Dynamic Appearance Modeling for Human Tracking*

Hwasup Lim[1], Vlad I. Morariu[3], Octavia I. Camps[1,2], and Mario Sznaier[1]
[1] Dept. of Electrical Engineering
[2] Dept. of Computer Science and Engineering
The Pennsylvania State University
University Park, PA 16802
[3] Dept. of Computer Science
University of Maryland
College Park, MD 20742

Abstract

Dynamic appearance is one of the most important cues for tracking and identifying moving people. However, direct modeling spatio-temporal variations of such appearance is often a difficult problem due to their high dimensionality and nonlinearities. In this paper we present a human tracking system that uses a dynamic appearance and motion modeling framework based on the use of robust system dynamics identification and nonlinear dimensionality reduction techniques. The proposed system learns dynamic appearance and motion models from a small set of initial frames and does not require prior knowledge such as gender or type of activity.

The advantages of the proposed tracking system are illustrated with several examples where the learned dynamics accurately predict the location and appearance of the targets in future frames, preventing tracking failures due to model drifting, target occlusion and scene clutter.

1. Introduction

Tracking and identifying moving humans in video sequences is a challenging task because the visual appearance of the targets changes dynamically due to their articulated body structure, and due to viewpoint and illumination changes. The performance of tracking algorithms often depends on how well they can estimate the visual information and efficiently handle its temporal variation. Several approaches have been introduced in the literature to address these issues.

Color information is widely used for tracking target objects. Perez et el. [16] proposed probabilistic tracking based on color histograms over time. Lim et el. [13] modeled dynamic variations of color histograms and predicted future color information using system identification techniques. Different objects with similar colors, however, may introduce ambiguities because local color and shape information are not taken into consideration by these approaches.

Black and Jepson [3] learned a set of view-based representation of the target objects using eigenspaces. Hager and Belhumeur [8] tracked the target object by estimating affine transformations of target templates. Neither of these approaches incorporate temporal variations of the appearance.

More recently, adaptive models incorporating both spatial and temporal variations of the appearance have been proposed. While these methods perform better than the previous approaches, in general, they lack accurate generative models. Jepson et el. [12] developed the WSL tracker, in which the appearance variation is divided into stable, lost, and wandering components. Ho et el. [10] used linear subspaces obtained by Gramm-Schmidt process and updated the orthogonal basis over time to handle temporal variations of local appearance information. Han and Davis [9] utilized a mean shift based algorithm for sequential density estimation of each pixel. All of these approaches search and validate the target appearance based on previous observations. They, however, may drift by incorporating the background into the target template and can fail to track in the presence of long term occlusion.

Figure 1. Generated walking sequence after the learning step, $t = 38, 40, \ldots, 48$. Top: from constant acceleration dynamics; Bottom: from dynamics identified using robust estimation

The work presented here is closely related to approaches

*This work was supported in part by AFOSR under grant FA9550. 05.1.0437 and NSF, under grants IIS.0117387, ECS.0221562 and ITR. 0312558.

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)

Figure 2. Dynamic appearance and occlusion by a background object, $t = 60, 75, 90,$ and 105. The white box is the predicted appearance and the black box is the probability map based on the predicted appearance.

that attempt to model visual appearance in low dimensional spaces. In the context of human appearance modeling, Lim and Kriegman [14] and Dimitrijevic et el. [5] used a set of prior learned templates which do not characterize individual differences. Elgammal and Lee [7] learned human appearance using LLE [20] and estimated the 3D pose of the target from the appearance using a nonlinear mapping. However, they did not provide a predictive mechanism. Rahimi et el. [18] proposed to use second order dynamics on appearance manifolds. While successful in many scenarios, this approach suffers from the fact that a tracker must rely on the assumed dynamics to produce estimates of the future appearances, introducing a potential source of fragility. A mismatch between this model and the actual dynamics will lead to incorrect predictions [4] [1]. This problem is illustrated in Figure 1 where it is seen that appearance templates generated using constant acceleration dynamics on the appearance manifold, deteriorate much more rapidly than templates generated using a dynamic model identified by a robust identification technique on the same manifold. Furthermore, it should also be noted that neither of the above approaches identifies the dynamics of the target motion and therefore do not produce accurate predictions of where the target should be sought for in the next frame. This makes trackers particularly fragile to clutter and occlusion as illustrated in Figure 2. Here, a Condensation tracker [11] using dynamic appearance templates in conjunction with constant velocity motion dynamics fails to find the target after the person is occluded by the tree.

In this paper, we present a robust human tracking system which learns dynamical appearance and motion models with accurate predictive power from a small set of initial frames. The learning module employs a robust system dynamic identification technique based on Caratheodory-Fejer (CF) interpolation [21]. The benefits of using this new identification method are multiple. Firstly, it does not require prior knowledge of a state space realization of the dynamic systems, or even their order. Secondly, it provides mechanisms to invalidate a priori assumptions about the dynamics 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 system state are valid and, in the context of *robust* filters such as mixed $\mathcal{H}_2/\mathcal{H}_\infty$ [19], to improve tracking robustness. Last but not least, because CF interpolation finds a model to fit the data instead of forcing an a priori model onto the data, the identified model can accurately predict *future* appearances and target position, even in the presence of *long term* occlusions. As a result, the predicted appearances are very effective as templates for robust tracking.

The remainder of the paper is organized as follows. Section 2 gives an overview of the proposed approach. Section 3 and 4 describe the learning modules of the robust dynamic appearance and motion model, respectively. Section 5 discusses how to use these models to track moving humans. Section 6 demonstrates experimental results. Finally, Section 7 presents conclusions and suggests future work.

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

2. Overview of the proposed approach

The overall tracking system is based on the following ideas:

1. At each time t, the target is located in the image and its current high dimensional appearance of the target in image space is represented by a point on a low dimensional manifold found using a nonlinear mapping that preserves the spatial and temporal neighborhoods and where the time evolution is governed by piecewise linear dynamics. This idea leads to a *separation type principle*, that allows for separating and independently identifying the appearance and target motion dynamics;
2. this low dimensional point is the output at time t, of a linear time invariant (LTI) dynamical system which is identified from a small set of frames using a robust identification procedure;
3. future outputs of the LTI system are *accurately* predicted by the dynamic evolution of the the system on the manifold;
4. future high dimensional appearances of the target in the image space are predicted by a nonlinear inverse mapping applied to these predicted outputs;
5. the location of the predicted appearance in the image space can be predicted by the output of a linear time

invariant (LTI) system which is identified from measurements of the target positions at the training images using a robust identification procedure; and

6. the predicted appearances at the predicted locations are used as dynamic templates for tracking, target identification, and occlusion detection in conjunction with a Kalman or a Condensation filter.

Each of these ideas are described in detail in the sequel.

3. DAM: Dynamic Appearance Modeling

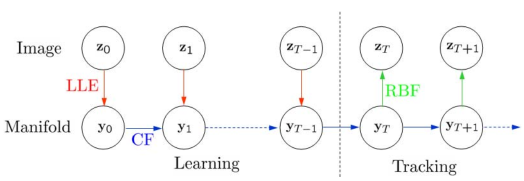

Figure 3. A generative model for learning and tracking appearance. $\mathbf{z}_t$ and $\mathbf{y}_t$ represent the appearances in the image and manifold space respectively.

The process of learning dynamic appearance models consists of three steps: dimensionality reduction, dynamics identification, and inverse mapping as illustrated in Figure 3. The proposed approach learns the dynamics in the manifold space and predicts future appearances based on the identified dynamics.

3.1. Nonlinear Dimensionality Reduction for DAM

Appearance changes due to human motions such as walking or running can often be represented by a small number of latent variables, which can be found by dimensionality reduction. Principal Component Analysis (PCA) is a simple and useful tool for dimensionality reduction. However, it can project faraway high dimensional data points to nearby points in the lower dimensional space if the data do not lie in a linear subspace.

Nonlinear dimensionality reduction methods are designed to preserve local neighborhood of the points by analyzing the underlying structure of the data. This is a highly desirable property when modeling temporal visual appearance since we would like to preserve temporal continuity in the low level representation of the data. In this paper, we utilize the Local Linear Embedding algorithm for nonlinear dimensionality reduction proposed by [20] to obtain a low dimensional appearance representation from a small video sequence. Other possibilities include Isomap [22], Laplacian Eigenmaps [2], and Hessian LLE [6].

Assume that a sequence of T frames of the target are collected by tracking the target while it is initially unoccluded. The frames are pre-processed and normalized to a size of $w \times h$ where w and h represent the width and height of a window capturing the entire target, respectively. The initial tracking and pre-processing are discussed in more detail in Section 6. Let $\mathbf{z}_t, t = 0, \ldots, T-1$ be a $(w * h) \times 1$ vector obtained by rasterizing the window at time t and let $\mathbf{y}_t$ be the $l \times 1$ vector of its low dimensional representation, where l is the dimension of the manifold.

The procedure to find the low dimensional representation by LLE is summarized as follows.

1. Find a set of nearest neighbors $\mathbf{z}_j$ for each vector $\mathbf{z}_t$.
2. Calculate the weights $\mathbf{w}_{tj}$ to best reconstruct $\mathbf{z}_t$, minimizing $\sum_t \|\mathbf{z}_t - \sum_j \mathbf{w}_{tj}\mathbf{z}_j\|^2$ where $\sum_j \mathbf{w}_{tj} = 1$.
3. Find the vector $\mathbf{y}_i$ best reconstructed using the weights $\mathbf{w}_{tj}$, minimizing $\sum_t \|\mathbf{y}_t - \sum_j \mathbf{w}_{tj}\mathbf{y}_j\|^2$.

The minimization in Step 3 can be efficiently solved by calculating the eigenvectors of $M = (I - W)^T(I - W)$ with the smallest eigenvalues, where I is the identity matrix and W is formed with the weights $\mathbf{w}_{tj}$.

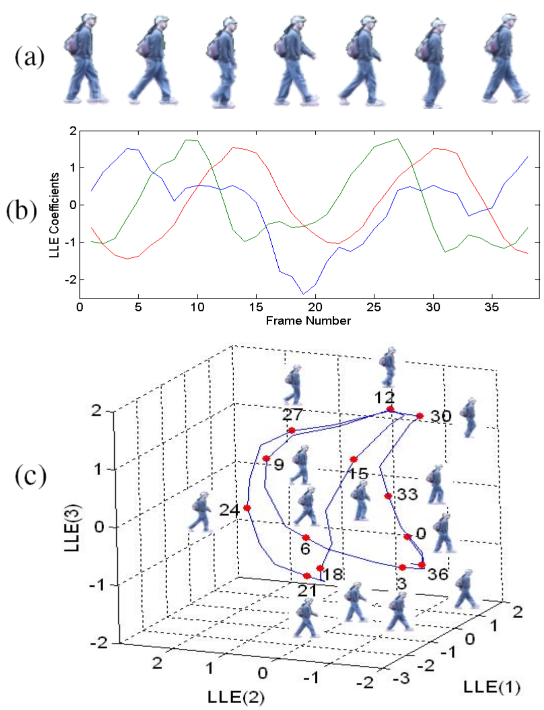

Figure 4. (a) Dynamic appearances of a walking motion for 37 frames, $t = 0, 6, 12, \ldots, 36$. (b) Low dimensional representation of the walking sequence by LLE. (c) 3 dimensional representation of the walking sequence by LLE. Periodicity of walking cycle is observed in the manifold space

Figure 4(b) shows the temporal evolution of the coordinates of the points on the LLE manifold for the first 37 frames of the walking sequence. Figure 4(c) illustrates how a set of collected appearances of a walking motion as the one shown in Figure 4(a) can be represented in a 3 dimensional space, such that the temporal ordering of the original

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

data is preserved. In the next step, we will find a dynamical model to predict future values of these coordinates, based on their past measurements.

3.2. Robust Dynamics Identification for DAM

Assume that the low dimensional representation of the target appearance at time t, $\mathbf{y}_t$ is related to the previous appearance representations by an ARMAX model of the form:

$$\mathbf{y}_t = \sum_{i=1}^{m} g_i \mathbf{y}_{t-i} + \sum_{i=0}^{m} h_i \mathbf{u}_{t-i} \quad (1)$$

where g_i, h_i are fixed coefficients and $\mathbf{u}_t$ denotes a stochastic input. This can be always assumed without loss of generality, since given m measurements of $\mathbf{y}$ and $\mathbf{u}$, there always exist a linear time invariant system such that (1) is satisfied ([21], Chapter 10). This system, in turn, can be represented using a state space description of the form:

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t \end{aligned} \quad (2)$$

where

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 0 & 0 & \dots & g_m \\ 1 & 0 & \dots & g_{m-1} \\ 0 & 1 & \ddots & g_{m-2} \\ \vdots & \vdots & \dots & g_1 \end{bmatrix} \mathbf{B} = \begin{bmatrix} h_m \\ h_{m-1} \\ h_{m-2} \\ \vdots \\ h_1 \end{bmatrix} \\ \mathbf{C} &= \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix} \ \mathbf{D} = h_0 \end{aligned} \quad (3)$$

Note that this (minimal) realization is unique, up to a coordinate transformation. In the sequel we will use the shorthand $\mathcal{F}$ to denote the operator $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ that maps the exogenous input $\mathbf{u}$ to the measured feature position $\mathbf{y}$, and use robust identification techniques to extract these matrices from the experimental data. The main advantages of this approach are the facts that (i) it fully exploits any a-priori information available about the motion modalities of the target, (ii) it does not necessitate making a-priori assumptions about either the order or the specific structure of the model, and (iii) it provides worst–case bounds on the identification and subsequent prediction errors. In order to apply this formalism, we will assume that the following a-priori information is available:

1. A bound of the norm of the measurement and process noises.
2. The operator $\mathcal{F}$ can be written as the sum of a *parametric* and a *non-parametric* component: $\mathcal{F} = \mathcal{F}_p + \mathcal{F}_{np}$ where $\mathcal{F}_p = \sum_{j=1}^{n} p_j \mathcal{F}^j$. Here $\mathcal{F}^j$ are known, given, not necessarily ℓ_2 stable operators that contain all the information available about possible modes of the system[2]. This information can be obtained by performing an FFT of the measured data or by learning appearances off-line while observing persons performing representative motions. The residual non-parametric operator $\mathcal{F}_{np}$, which will be obtained as a by–product of the identification, provides a measure of the quality of the approximation $\mathcal{F}_p$.

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

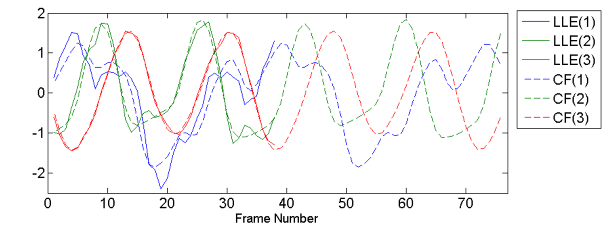

Figure 5. Dynamics Identification by CF interpolation and prediction for next 38 frames.

It can be shown that the problem of finding the coefficients p_j and a state–space representation of $\mathcal{F}_{np}$ can be reduced to a tractable convex optimization problem and efficiently solved using commercially available software. Further details can be found in [21, 15, 4, 13]. Figure 5 illustrates the interpolation and prediction of the three coefficients for the LLE manifold shown in Figure 4(b).

3.3. Recovering Dynamic Appearance from the Manifold Space

An inverse mapping from the manifold to the image space – estimating each pixel value at time t from the low dimensional appearance representation $\mathbf{y}_t$ – can be learned from the training data by employing using Radial Basis Function (RBF) [17]. Once the RBF network learns the nonlinear mapping between pairs of manifold points and appearance pixels, intermediate and future image appearances can be generated from points on the manifold corresponding to the current/future state.

Let z_t^k represent the value of the k^{th} pixel of the appearance vector $\mathbf{z}_t$ at time t, $k = 0, \dots, w*h-1$ and let $\mathbf{y}_t^e$ be the estimated low dimensional representation for $\mathbf{z}_t$. The mapping function from $\mathbf{y}_t$ to z_t^k is defined as $f^k(\mathbf{y}_t^e) = z_t^k$ such that

$$f^k(\mathbf{y}_t^e) = \sum_{i=0}^{N-1} c_i^k h(\|\mathbf{y}_t^e - \mathbf{y}_i^c)\|) + \sum_{i=0}^{M-1} d_i^k p_i(\mathbf{y}_t^e) \quad (4)$$

where $\sum_{i=0}^{N-1} c_i^k p_i(\mathbf{y}_t^e) = 0$, $h(\cdot)$ is a radial basis function, $p_i(\cdot)$ is a linear polynomial, and $\mathbf{y}_i^c$ are N centers of $\mathbf{y}_t$, and the coefficients c_i and d_i can be calculated by solving a linear system [7, 17].

Figure 6 shows appearances for the walking sequence of Figure 4(a) recovered by the interpolated and predicted states on the manifold space. It is seen there, that the learned

Figure 6. Generated walking sequence for 79 frames by RBF, $t = 0, 6, 12, \ldots, 78$.

RBF network successfully interpolates the unseen appearances.

4. DMM: Dynamic Motion Modeling

Given a set of measurements of the position of the target in image space, one can use the same techniques described in section 3.2 to identify the motion dynamics of the target. This dynamics in turn, can be used to accurately predict the location of the target in future frames, even in the presence of occlusion.

5. Human Tracking using DAM & DMM

Unlike other approaches, our approach utilizes the motion dynamics of the target in the image space to predict the location of the target *and* the dynamics of the appearance nonlinear projections on the manifold space to generate the dynamic appearance templates that can be sought at the predicted locations. The advantage of our approach is that the dynamics, which are identified on line from a small number of training frames, provide accurate location and visual information of the target over time.

Once the target location and its appearance are predicted, template-based tracking techniques such as Condensation [11] or affine region tracking [8] can be used to follow the target. In this paper we used a Condensation filter where the likelihood of the target location and appearance was measured by evaluating the pixel-wise color similarity between the tracked region and the predicted template.

5.1. Occlusion Handling

Two cases of occlusions are taken into account in this paper; occlusion by background objects and occlusion by other moving people. In the first case, occlusion is detected by comparing the likelihood against a threshold. The appearance and position of the template are predicted based on the identified dynamics. The predictions of the position and appearance are conducted until the target person is detected again using color similarity. As mentioned before, the success of the proposed approach hinges on the close fit of the system dynamics to the data provided by robust identification [4].

In the second case, the occlusion is detected by the collision of multiple templates of moving people. If the occlusion is detected, the probability that the template is likely to be the actual appearance within the region overlapping the two templates decides between the front or rear layer. The likelihood is measured by color similarity within the overlapped region. If the template is assigned to the front layer, the tracker for that person continues to track. On the other hand, if the template is assigned to the rear layer, the tracker handles the occlusion as in the first case – e.g., the person in the front layer is treated as a background object with respect to the person in the rear layer.

6. Experimental Results

The proposed approach was tested on several outdoor scenes captured with stationary and moving color cameras. In every case, the dynamic appearance of the target was learned from frames from the beginning of the sequences, where it was assumed that the target was not occluded and that the camera was stationary while learning.

6.1. Implementation Details

The foreground regions were extracted by background subtraction and morphological operations were applied to the binary foreground to remove noise and fill in regions. The target appearance was then obtained using this binary mask. The meanshift algorithm was then applied to capture the appearance into a rectangular window, which was normalized to a given size. To reduce the dimensionality of the appearance vector, YCbCr 4:2:0 color compression was applied to reduce the image data by a half while preserving almost all color information. After applying color compression, the appearance vector $\mathbf{z}_t$ at time t was obtained by rasterizing the preprocessed window.

The corresponding 3 dimensional representation $\mathbf{y}_t$ was then obtained by applying the LLE algorithm using a set of the appearances $\mathbf{z}_t$ for T frames. Next, the dynamic evolution of the low dimensional representation was identified using CF interpolation. For the experiments reported here, the parametric component was set to $F_p \in$ span $[\frac{z}{z+1}, \frac{z^2-\cos\omega z}{z^2-2\cos\omega z+1}, \frac{\sin\omega z}{z^2-2\cos\omega z+1}]$ where $w = [w_0/2, w_0, 2w_0, 4w_0]$. The basis frequency w_0 was found from the Fourier Transform of $\mathbf{y}_t$ for T frames.

Finally, a RBF network was used to reconstruct the high dimensional representation. Each pixel z_k^t was reconstructed by nonlinear mapping from the 3 dimensional manifold space.

6.2. Tracking Results

Three experimental results are reported here; tracking moving people in the presence of occlusion by a scene ob-

Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)
0-7695-2597-0/06 $20.00 © 2006 IEEE

Figure 7. Handling occlusion by the background object, $t = 60, 75, 90$, and 105. The white box in the second row is the predicted appearance and the black box is the probability map based on the predicted appearance.

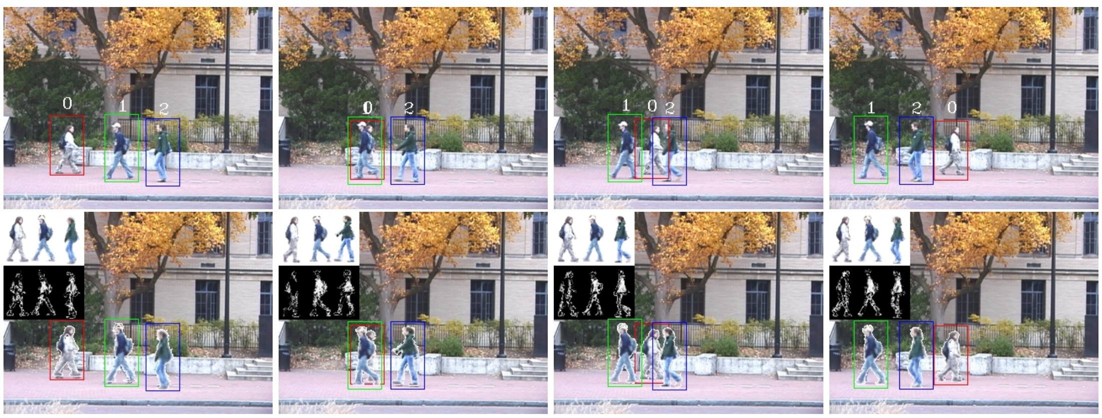

Figure 8. handling occlusion by other moving people, $t = 40, 60, 70$, and 85. The layer is automatically selected based the likelihood of each template to the observed appearance with the overlapped regions.

ject, occlusion by other moving people, and tracking using a moving camera.

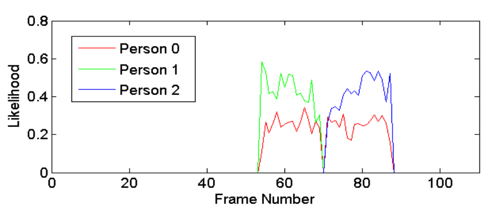

Figure 9. Layer selection by the likelihood to the observed appearance in the overlapped region by two templates. The higher likelihood represents the front layer and vice versa.

Figure 7 shows again the video sequence introduced in Figure 2 but using the proposed approach. It is seen that in this case, the appearances and their location are accurately predicted even during long term occlusion by the tree. These predictions allow the tracker to “hallucinate” the target while it is occluded and to recover it after the occlusion is over. It also should be noted that the template remains uncorrupted by the occlusion, unlike all adaptive appearance techniques, preventing tracking drift.

Figure 8 shows the use of the predicted appearances for layer selection among people. As illustrated in the Figure 9, the likelihood of the person 0 is lower than the ones for person 1 and 2, while they walk across each other, respectively. Thus, person 0 is assigned to the rear layer and persons 1 and 2 are assigned to the front layer while occluded. Since the dynamic appearance is individually learned for each person, the trackers can track and identify each target person even during occlusion and in the presence of clutter.

Finally, Figure 10 shows that once the dynamic appearance is learned, the tracker can track the target person even though the camera moves.

7. Conclusions and Further Work

In this paper, we formulated dynamic appearance and motion modeling for human tracking as a three step process: dimensionality reduction, dynamics identification, and inverse mapping. The proposed approach predicts the loca-

Figure 10. Moving camera after learning appearance, $t = 170, 180, 190,$ and 225. The tracker robustly track the target person even though the camera moves.

tion of the target and predicts its future appearance based on accurately identified dynamics learned from a small set of initial frames. The predicted appearances can then be used as dynamic templates for tracking and identifying moving people even in the presence of occlusion and clutter. We are currently working on incorporating the worst identification error bounds provided by CF to perform model (in)validation and decide when the current appearance model is not longer valid (e.g. if the target changes activity from walking to running) and to decide when a new model should be obtained.

References

[1] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice Hall, New Jersey, 1979.

[2] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

[3] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *IJCV*, 26(1):63–84, 1998.

[4] O. Camps, H. Lim, C. Mazzaro, and M. Sznaier. A caratheodory-fejer approach to robust multiframe tracking. In *ICCV*, volume 2, pages 1048–1055, October 2003.

[5] M. Dimitrijevic, V. Lepetit, and P. Fua. Human body pose recognition using spatio-temporal templates. In *ICCV workshop on Modeling People and Human Interaction*, 2005.

[6] D. L. Donoho and C. E. Grimes. Hessian eigenmaps: locally linear embedding techniques for high-dimensional data. In *Proceedings of the National Academy of Arts and Sciences*, volume 100, pages 5591–5596, 2003.

[7] A. Elgammal and C.-S. Lee. Inferring 3D body pose from silhouettes using activity manifold learning. In *CVPR*, volume 2, pages 681–688, 2004.

[8] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 20(10):1025–1039, 1998.

[9] B. Han and L. Davis. On-line density-based appearance modeling for object tracking. In *ICCV*, 2005.

[10] J. Ho, K.-C. Lee, M.-H. Yang, and D. Kriegman. Visual tracking using learned linear subspaces. In *CVPR*, volume 1, pages 782–789, 2004.

[11] M. Isard and A. Blake. Condensation-conditional density propagation of visual tracking. *IJCV*, pages 5–28, 1998.

[12] A. D. Jepson, D. J. Fleet, and T. F. El-Maraghi. Robust online appearance models for visual tracking. *PAMI*, 25(10):1296–1311, 2003.

[13] H. Lim, O. Camps, and M. Sznaier. A caratheodory-fejer approach to dynamic appearance modelling. In *CVPR*, pages 301–307, June 2005.

[14] J. Lim and D. Kriegman. Tracking humans using prior and learned representations of shape and appearance. In *IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pages 869–874, 2004.

[15] C. Mazzaro, M. Sznaier, and O. Camps. A model (in)validation approach to gait recognition. In *Proc. IEEE Int'l Symp. on 3D Data Processing, Visualization and Transmission*, pages 700–703, 2002.

[16] P. Perez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *ECCV*, pages 661–675, 2002.

[17] T. Poggio and F. Girosi. Networks for approximation and learning. In *Proc. of IEEE*, volume 78, pages 1481–1497, 1990.

[18] A. Rahimi, B. Recht, and T. Darrell. Learning appearance manifolds from video. In *CVPR*, volume 1, pages 868–875, 2005.

[19] H. Rotstein, M. Sznaier, and M. Idan. h_2/h_∞ filtering: Theory and an aerospace application. *International Journal of Robust and Non–Linear Control, Special Issue on h_∞ and Robust Estimation*, 6(4):347–366, 1995.

[20] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science Magazine*, 290(5500):2323–2326, Dec. 2000.

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

[22] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

0-7695-2597-0/06 $20.00 © 2006 IEEE