Markov-Based Failure Prediction for Human Motion Analysis

Shiloh L. Dockstader[†‡], Nikita S. Imennov[#], and A. Murat Tekalp[‡*]
[‡]*Dept. of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627*
[*]*College of Engineering, Koç University, Istanbul, Turkey*
[#]*Dept. of Biomedical Engineering, University of Rochester, Rochester, NY 14627*

Abstract

This paper presents a new method of detecting and predicting motion tracking failures with applications in human motion and gait analysis. We define a tracking failure as an event and describe its temporal characteristics using a hidden Markov model (HMM). This stochastic model is trained using previous examples of tracking failures. We derive vector observations for the HMM using the noise covariance matrices characterizing a tracked, 3-D structural model of the human body. We show a causal relationship between the conditional output probability of the HMM, as transformed using a logarithmic mapping function, and impending tracking failures. Results are illustrated on several multi-view sequences of complex human motion.

1. Introduction

One of the greatest limitations in human motion analysis is the underlying difficulty of tracking the human body for subsequent interpretation [1]. We focus specifically on the application of gait analysis in which a number of relevant gait variables must be extracted from a moving structural model of the human body [2]. As is the case with many applications, successful gait analysis demands a robust tracking algorithm to enable the extraction of useful features and variables. Rather than addressing the tracking issue directly, as has been done previously [3], we turn our attention to an equally important concern: the detection and prediction of motion tracking failures. Armed with an effective prediction mechanism, one can then leverage any number of self-correcting or model-switching methods to enhance the longevity of the principal tracking and feature extraction.

Despite an obvious need by applications in motion tracking, little attention has focused on the detection, prediction, and analysis of terminal failures. Pasqual *et al*. [4] introduce an algorithm that explicitly addresses the uncertainty of tracking. They suggest a method of feature substitution using optical flow, texture, and implicit depth and then switch modalities as needed to prolong and enhance tracking performance. Darrell *et al*. [5] present an interesting approach to tracking using depth estimation, color segmentation, and intensity pattern classification. The method effectively increases tracking robustness using multiple modalities, but does not explicitly address the detection or prediction of tracking failures. In an attempt to leverage the knowledge of tracking failures, Shearer *et al*. [6] employ complementary region- and edge-based algorithms for tracking objects. The approach includes simultaneous monitoring for tracking confidence and uses this information to share data between the two algorithms. The method of detecting failures, however, is quite fundamental, not entirely robust, and does not lend itself to prediction. Dockstader *et al*. [7] introduce a probabilistic model for quantifying, but not predicting or detecting, tracking failures.

The majority of previous work in the field of failure detection and prediction has occurred not in the vision community, but elsewhere [8]. Dobra and Festila [9] develop a technique for detecting failures based on coefficient changes and statistical decision methods. Mehra *et al*. [10] use an Interacting Multiple Model Extended Kalman Filter (IMM-EKF) to detect and identify failure modes. The method represents each failure mode by a model and combines the outputs of the models to detect failures. Similarly, Cho and Paolella [11] suggest a constant-gain extended Kalman filter as a vehicle for failure detection. The proposed algorithm is robust to new dynamics and process noise while remaining effective in detecting failures. With applications in control theory, Doraiswami *et al*. [12] propose a three-stage process for failure detection and isolation. The method isolates faults by computing the maximum correlation between residual measurements and estimates of the residual generated using a number of failed hypotheses.

[†] Corresponding Author: dockstad@ieee.org

Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV 2003) 2-Volume Set
0-7695-1950-4/03 $17.00 © 2003 IEEE

Unlike previous work, this research introduces an effective means of detecting and predicting tracking failures using a combination of stochastic modeling and logarithmic mapping. The approach assumes a parametric, structural model of the human body, characterized by a set of noise covariance measurements. We define a tracking failure as an event and build a corresponding hidden Markov model as its representation. In particular, the observation sequence for the HMM is constructed from the time-varying noise covariance matrices associated with a 3-D Kalman filtering algorithm. We show an ability to detect tracking failures based on the conditional output probability of the trained HMM and an ability to reliably predict failures by mapping this probability through a logarithmic transformation function.

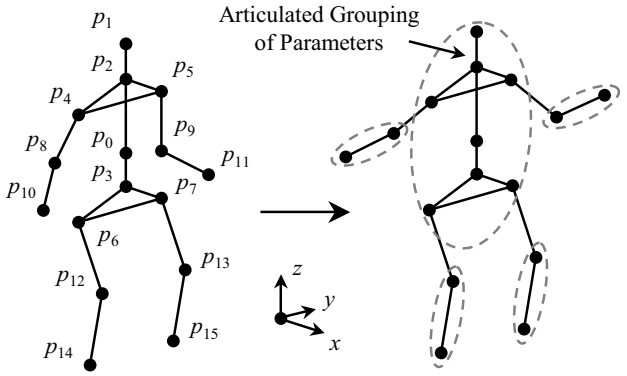

Figure 1. Structural model of the human body.

A hidden Markov model is a natural tool for representing the time-varying correlations in stochastic processes. HMMs have been successfully used in the areas of speech recognition [13], motion recognition [14], gesture recognition [15][16], and general image analysis [17]. In this paper we extend their application to the modeling of motion tracking failures.

2. Theory

The proposed method of modeling tracking failures is built upon an occlusion-adaptive, multi-view algorithm for feature tracking [3]. The tracking is applied to a 3-D structural model of the human body and characterized by stochastic kinematic constraints that limit the variability and improve the accuracy of the underlying body parameter estimates [2].

2.1. Structural Modeling

The suggested structural model employs fifteen parameters ($p_1 \ldots p_{15}$) that are measured in three-dimensional, body-centered coordinates, as indicated in Figure 1. The origin of the coordinate system, p_0, corresponds to a fixed position on the 3-D model. The time-varying coordinate axes are uniquely determined at each frame, k, by interpreting the velocity of the origin. We assume that during a normal gait cycle, the body moves forward, tangential to the transverse (x-y) and sagittal (x-z) planes and orthogonal to the coronal plane (y-z).

The tracking algorithm is implemented using a Kalman filter due to its convenient application of dynamics via linear systems theory. Any number of techniques might also be considered, however [18]. We introduce a time-varying state vector as

$$\boldsymbol{\sigma}[k] \equiv \left[\boldsymbol{\sigma}_1[k]\ \boldsymbol{\sigma}_2[k]\ \cdots\ \boldsymbol{\sigma}_m[k]\ \cdots\ \boldsymbol{\sigma}_{2N+1}[k]\right]^T . \quad (1)$$

Here, $\boldsymbol{\sigma}_m[k]$, $m \le N$ denotes the 3-D position of the m^{th} parameter in our body-centered coordinate system, while

$$\boldsymbol{\sigma}_{m+N+1}[k] = \left.\frac{\partial \boldsymbol{\sigma}_m[k]}{\partial k}\right|_{m \le N} \quad (2)$$

indicates an approximation of the true velocity of the m^{th} parameter. The corresponding state equation is given by

$$\hat{\boldsymbol{\sigma}}[k] = \boldsymbol{\Psi}[k]\hat{\boldsymbol{\sigma}}[k-1] . \quad (3)$$

We develop an error covariance matrix, $\boldsymbol{\Gamma}[k \mid k-1]$, that depicts our confidence in the predictions of the state estimates. The update equation is indicated by

$$\boldsymbol{\Gamma}[k \mid k-1] = \boldsymbol{\Psi}[k]\boldsymbol{\Gamma}[k-1]\boldsymbol{\Psi}^T[k] + \mathbf{Q}[k] , \quad (4)$$

where $\mathbf{Q}[k]$ represents a Gaussian noise covariance matrix which is iteratively modified over time to account for the deviations between the predictions and corrections of the state estimates. The system then defines a Kalman gain matrix, $\mathbf{D}[k]$, according to

$$\mathbf{D}[k] = \boldsymbol{\Gamma}[k \mid k-1]\boldsymbol{\Phi}^T[k] \cdot \left(\boldsymbol{\Theta}[k] + \boldsymbol{\Phi}[k]\boldsymbol{\Gamma}[k \mid k-1]\boldsymbol{\Phi}^T[k]\right)^{-1} , \quad (5)$$

where $\boldsymbol{\Phi}[k] = \left[\mathbf{I} \vdots \mathbf{0}\right]_{N \times (2N+1)}$ indicates the observation matrix and $\boldsymbol{\Theta}[k]$ is a recursively updated observation noise covariance matrix. The remaining steps of the process include

$$\hat{\boldsymbol{\sigma}}[k] = \hat{\boldsymbol{\sigma}}[k \mid k-1] + \mathbf{D}[k]\left(\hat{\mathbf{y}}[k] - \boldsymbol{\Phi}[k]\hat{\boldsymbol{\sigma}}[k \mid k-1]\right) \quad (6)$$

and

$$\boldsymbol{\Gamma}[k] = \left(\mathbf{I} - \mathbf{D}[k]\boldsymbol{\Phi}[k]\right)\boldsymbol{\Gamma}[k \mid k-1] , \quad (7)$$

where $\hat{\mathbf{y}}[k]$ is a vector of three-dimensional, image-derived observations. It is this final noise covariance at each frame that lays the foundation for the prediction of tracking failures. For a more thorough treatment of this theory, including an extension to gait variable extraction, we refer the reader to [2] and [3].

2.2. Tracking Failure Definition

A terminal tracking failure is defined as an immediate and sustained loss in tracking accuracy at one or more

structural model parameters. The acceptable magnitude of such a loss is application dependent and, in the case of this research, driven by the ultimate accuracy required in gait variable extraction. We quantify an immediate loss via the distance between a parameter's estimated and ground-truth values and a sustained loss via the RMS error between the same measures taken over a period of T frames. We designate for the m^{th} model parameter p'_m and p''_m as the error thresholds for immediate and sustained losses in tracking accuracy, respectively.

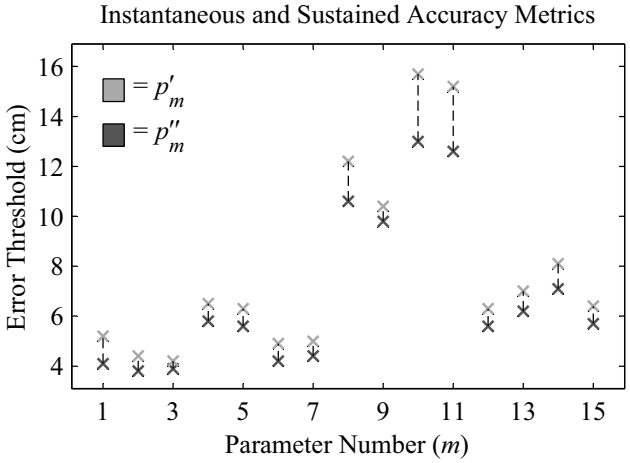

Figure 2. Tracking accuracy threshold parameters.

Our own visual analysis of typical image sequences and mathematical analysis of Kalman noise covariance parameters shows that most tracking failures are preceded by $T \equiv 30$ or fewer frames of relevant data. The accuracy parameters, p'_m and p''_m, corresponding to such failures are empirically derived and summarized in Figure 2.

2.3. Markov Modeling

Tracking failures are not easily characterized by the changing positions of model parameters over the course of time, but are correlated, however, with temporal changes in noise covariance measurements. The matrix $\mathbf{\Gamma}[k]$ is a $(2N+1)\times(2N+1)$ block diagonal matrix in which the 3×3 matrices along the diagonal of the first $N\times N$ quadrant represent the 3-D noise distributions for each of the m model parameters. Let the determinant of the m^{th} matrix at frame k be denoted by $o_m[k] = |\mathbf{\Gamma}_m[k]|$, and let

$$\mathbf{o}_k \equiv \left[o_1[k]\ o_2[k]\ \cdots\ o_m[k]\ \cdots\ o_N[k]\right]^T \tag{8}$$

indicate the vector observation for the entire structural model at frame k. A corresponding observation sequence extracted over T adjacent frames is denoted by $\mathbf{O} = (\mathbf{o}_1\ \mathbf{o}_2\ \cdots\ \mathbf{o}_T)$.

We introduce a nearly ergodic HMM, $\lambda_S = (A, B, \zeta)$, to describe the stochastic properties associated with tracking failures. Using sequences of length $T = 30$ and an observation vector of dimension $N = 15$, one finds in practice that the vector observations are easily clustered into $M = 1024$ discrete symbols, thus yielding a 10-bit HMM codebook. The number of states, R, used by the model is motivated by the articulated structure of the suggested body model. In the majority of training sequences, failures show a dependency on the confidence of the torso parameters as well as the number of accompanying body limbs being successfully tracked. This is in contrast to a dependency on the actual parameters of the various body limbs. Thus, an appropriate number of states is the minimum best describing this phenomenon, or $R = 2\cdot5 = 10$ (the torso + zero to four limbs). This grouping of body parameters and the corresponding HMM topological description are illustrated in Figures 1 and 3, respectively.

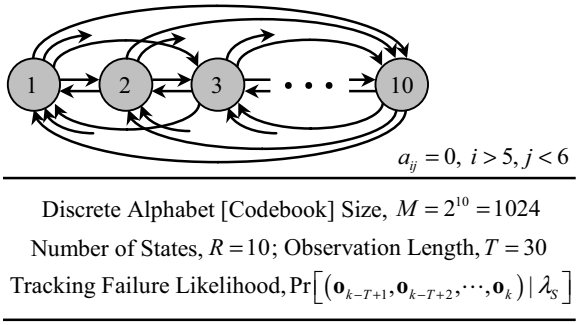

Figure 3. HMM topology.

Estimating the remaining parameters of λ_S is performed using the Baum-Welch method. The only restriction on the topology of the hidden Markov model is that, where $A = \{a_{ij}\}$,

$$a_{ij} = 0,\ i > 5, j < 6\,. \tag{9}$$

This constraint allows for the progression of noise covariance changes from which the algorithm cannot recover (e.g., a tracking degradation within the parameters of the torso). With this restriction, the model is estimated according to

$$\hat{\lambda}_S = \arg\max_{\lambda_S}\left(\prod_i \Pr\left[\mathbf{O}^{(i)} \mid \lambda_S\right]\right), \tag{10}$$

where $\mathbf{O}^{(i)}$ is the i^{th} observation training sequence. The optimization procedure supports the initial choices of $R = 10$, $M = 1024$, and $T = 30$ by producing, at least, a local maximum at these values. The initial state distribution, ζ, the state-transition probabilities, A, and the observation symbol probabilities, B, are then calculated accordingly. Initial values for these parameter estimates are based on a uniform distribution. Figure 4 illustrates the sensitivity, as defined in Table 1 (*Sens*), of the proposed failure detection algorithm as a function of the number of states.

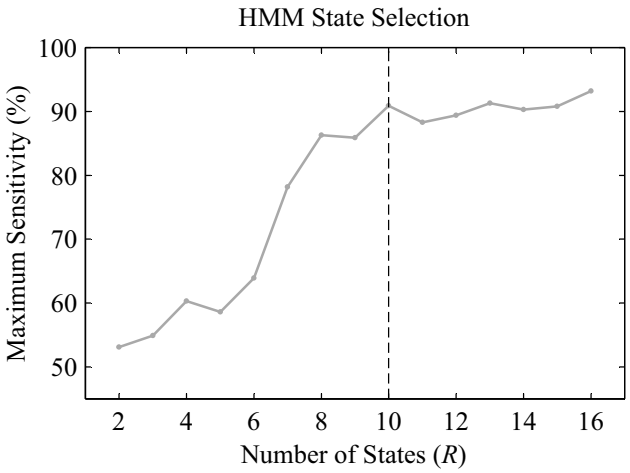

Figure 4. HMM detection sensitivity.

2.4. Tracking Failure Prediction

For all sequences introduced after the initial training set, the measure $\Pr\left[\left(\mathbf{o}_{k-T+1},\cdots,\mathbf{o}_k\right)\mid\lambda_S\right]$ or, alternatively, $\Pr\left[\mathbf{O}\mid\lambda_S\right]$ may be used to test the likelihood that the observation, in which $T=30$, was produced by λ_S. A greater likelihood suggests a more confident measure that, in turn, implies a greater correlation between the observed sequence and those known to correspond with imminent tracking failures. A simple threshold placed on this output probability is a sufficient mechanism for detecting such events. Thus, where a single fixed-state sequence is denoted by $\mathbf{r}=\left(r_1\ r_2\ \cdots\ r_T\right)$, we have

$$\Pr\left[\mathbf{O}\mid\lambda_S\right]\equiv\sum_{\text{all }\mathbf{r}}\Pr\left[\mathbf{O}\mid\mathbf{r},\lambda_S\right]\cdot\Pr\left[\mathbf{r}\mid\lambda_S\right]\ \underset{\text{No Failure}}{\overset{\text{Failure}}{\gtrless}}\ \lambda_S'. \quad (11)$$

The above output probability is estimated using the well-known forward estimation procedure.

An analysis of (11) taken as a function of frame number reveals that there exists between the two a monotonically inverse and somewhat logarithmic relationship. This suggests that the HMM could not only be used to detect failures, but to predict them as well. Thus, we introduce a parametric mapping function between the output probability, $\Pr\left[\mathbf{O}\mid\lambda_S\right]$, and the expected number of frames until the next tracking failure, k_f, according to

$$k_f=a_0+a_1\ln\left(\Pr\left[\mathbf{O}\mid\lambda_S\right]+a_2\right). \quad (12)$$

The parameters (a_0,a_1,a_2) are solved using weighted non-linear least squares, implemented via the Levenberg-Marquardt approach. The weight function used for our observations is exponential in nature and defined as

$$e^{-a_3\left(k_f-1\right)}. \quad (13)$$

The parameter a_3 is built using prior information regarding the known dependence of tracking failures on previous data. In brief, about $a_4=90\%$ of the applicable observations occur in the first half of a $T=30$ frame HMM observation sequence – the remaining correlated measurements occur earlier. It is simple to show that this proportion is satisfied when

$$a_4=\frac{\int_1^{T/2}\exp\left[-a_3\left(k-1\right)\right]dk}{\int_1^{\infty}\exp\left[-a_3\left(k-1\right)\right]dk} \quad (14)$$

or equivalently, when

$$a_3=\tfrac{2}{2-T}\ln\left(1-a_4\right). \quad (15)$$

This yields a parameter of $a_3\approx 0.1645$ and a least squares weighting function of $\exp\left[-0.1645\cdot\left(k_f-1\right)\right]$. This function accounts for the considerable estimation error and little interest associated with predicting tracking failures too far in advance.

3. Experimental Results

To test the contribution we collect a number of video sequences of complex human motion, captured at $\Delta t=1/30$ sec using three or four views of a single scene. Training of both the HMM and the vector quantization scheme is based on approximately 62000 frames of relevant video data, while testing is based on approximately 18000 frames of data. Ground truth measurements are based on a number of methods, including the use of markers placed on the body as well as manual interpretation of feature points. Figure 5 illustrates the tracking results immediately preceding a detected tracking failure. Each column presents data for a unique sequence, while each row shows a progression of frames preceding a correctly detected failure by the trained HMM.

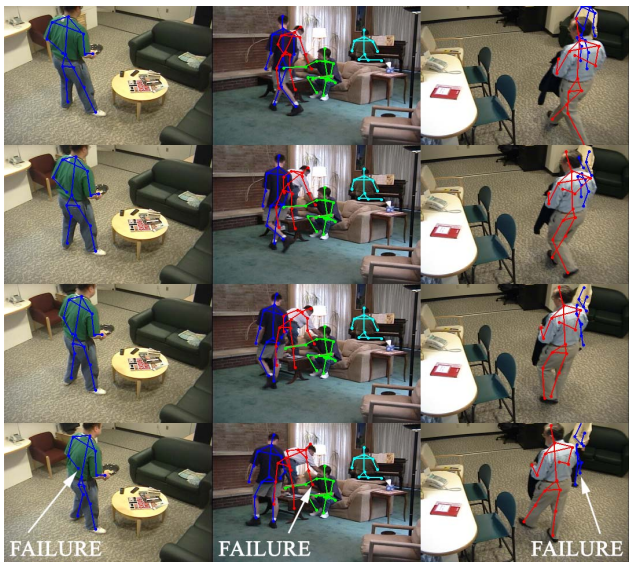

Figure 5. Failed tracking results taken from multiple views and correctly identified as such by HMM.

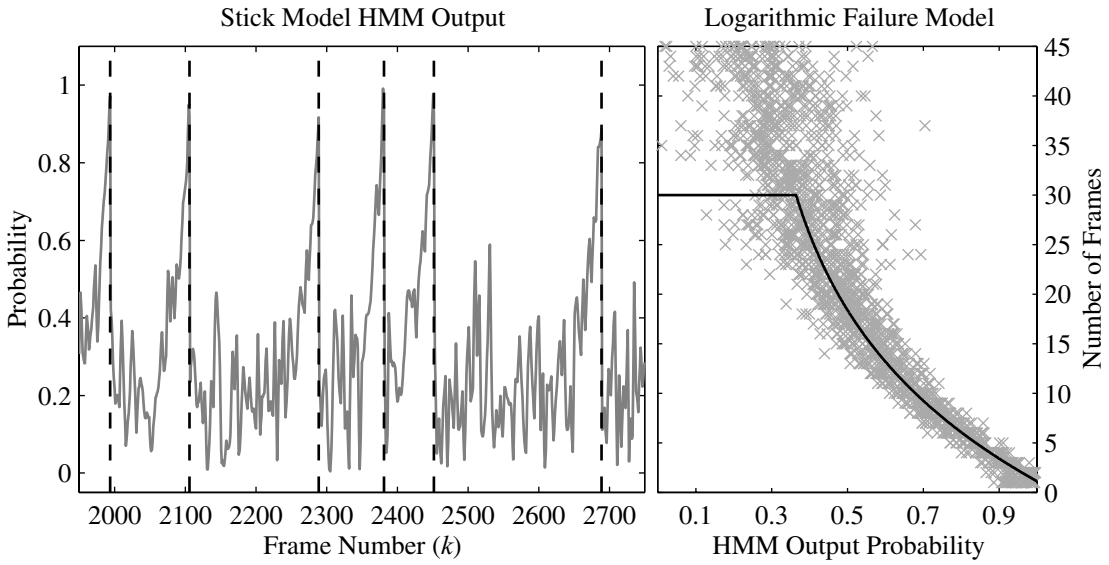

Figure 6. Probability as a function of frame on the left; logarithmic fit to the inverse observations on the right.

The output probability, $\Pr[\mathbf{O} \mid \lambda_S]$, of the HMM for a particular sequence is shown as a function of time in the left side of Figure 6. The vertical dashed lines show known tracking failures, according to our earlier definition, while the gray curve shows the HMM output. There is little correlation between the failures and the output taken more than 15-30 frames in advance. However, as a tracking failure draws closer, the characteristic changes in the Kalman noise covariance measurements drive the output probability to a more deterministic and correlated state.

Table 1. HMM accuracy statistics.

	TP	*TN*	*FP*	*FN*	$\frac{TN}{TN+FN}$ *(Spec)*	$\frac{TP}{TP+FP}$ *(Sens)*	$\frac{TP+TN}{TP+FP+TN+FN}$ *(Accuracy)*
$\|\mathbf{o}_k\|$	1486	14429	1441	144	99.0%	50.8%	90.9%
$\bar{\mathbf{o}}_k$	1491	14504	1366	139	99.1%	52.2%	91.4%
λ_S	1503	15718	152	127	99.2%	90.8%	98.4%

To demonstrate the existence of a temporal correlation in the noise covariance features and, ultimately, the utility of the Markov assumption, we construct more fundamental metrics for comparison. In particular, we consider both $\|\mathbf{o}_k\|$ and

$$\bar{\mathbf{o}}_k \equiv 1/T \sum_{j=k-T}^{k-1} \|\mathbf{o}_{j+1}\| \qquad (16)$$

at each frame k. In all cases, we develop a simple threshold, λ_S', one each for $\Pr[\mathbf{O} \mid \lambda_S]$, $\|\mathbf{o}_k\|$, and $\bar{\mathbf{o}}_k$ that maximizes detection accuracy given a specificity of 99% or greater. This criteria maximizes accuracy without sacrificing specificity or, equivalently, missing too many true failures. Using such thresholds yields the results in Table 1. The proposed Markov scheme produces a maximum accuracy of about 98%, while the alternate metrics each generate only 91%. In the instances of the alternate metrics, the cost of maintaining a specificity of 99% is a lower threshold yielding far too many false positives and, thus, lower sensitivities of only 51% and 52%.

Table 2. Failure prediction distribution.

		Actual Failure Position (%)						
	Frames	1-2	3-4	5-6	7-8	9-10	11-12	13-14
Predicted (%)	1-2	63	28	6	1	0	0	0
	3-4	18	46	20	9	2	1	0
	5-6	3	17	44	19	8	2	1
	7-8	0	2	19	40	23	7	1
	9-10	0	0	1	16	37	21	15
	11-12	0	0	0	8	14	33	21
	13-14	0	0	0	0	3	16	28

The Markov model output probability is transformed using the proposed logarithmic mapping with parameters $(-3.208, -16.367, -0.233)$ and is shown plotted against our observations in the right side of Figure 6. We quantify the accuracy of the logarithmic model in Table 2. Each row represents the number of frames until the next tracking failure, as predicted by the algorithm, while each column indicates the actual number of frames, based on annotated ground truth. For example, the data in row 1 - column 1 states that for all instances in which a failure was predicted to occur in 1-2 frames, a failure *actually* occurred in 1-2 frames 63% of the time. For the same prediction, failures occurred within 3-4 frames 28% of the

time (r1 - c2) and within 5-6 frames 6% of the time (r1 - c3). Naturally, a perfect prediction scheme would yield a diagonal table; errors or deviations from this ideal are indicated in the off-diagonal table elements.

The data in Tables 1 and 2 demonstrates the clear advantage of using the temporal correlations inherent in tracking failure events. The prediction, although prone to some estimation error, is both reliable and robust. Only as the temporal distance to the next failure increases beyond 10-15 frames does the distribution exhibit a considerable amount of variance and skew. While the increase prevents a fully accurate prediction, it at least supports a soft lower bound on k_f for low values of $\Pr[\mathbf{O} \mid \lambda_S]$.

4. Conclusions and Future Work

This research introduces a new method of detecting and predicting motion tracking failures using hidden Markov and logarithmic modeling. The approach defines a failure as an event and uses the output probability of a trained HMM to detect and a logarithmically transformed probability to predict such events. The vector observations for the model are derived from the time-varying noise covariance matrices of a Kalman filter that tracks the parameters of a structural model of the human body. The results clearly show the correlation between the proposed Markov metric and subsequent tracking failures as well as the utility of the Markov model over more fundamental approaches. The proposed theory is demonstrated on several multi-view sequences of complex human motion in support of applications in gait and human motion analysis. Future work in this area will investigate the application of this analysis to automated model switching and increased duration of gait variable and motion signature extraction.

Acknowledgments

This research was supported in part by grants from the Center for Future Health, Eastman Kodak Company, and the RUBI program (NSF-REU# EEC-0097470). The authors also thank Michel Berg for his many suggestions.

References

[1] J. Davis and S. Taylor, "Analysis and recognition of walking movements," *Proc. of the Int. Conf. on Pattern Recognition*, Québec City, Canada, 11-15 August 2002, pp. 315-318.

[2] S. L. Dockstader, K. A. Bergkessel, and A. M. Tekalp, "Feature extraction for the analysis of gait and human motion," *Proc. of the Int. Conf. on Pattern Recognition*, Québec City, Canada, 11-15 August 2002, pp. 1441-1455.

[3] S. L. Dockstader and A. M. Tekalp, "Multiple Camera Tracking of Interacting and Occluded Human Motion," *Proc. of the IEEE*, vol. 89, no. 10, pp. 1441-1455, 2001.

[4] A. A. Pasqual, K. Aizawa, and M. Hatori, "Use of multiple visual features for object tracking," *Proc. of SPIE*, San Jose, CA, 25-27 January 1999, vol. 3653, pp. 946-955.

[5] T. Darrell, G. Gordon, M. Harville, and J. Woodfill, "Integrated Person Tracking Using Stereo, Color, and Pattern Detection," *Int. J. of Computer Vision*, vol. 37, no. 2, pp. 175-185, 2000.

[6] K. Shearer, K. D. Wong, and S. Venkatesh, "Combining Multiple Tracking Algorithms for Improved General Performance," *Pattern Recognition*, vol. 34, no. 6, pp. 1257-1269, 2001.

[7] S. L. Dockstader, M. J. Berg, and A. M. Tekalp, "Performance analysis of a kinematic human motion model," *Proc. of the Int. Conf. on Multimedia and Expo*, Lausanne, Switzerland, 26-29 August 2002, pp. 885-888.

[8] V. B. Vagapov, "Determination of Tracking Failure Probability in Multidimensional Radio Tracking Systems," *Radiotekhnika*, vol. 37, no. 12, pp. 40-42, 1982.

[9] P. Dobra and C. Festila, "Fault detection and diagnosis for continuous time process," *Proc. of the IFAC Workshop on Intelligent Manufacturing Systems*, Bucharest, Romania, 24-26 October 1995, pp. 389-393.

[10] R. Mehra, C. Rago, and S. Seereeram, "Autonomous failure detection, identification and fault-tolerant estimation with aerospace applications," *Proc. of the IEEE Aerospace Conf.*, Aspen, CO, 21-28 March 1998, vol. 2, pp. 133-138.

[11] D. Cho and P. Paolella, "Model-based failure detection and isolation of automotive powertrain systems," *Proc. of the American Control Conf.*, San Diego, CA, 23-25 May 1990, vol. 3, pp. 2898-2905.

[12] R. Doraiswami, C. P. Diduch, and J. Kuehner, "Failure detection and isolation: A new paradigm," *Proc. of the American Control Conf.*, Arlington, VA, 25-27 June 2001, vol. 1, pp. 470-475.

[13] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. of the IEEE*, vol. 77, no. 2, pp. 257-286, 1989.

[14] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using hidden Markov model," *Proc. of the Conf. on Computer Vision and Pattern Recognition*, Champaign, IL, 15-18 June 1992, pp. 379-385.

[15] A. D. Wilson and A. F. Bobick, "Parametric Hidden Markov Models for Gesture Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 884-890, 1999.

[16] H.-K. Lee and J. H. Kim, "An HMM-Based Threshold Model Approach for Gesture Recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 961-973, 1999.

[17] K. Aas, L. Eikvil, and R. B. Huseby, "Applications of Hidden Markov Chains in Image Analysis," *Pattern Recognition*, vol. 32, no. 4, pp. 703-713, 1999.

[18] M. Isard and A. Blake, "Condensation - Conditional Density Propagation for Visual Tracking," *Int. J. of Computer Vision*, vol. 29, no. 1, pp. 5-28, 1998.