modelling. In *Proc. 2005 CVPR*, pages 301–307.

7) W. Ma, H. Lim, M. Sznaier, and O. Camps. Risk adjusted idenitification of wiener sys-

Figure 8 Top: original sequence. Middle: observed and estimated descriptors. Bottom: inpainted sequence.

tems. In *45th CDC*, San Diego, CA, 2006. pp. 2512-2515.

8) V. Morariu and O. Camps. Modelling correspondences for multi camera tracking using nonlinear mani- fold learning and target dynamics. In *Proc. 2006 CVPR,* pages 545–552.

9) 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.

10) P. A. Parrilo, R. S. Sanchez Pena, and M. Sznaier. A parametric extension of mixed time/frequency domain based robust identification. *IEEE Trans. Autom. Contr.*, 44(2):364–369, 1999.

11) K. A. Patwardhan, G. Sapiro, and M. Bertalmio. Video inpainting of occluding and occluded objects. In *Proc. of ICIP 2005*, volume 2, pages 69–72. IEEE, 2005.

12) R. Sánchez Peña and M. Sznaier. *Robust Systems Theory and Applications.* Wiley & Sons, Inc., 1998.

13) M. Sznaier, M. C. Mazzaro, and O. Camps. Semi-blind model (in)validation with applications to texture classification. In *44th CDC–ECC'05*, pages 6065–6070, Seville, Spain, 2005.

14) C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

15) Jing Xiao, Jinxiang Chai, and Takeo Kanade. A closed-form solution to non-rigid shape and motion recovery. In *The 8th European Conference on Computer Vision (ECCV 2004)*, May 2004.

16) L. Zelnik-Manor and M. Irani. Degeneracies, dependencies and their implications in multi-body and multi-sequence factorization. In *Proc. 2003 CVPR*, pages 287–293.