NonStationary "Shape Activities"

Namrata Vaswani
Dept. of Electrical and Computer Engineering
Iowa State University
Ames, IA 50011, USA
namrata@iastate.edu

Rama Chellappa
Dept. of Electrical and Computer Engineering
University of Maryland
College Park, MD 20742, USA
rama@cfar.umd.edu

***Abstract*— The changing configuration of a group of moving landmarks can be modeled as a moving and deforming shape. The landmarks defining the shape could be moving objects(people/vehicles/robots) or rigid components of an articulated shape like the human body. In past work, the term "shape activity" has been used to denote a particular stochastic model for shape deformation. Dynamical models have been proposed for characterizing stationary shape activities (assume constant mean shape). In this work we define stochastic dynamic models for non-stationary shape activities and show that the stationary shape activity model follows as a special case of this. Most activities performed by a group of moving landmarks (here, objects) are not stationary and hence this more general model is needed. We also define a piecewise stationary model with non-stationary transitions which can be used to segment out and track a sequence of activities. Noisy observations coming from these models can be tracked using a particle filter. We discuss applications of our framework to abnormal activity detection, tracking and activity sequence segmentation.**

I. Introduction

The changing configuration of a group of moving landmarks (here point objects) can be modeled as a moving and deforming shape. Shape of a group of discrete points (known as 'landmarks') is defined by Kendall [1] as all the geometric information that remains when location, scale and rotational effects are filtered out. The deformation of a moving and deforming shape can be split into rigid motion of an average shape and its non-rigid deformations [2]. In past work [3], we have used the term "shape activity" to denote a particular stochastic model for shape deformation. Dynamical models were proposed for characterizing stationary shape activities (assume constant "mean shape") and statistics defined to detect "abnormality" [4]. In this work we define stochastic dynamic models for nonstationary shape activities and show that the stationary shape activity model follows as a special case of this. Most activities performed by a group of moving landmarks (here, objects) are not stationary and hence this more general model is needed. If the activity is actually stationary it still gets tracked by the non-stationary model. We use our model to track noisy observations using a particle filter [5], [6]. The nonstationary model, being more general, is also more robust to model error and is able to track abnormalities in the activity (which have not been modeled in the training data). Abnormality is detected by using the ELL statistic defined in [4], [7]. Finally, we also define a piecewise stationary model which can be used to segment a given activity sequence into different stationary shape activities and track them. Noisy observations coming from these models can be tracked using a particle filter. We discuss applications to abnormal activity detection, tracking and activity sequence segmentation.

We use a shape based dynamical model for activity because it makes our approach invariant to camera motion, under the weak perspective model (also referred to as the scaled orthographic camera) [8] assumption. The weak perspective model is a valid assumption when the scene depth is much smaller compared to distance from the camera. This is often also the case when the moving objects in the scene are small enough to be treated as point objects, for example in surveillance applications, see Figure 6(a).

The "shape activity" is a generic framework which can used to model dynamics of moving configurations in many applications depending on what is treated as the landmark. The "landmark" can be a person or a vehicle or any moving object. One can learn a shape dynamical model for an activity performed by a group of moving people or model moving traffic and use it to detect abnormal (suspicious) behavior [3], e.g. see Figure 6(a). The "landmark" could be a robot and this framework can be extended to apply feedback control to a group of robots to perform a certain task. Alternatively, the "landmarks" could be the various rigid parts of the human body (see figure 4(a)). Our framework can be used to learn models for the actions and detect and track abnormality in the action. This ability could be useful to medical professionals trying to analyze motion disorders in their patients. Also, the piecewise stationary framework can be used to segment a long sequence into piecewise stationary actions. Our approach is sensor independent. The landmark observations could be obtained by tracking moving objects in low resolution video or using radar, acoustic or infra-red sensors, and only the observation model changes.

We discuss related work and the shape representation used in the next two subsections. The shape activity models are presented in section II. Abnormality detection and tracking are discussed in section III. Shape activity segmentation is explained in section IV. Results are presented in section V and discussion and conclusion in section VI.

A. Related Work

There are many representations for the shape of continuous curves - Fourier descriptors [9], B-splines [10], angle function or curvature based representations [11], deformable snakes and level sets [12]. But in our work, we are trying to model the dynamics of a group of discrete landmarks and so the data is inherently finite dimensional. Hence we use Kendall's representation of shape of a group of discrete landmarks [1], [13]. Our approach is invariant to scaled orthographic camera motion. Other view invariant and quasi view invariant approaches for modeling actions are [14], [15]. Our approach can be made invariant to affine camera motion by replacing the regular shape space by affine shape spaces (chapter 12 of [13]). Also, both [14] and [15] are non-parametric approaches, while we define a shape based parametric model for representing group activity or human actions. Another work which also models human motion using a dynamical model is [16]. They learn a linear dynamical model for the gait of different subjects and use the distance between dynamical models as a metric for gait recognition. Our dynamical model is similar in spirit to [17] where the authors use piecewise geodesic priors to define models for motion on Grassmann manifolds and track them using a particle filter. The application considered there is time-varying subspace estimation.

B. Shape Representation

We use a discrete representation of shape of a group of k **landmarks**. The various moving objects (point objects) in group activity or the rigid parts of human body in an action form the "landmarks". The **configuration** is the set of landmarks, in the 2D case it is the x and y coordinates of the landmarks which can be represented as a k dimensional complex vector [13].

The raw configuration, Y_{raw}, can be normalized for translation (moving origin to the centroid of the configuration) and then for scale (normalizing the translation normalized vector by its Euclidean norm) to yield the **pre-shape**, denoted by w. A configuration of k points after translation normalization, denoted by Y, lies in $\mathcal{C}^{k-1}$ ((k-1)-dimensional complex space) while the pre-shape, w, lies on a hyper-sphere in $\mathcal{C}^{k-1}$. A pre-shape w_1 can be aligned with another pre-shape w_0 by finding the rotation angle for the best fit (minimum mean square error fit) and this gives the **Procrustes fit** of w_1 onto w_0 [13]. This is the **shape** of w_1 w.r.t. w_0. The **Procrustes distance** between preshapes w_1 and w_0 is the Euclidean distance between the Procrustes fit of w_1 onto w_0. The **Procrustes mean** of a set of preshapes $\{w_i\}$ is the minimizer of the sum of squares of Procrustes distances from each w_i to an unknown unit size mean configuration μ [13]. Any pre-shape of the set can then be aligned w.r.t. this Procrustes mean to return the **shape** (denoted by z) w.r.t. the mean shape, μ [13].

Definition 1: The term **"mean shape"**, in this entire paper, is used to denote the minimizer of the expected value (w.r.t. a probability distribution P on the shape space, $\mathcal{M}$) of the squared Procrustes distance of any shape from this shape, i.e. $\mu = \arg\min_\mu E_P[d^2(z,\mu)] = \arg\min_\mu \int_{z\in\mathcal{M}} d^2(z,\mu)P(dz)$, where d is the Procrustes distance.

The shape space, $\mathcal{M}$, is a manifold in $\mathcal{C}^{k-1}$ and hence its actual dimension is $\mathcal{C}^{k-2}$. Thus the tangent plane at any point of the shape space is a $\mathcal{C}^{k-2}$ dimensional hyperplane in $\mathcal{C}^k$ [13]. The tangent coordinate (denoted by v) w.r.t. μ, of a configuration, Y_{raw}, is evaluated [13] as follows[1]:

$$\begin{aligned} Y &= CY_{raw}, \quad \text{where } C \triangleq I_k - 1_k 1_k^T/k \\ s &\triangleq s(Y) = ||Y||, \quad w = Y/s, \\ \theta &\triangleq \theta(Y,\mu) = -\arg(w^T\mu), \qquad z(Y,\mu) = we^{j\theta} \end{aligned} \tag{1}$$

$$v \triangleq v(Y,\mu) = [I_k - \mu\mu^T]z = [I_k - \mu\mu^T]\frac{Ye^{j\theta}}{s} \tag{2}$$

II. The Shape Activity Model

The distinction between motion and deformation of a moving and deforming shape is not clear. We model the motion/deformation of a deforming shape as scaled Euclidean motion of the "mean shape" (translation, rotation, isotropic scaling) plus its non-rigid deformation. The term "shape activity" is used to denote a particular stochastic model for shape deformation [3]. We define a "stationary shape activity" as one for which the shape vector is stationary i.e. the "mean shape" remains constant with time and the deformation model is stationary. Since the "mean shape" is constant and assuming small enough variance, the dynamics in shape space can be approximated by dynamics in the tangent to shape space at the mean (see Figure 1(a)). A partially observed and non-linear model for representing a stationary shape activity was proposed in [3]. It used tangent coordinates of shape w.r.t mean, and the motion parameters (scale, rotation) as the state.

In this work, we define a "non-stationary shape activity" model for which the "mean shape" is time-varying and hence modeling the shape dynamics requires a tangent space (see figure 1(b)) defined w.r.t the current shape. Thus the state space now consists of the "mean shape" at t (given X_{t-1}), the tangent coordinate w.r.t. the current "mean shape" ("shape velocity") and motion parameters - s_t, θ_t. Our model can be understood as a Markov model on "shape velocity" which is parallel transported at each t to the tangent space at the current "mean shape". The stationary shape activity model of [3] is a special case of this nonstationary model and is discussed in Section II-B. We also define a piecewise stationary shape activity model in Section II-C to either model a shape activity with slowly varying "mean shape" or to segment and track a sequence of activities each of which is stationary.

A. Non-stationary Shape Activity (NSSA)

The observed configuration of landmarks after translation normalization, Y_t, forms the observation vector. The "mean shape" at time t, μ_t, the coefficients vector (of the tangent

[1]Note for complex numbers (or vectors), ^{T} denotes conjugate transpose

coordinate of shape w.r.t. the current mean shape), c_t, and the motion parameters (scale s_t, rotation θ_t) form the state vector, i.e. state $X_t = [\mu_t, c_t, s_t, \theta_t]$. Denote the tangent space at μ_t by T_{μ_t}. We then have the following dynamics:

The shape at the previous time instant is used as the current mean shape, i.e. $\mu_t = z_{t-1}$ and so $T_{\mu_t} = T_{z_{t-1}}$. The tangent coordinate of z_t in $T_{z_{t-1}}$ defines a **"shape velocity"**. Since the tangent plane is a $(k-2)$-dim hyperplane in $\mathcal{C}^k$, a tangent vector has only $(k-2)$ independent (complex) coefficients. We perform an SVD (Singular Value Decomposition) [18] of the tangent projection matrix, $[I_k - \mu_t \mu_t^T]C$, to obtain a $(k-2)$-dim orthogonal basis for T_{μ_t}. The basis vectors, $\{\underline{u}_{t,i}\}_{i=1}^{k-2}$, are arranged as column vectors of a matrix, $U_t(\mu_t)$, i.e. ${U_t}^{k\times(k-2)} = [\underline{u}_{t,1}, \underline{u}_{t,2} \cdots \underline{u}_{t,k-2}]$ [2]. The vector of coefficients ($(k-2)$-dim) along these basis directions, $c_t(z_t, \mu_t)$, is thus a canonical representation of the tangent coordinate of z_t in T_{μ_t}. The tangent coordinate is given by $v_t(z_t, \mu_t) = U_t c_t$.

Now, the coefficient vector, c_t is the coefficient vector of the shape velocity, and is thus the multivariate analog of one dimensional speed. We can assume c_t (shape speed) to be i.i.d. Gaussian or define a linear Gauss-Markov model on it. Both these can be summarized by the following model.

$$\begin{aligned} \mu_t &= z_{t-1} \\ c_t &= A_{c,2,t} c_{t-1} + n_t, \quad n_t \sim \mathcal{N}(0, \Sigma_{n,c,2,t}) \\ v_t &= U_t c_t, \quad U_t = \text{orthogonal basis}(T_{\mu_t}) \\ z_t &= (1 - v_t^T v_t)^{1/2} \mu_t + v_t. \end{aligned} \tag{3}$$

One thing to note is that a Markov model on the shape speed corresponds to a second order Markov model on shape, z_t (hence the subscript '2' on the parameters). Some special cases are $A_{c,2,t} = 0$ or i.i.d. speed (first order Markov model on shape); $A_{c,2,t} = I$ which corresponds to i.i.d. acceleration and $A_{c,2,t} = A_{AR}$ or stationary speed.

Motion dynamics can be defined as in [3] or differently depending on the application. We use a Markov log-normal model for the scale parameter, s_t, and a Markov uniform model for θ_t. Note that θ_t here is the rotation angle of current configuration w.r.t. the current "mean shape" $\mu_t = z_{t-1}$ and hence is a measure of rotation speed while in [3] it denotes rotation of current configuration w.r.t. the constant mean. The motion model equations are:

$$\begin{aligned} \log s_t &= \alpha_s \log s_{t-1} + (1 - \alpha_s)\mu_s + n_{s,t} \\ \log s_0 &\sim \mathcal{N}(\mu_s, \sigma_s^2), \quad n_{s,t} \sim \mathcal{N}(0, \sigma_r^2) \\ \theta_t &= \alpha_\theta \theta_{t-1} + n_{\theta,t}, \quad n_{\theta,t} \sim Unif(-a, a) \end{aligned} \tag{4}$$

The shape and motion model (equations (3), (4)) form the **system model**. The **observation model** is as follows:

$$\begin{aligned} Y_t &= h(X_t) + \zeta_t, \quad \zeta_t \sim \mathcal{N}(0, \Sigma_{obs,t}) \\ h(X_t) &= z_t s_t e^{-j\theta_t}. \end{aligned} \tag{5}$$

[2] ${U_t}^{k\times(k-2)} = \text{orthogonal basis}(T_{\mu_t})$ is evaluated as : $U_t = U_{full,t} Q$ where $U_{full,t} S U_{full,t}^T = [I_k - \mu_t \mu_t^T]C$, and $Q = [I_{(k-2)\times(k-2)}, 0_{(k-2)\times 2}]^T$

1) Training: Given a training sequence of centered (translation normalized) configurations, $\{Y_t\}_{t=1}^T$, we first evaluate $\{c_t, v_t, s_t, \theta_t\}_{t=1}^T$ as follows [3] :

$$\begin{aligned} \mu_t &= z_{t-1} \\ s_t &= ||Y_t||, \quad w_t = Y_t / s_t, \\ \theta_t(Y_t, \mu_t) &= -angle(w_t^T z_{t-1}), \quad z_t(Y_t, z_{t-1}) = w_t e^{j\theta_t}, \\ v_t(Y_t, \mu_t) &= [I_k - z_{t-1} z_{t-1}^T] z_t, \\ c_t(Y_t, \mu_t) &= U_t(z_{t-1})^T z_t. \end{aligned} \tag{6}$$

If we assume a time invariant Markov model on c_t, we can use $\{c_t\}_{t=1}^T$ to learn its parameters [3], [18].

B. Stationary Shape Activity (SSA)

For a stationary shape activity, the "mean shape" is constant with time, $\mu_t = \mu_0$, and the shape sequence is clustered around the "mean shape" (see Figure 1(a)). Hence the shape deformation dynamics can be defined in a single tangent space at the mean (which can be learnt as the Procrustes mean [13] of the training data). The SVD of the tangent projection matrix $U_t = U_0 = basis(T_{\mu_0})$ is constant too. $v_t = v_t(Y_t, \mu_0) = U_0 c_t(Y_t, \mu_0)$ is the tangent coordinate w.r.t. the "mean shape" (not tangent velocity) and $\theta_t = \theta_t(Y_t, \mu_0)$ is rotation angle w.r.t. the constant mean (not rotation speed). Since there is a single mean shape, it does not need to be part of the state vector. Thus the state vector is $X_t = [c_t, s_t, \theta_t]$. The dynamics on c_t is defined by the autoregression, $c_t = A_{c,1} c_{t-1} + n_t$. Note that in this case $c_t(Y_t, \mu_0)$ are the tangent coordinates for the shape, z_t and hence the above model corresponds to a first order Markov model on shape, z_t. Also note that in this case, v_t and c_t are related by a constant orthogonal transformation.

C. Piecewise Stationary Shape Dynamics

When the shape is not stationary but is slowly varying, one could model the "mean shape" as being piecewise constant. Now in SSA, the "mean shape" is constant i.e. $\mu_t = \mu_0$ for all t and hence all the dynamics can be described in a single tangent space while in NSSA, the tangent space changes at each time instant: $\mu_t = z_{t-1}$ is the pole of the tangent space at time t. But for PSSA we let the mean μ_t (and hence also the tangent space) be piecewise constant.

Let the "mean shape" change times be $t_1, t_2, t_3, ...$ and the corresponding means be $\mu_1, \mu_2, \mu_3, ...$ Then we have the following dynamics: Between $t_{j-1} < t < t_j$, $\mu_t = \mu_{t-1}$ and so $c_{t-1}(z_{t-1}, \mu_t) = c_{t-1}(z_{t-1}, \mu_{t-1})$. Hence in this interval, the dynamics is similar to that for an SSA, i.e.

$$\begin{aligned} c_t(z_t, \mu_t) &= A_{c,1,t} c_{t-1}(z_{t-1}, \mu_t) + n_t, \\ v_t &= U(\mu_t) c_t, \\ z_t &= (1 - v_t^T v_t)^{1/2} \mu_t + v_t. \end{aligned} \tag{7}$$

At the change time instant, $t = t_j$, $\mu_t = \mu_j$ and so the tangent coefficient c_{t-1} needs to be recalculated in the new

[3] Note, the last equation, $c_t = {U_t}^T z_t$, holds because $c_t = {U_t}^T v_t = {U_t}^T [I - z_{t-1} z_{t-1}^T] z_t = {U_t}^T [I - z_{t-1} z_{t-1}^T] C z_t = {U_t}^T U_t {U_t}^T z_t = {U_t}^T z_t$.

tangent space w.r.t. $\mu_t = \mu_j$. This is achieved as follows:

$$\begin{aligned} c_{t-1}(z_{t-1}, \mu_t) &= U(\mu_t)^T z_{t-1} e^{j\theta(z_{t-1}, \mu_t)} \\ c_t(z_t, \mu_t) &= A_{c,1,t} c_{t-1}(z_{t-1}, \mu_t) + n_t, \\ v_t &= U(\mu_t) c_t, \\ z_t &= (1 - v_t^T v_t)^{1/2} \mu_t + v_t. \end{aligned} \tag{8}$$

Note that in NSSA, v_t is a tangent coordinate w.r.t. $\mu_t = z_{t-1}$ and hence it measures shape velocity while in this case, v_t (and hence also c_t) is a tangent shape coordinate w.r.t. the current "mean shape" μ_t. Hence like in SSA, here also we have a first order Markov model on shape. Hence the subscript '1' on $A_{c,1,t}$.

The times at which the changes occur and the changed means could both be unknown or known or one of them could be unknown. When both change times and the corresponding means are known, PSSA can be used for tracking a sequence of stationary shape activities (each with its known shape mean and known transition times) and detecting abnormality. Abnormality can be defined as ELL w.r.t. the current "mean shape" exceeding a threshold. When times at which the changes occur are unknown, one can use ELL [4], [7] w.r.t. the current "mean shape" to detect a change. This is useful for activity sequence identification (figuring out when one activity ends and the next one starts) and tracking. Both cases are discussed in Section III-C.

When both change times and changed system means are not known, one can detect the change using ELL. The "best" estimate of the shape at the t based on observations $Y_{1:t}$ can be used as the new shape mean. Now since the shape space is nonlinear, the expected value of shape given observations, $E_{\pi_t^N}[z_t]$ (the MMSE estimate), may not lie in the shape space at all. But we can instead estimate a Procrustes mean [13] of the shape which is the minimum mean Procrustes distance square estimator ("mean shape" w.r.t. posterior distribution). It can be evaluated as the largest eigenvector of the matrix $S \triangleq E_{\pi_t^N}[z_t z_t^T] = \frac{1}{N}\sum_{i=1}^{N} z_t^i z_t^{iT}$ [13]. Note that the Procrustes mean is an intrinsic mean for the shape manifold. One can also evaluate the extrinsic mean [11] which is the projection of the Euclidean mean of tangent coordinates, $E_{\pi_t^N}[v_t]$, onto the shape space, i.e. $\mu_t^{extrinsic} = (1 - E_{\pi_t^N}[v_t]^T E_{\pi_t^N}[v_t])^{1/2}\mu_{t-1} + E_{\pi_t^N}[v_t]$. Setting the mean this way will be valid as long as the tracking error (or equivalently the observation likelihood, OL [4], [7]) is still below the tracking error threshold (the posterior π_t^N is estimated correctly). This follows from theorem 4 in chapter 2 of [7]. This form of PSSA can be used for activity sequence segmentation and tracking as discussed in Section IV.

III. Abnormal Activity Detection and Tracking

Now, in the previous section, we have defined stochastic dynamic models for shape and motion dynamics with noisy observations of the configurations forming the observation vector. Filtering needs to be performed to estimate (filter out) the posterior probability distribution of shape (state) given the noisy observations. Since the model is nonlinear, we use a particle filter (PF) [5] which is a sequential Monte Carlo approximation of the optimal non-linear filter. The particle filter [6] is a recursive algorithm which produces at each time t, a cloud of N particles, $\{x_t^{(i)}\}_{i=1}^N$, whose empirical measure (denoted by $\pi_t^N(dx)$) closely "follows" $\pi_t(dx_t|Y_{0:t})$, the posterior distribution of the state given past observations (denoted by $\pi_t(dx)$ in the rest of the paper).

A. Tracking to Obtain Observations

The particle filter also provides at each time instant the prediction distribution, $\pi_t(X_t|Y_{1:t-1})$, which can be used to predict the expected configuration at the next time instant using past observations, i.e. $\hat{Y}_t \triangleq E[Y_t|Y_{0:t-1}] = E_{\pi_{t|t-1}}[h(X_t)]$. We can use this information to improve the measurement algorithm used for obtaining the observations (a motion detector [19] in our case). Its computational complexity can be reduced and its ability to ignore outliers can be improved by using the predicted configuration and searching only locally around it for the current observation[4]. As we show in section V, the observed configuration is close to its prediction when there is no abnormality or change and hence the prediction can be used to obtain the observation. Also, if the configuration is a moving one, then the predicted motion information can be used to translate, zoom or rotate the camera (or any other sensor) to better capture the scene but in this case, one would have to alter the motion model to include a control input.

B. Abnormal Activity Detection

An abnormal activity (suspicious behavior in our case) is defined as a change in the system model, which could be slow or drastic, and whose parameters are unknown. Given a test sequence of observations and a "shape activity" model, we use the change detection statistics defined in [4], [7] to detect a change (i.e. detect when observations stop following the given shape activity model). A change being drastic or slow depends on the system model used in particle filtering. A more general system model can track a lot more changes and hence the nonstationary shape activity model does a better job of tracking abnormal observations than the stationary one. *Whenever changed observations get tracked correctly, the ELL detects the change while if the PF loses track, the tracking error detects the change [4], [7].*

Now for abnormality detection, the normal activity needs to be characterized first. We can either use shape velocity or shape or both to represent normalcy depending on the practical problem being dealt with. To use shape to detect abnormality, we represent a normal activity by a stationary shape activity model or by a PSSA model (whichever is appropriate for a given problem). For simplicity, assume an SSA model for normal activity. Then the normal prior is a time invariant Gaussian distribution of the tangent

[4]One thing to note here is that in certain cases (for example, if the posterior of any state variable is multimodal), evaluating the posterior expectation as a prediction of the current observation is not the correct thing to do. In such a case, one can track the observations using the CONDENSATION algorithm [10] which searches for the current observation around each of the possible $h(\bar{x}_t^i), i = 1, 2...N$.

coordinates w.r.t. μ_0 (the normal activity mean shape), $\mathcal{N}(0, \Sigma_{v,0})$. Now for a Gaussian prior, the discriminating term of ELL reduces to expectation, under the posterior, of the Mahalonobis distance from the prior's mean. We evaluate it as follows: We project the filtered shape of the observations at time t into T_{μ_0} to obtain $v(z_t, \mu_0)$ and evaluate $E_{\pi_t}[v(z_t,\mu_0)^T \Sigma_{v,0}^{-1} v(z_t,\mu_0)]$. Thus given the particle filtered shape distribution $\pi_t^N(dz_t) \triangleq \sum_{i=1}^N \frac{1}{N}\delta_{z_t^{(i)}}(dz_t)$ (which approximates $\pi_t(dz_t)$), we evaluate

$$\pi_t^N(dv_{t,\mu_0}) \triangleq \sum_{i=1}^N \frac{1}{N}\delta_{v_{t,\mu_0}^{(i)}}(dv_{t,\mu_0}), \quad \text{where}$$

$$v_{t,\mu_0}^{(i)} \triangleq v(z_t^{(i)},\mu_0) = [I_k - \mu_0\mu_0^*]z_t^{(i)} e^{j\theta(z_t^{(i)},\mu_0)}. \quad (9)$$

ELL(Shape) is then approximated as

$$ELL^N(Shape) = \frac{1}{N}\sum_{i=1}^N v_{t,\mu_0}^{(i)\ T} \Sigma_{v,0}^{-1} v_{t,\mu_0}^{(i)}. \quad (10)$$

If PSSA is used to define a normal activity, the prior is a Gaussian distribution on the tangent coordinates in the tangent space of the current mean μ_t.

Depending on the practical problem, one might want to use shape velocity (rate of change of shape) to define normalcy. Given that a stationary Gauss Markov model has been defined for the shape velocity, v_t, with parameters $\Sigma_{v,2}, A_{v,2}, \Sigma_{n,v,2}$, the change detection statistic will simplify to $E_{\pi_t^N}[v_t^T \Sigma_{v,2}^{-1} v_t]$ where $v_t = v(z_t, z_{t-1})$ denotes shape velocity[5]. We refer to this statistic as "ELL (Shape Velocity)". Many times, the learnt covariance matrices can be much smaller than the actual variance of v_t and in such cases, a better solution is to use unweighted shape velocity norms.

C. Activity Sequence Identification and Tracking

Consider two possible situations for tracking a sequence of activities. Assume each activity is represented by an SSA so that the sequence of activities is characterized by a PSSA. The "mean shape" of each SSA component is known but the transition times are unknown.

First consider the case when there are two possible activities and their order of occurrence is known, only the change time is unknown. In this case, one can detect the change using ELL (before the particle filter loses track) and start tracking it with the second activity's transition model.

Now consider the general case when a sequence of activities occur, and we do not know the order in which they occur. In this case, we can use a discrete mode variable as part of the state vector to denote each activity type. We make the state transition model a mixture distribution and keep the mode variable as a state. Whenever a change occurs, it takes the mode variable a few time instants to stabilize to the correct mode. One could replace the multimodal dynamics with that of the detected mode once the mode variable has stabilized. Also, in this case we can declare an activity to be abnormal (i.e. neither of the known activity types) if the ELL w.r.t all known models exceeds a threshold.

[5]Note that $v(z_t, \mu_0)$ denotes the tangent shape coordinate of z_t w.r.t. μ_0 while $v_t = v(z_t, z_{t-1})$ denotes the shape velocity

IV. Shape Activity Sequence Segmentation

The PSSA model with unknown mean shapes and unknown change times can be used along with ELL for activity sequence segmentation as follows:

- Track observations using PSSA, until the ELL of tangent coordinates w.r.t. the current μ_t, $ELL(\mu_t) = E_{\pi_t}[v_t^T \Sigma_{v,t}^{-1} v_t]$ exceeds the change detection threshold.
- Use time instants when $ELL(\mu_t)$ exceeds its threshold, as segmentation boundaries.
- If at time t, $ELL(\mu_t)$ has exceeded its threshold but the tracking error is still below its threshold (PF is still in track, i.e. π_t^N approximates π_t^c correctly), then set μ_{t+1} as the posterior Procrustes mean of the shape at t, given past observations, $Y_{1:t}$. This is explained in the last paragraph of section II-C.
- Recalculate v_t and c_t in the new tangent space at μ_{t+1} (as discussed in section II-C).

V. Experimental Results

A. Simulated Shape Sequence

We first simulated a shape activity sequence, starting with a regular hexagon as the mean. The sequence was stationary for the first 40 frames (around the regular hexagon) and for the next 40 frames, a bias was added to the tangent coordinate at every frame, which resulted in unmodeled non-stationary deformations of the shape (abnormality). We also scaled and rotated each frame according to Markov log-normal and uniform models. Four pixel and nine pixel i.i.d. white Gaussian observation noise was added to each frame to produce the observations. Another sequence of training data was generated this time without adding any bias in tangent space (no abnormality). The parameters of both SSA and NSSA were learnt using this normal sequence and with no observation noise added.

We attempted to track the abnormal noisy observations using both SSA and NSSA models. Both SSA and NSSA track the normal observations equally well, see Figure 2(a). But within a few frames of introducing the abnormality SSA loses track, while NSSA is able to remain in track, see Figure 2(b). Even in 9-pixel noise, NSSA is able to track the abnormality, we show the distance from ground truth in Figure 2(c). We also plot the abnormality detection statistics in Figure 3. Both SSA and NSSA are able to detect abnormality using both shape and shape velocity statistics. We show ELL (Shape), ELL (Shape Velocity, Unweighted) and ELL (Shape) for 9-pixel observation noise in 3 (a), (b) and (c) respectively. All statistics have been normalized by their maximum value (to be able to plot SSA and NSSA in one figure).

B. Human Actions

Next we attempted to track human actions and track as well as detect abnormality in the action. We show here results on tracking a figure skater, shown in Figure 4(a). We had

observation noise-free locations of landmarks in the normal skater sequence. The 10 landmarks used were [head, torso, both elbows, hands, knees and feet]. The abnormality was the knee deviating too far away. As before, we used the normal sequence for training SSA and NSSA models; added observation noise to the abnormal one and attempted to track it. We show the tracks (of the landmark locations) along with the ground truth in Figure 4(b) and (c). SSA is able to track the normal sequence better than NSSA while it completely fails for the abnormality. But NSSA is able to track both. In Figure 5, we show the tracking error, the ELL (Shape) and ELL (Shape velocity) plots. NSSA is able to detect using both ELL(Shape) and ELL(Shape Velocity), while SSA can detect only using tracking error.

C. Group Activity

We also show results on a video sequence of people deplaning and moving towards the terminal with the abnormality being either a person stopped in the path or a person walking away in a weird direction [3]. In [3], we used simulated observation noise, but we show here results on real observations. Very noisy observations were obtained by using a motion detection algorithm [19] which we tried to track using NSSA and SSA. We show one image of the moving people video in Figure 6(a). In Figure 6(b), we show the tracking error for a sudden abnormality which is smaller using NSSA than SSA. Also (we have not shown the figure) both NSSA and SSA are able to detect the abnormality. In Figure 6(c), we show the ELL (Shape) plot for detecting a slow abnormality. It can be seen that the NSSA model is able to detect slow changes as well as SSA. Detecting a slow change and tracking a sudden change are the more difficult problems and NSSA can handle both.

VI. Discussion and Conclusion

We have proposed a non-stationary "shape activity" model which has a time varying "mean shape" (and hence a time varying tangent space) and compared it to the stationary shape activity model, which was first proposed in [3]. We have shown application of NSSA to modeling human actions and also to modeling "group activity" (performed by a group of moving objects) and detecting abnormality. We show that the NSSA model can track and detect the abnormality while SSA can only detect it. We characterize a normal activity by a SSA (or a PSSA) model and learn its parameters using training data. Test observations are tracked using NSSA. The ELL statistic [4], [7] is used for abnormality detection. Finally, we have also proposed piecewise stationary shape activity models for modeling a slowly changing mean shape. The piecewise stationary model can be used in conjunction with ELL, to segment a long sequence of shape activities into piecewise stationary segments and to simultaneously track it (explained in section IV).

The NSSA system model is more general than the SSA model and so able to track larger than normal changes in the system model without losing track. Hence it can track more sudden abnormalities (and yet detect them) than a SSA model. For the same reason, it is also more robust to modeling error in learning the system model parameters as long as the observations are good. The SSA model on the other hand is more specific to the activity it has been learnt for and because of this is more robust to observation noise in the data (for normal sequences) than the more general NSSA model. The PSSA model offers a good compromise between the two models, the "mean shape" changes only when the ELL from the current mean exceeds a threshold (the current stationary model is unable to track the observations). We hope to apply the PSSA model to segment real activity sequences using PSSA and track them as part of future work. Also, we intend to perform a theoretical analysis of the stability of particle filtering under system model error, with the various models proposed in this work.

References

[1] D.G. Kendall, D. Barden, T.K. Carne, and H. Le, *Shape and Shape Theory*, John Wiley and Sons, 1999.

[2] S. Soatto and A.J. Yezzi, "Deformotion: Deforming motion, shape average and the joint registration and segmentation of images," in *European Conference on Computer Vision*, Copenhagen, Denmark, May 2002, p. III: 32 ff.

[3] N. Vaswani, A. RoyChowdhury, and R. Chellappa, "Activity recognition using the dynamics of the configuration of interacting objects," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Madison, WI, June 2003.

[4] N. Vaswani, "Change detection in partially observed nonlinear dynamic systems with unknown change parameters," in *American Control Conference (ACC)*, 2004.

[5] N.J. Gordon, D.J. Salmond, and A.F.M. Smith, "Novel approach to nonlinear/nongaussian bayesian state estimation," *IEE Proceedings-F (Radar and Signal Processing)*, pp. 140(2):107–113, 1993.

[6] A. Doucet, N. deFreitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer, 2001.

[7] N. Vaswani, *Change Detection in Stochastic Shape Dynamical Models with Applications in Activity Modeling and Abnormality Detection*, Ph.D. Thesis, ECE Dept, University of Maryland at College Park, August 2004.

[8] D. Forsyth and J. Ponce, *Computer Vision - A Modern Approach*, Prentice Hall, 2003.

[9] C.T. Zahn and R.Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Transactions on Computers*, vol. C-21, pp. 269–281, 1972.

[10] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," *European Conference on Computer Vision*, pp. 343–356, 1996.

[11] A. Srivastava, W. Mio, E. Klassen, and S. Joshi, "Analysis of planar shapes using geodesic paths on shape spaces," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pp. 372–383, March 2004.

[12] D. Adalsteinsson and J. A. Sethian, "A fast level set method for propagating interfaces," *Journal of Computational Physics*, vol. 118, pp. 269–277, 1995.

[13] I.L. Dryden and K.V. Mardia, *Statistical Shape Analysis*, John Wiley and Sons, 1998.

[14] C. Rao, A. Yilmaz, and M. Shah, "View-invariant representation and recognition of actions," *Intl. J. of Computer Vision, vol. 50, no. 2*, 2003.

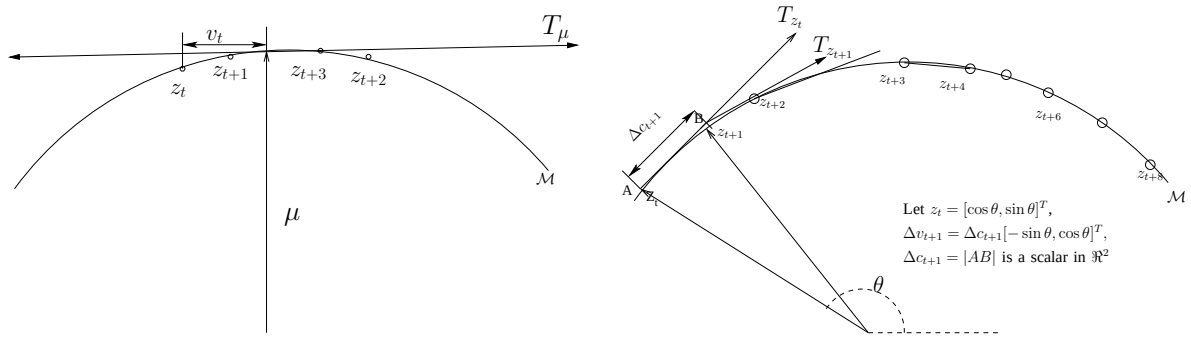
[15] V. Parmeswaran and R. Chellappa, "Action recognition based on view invariant spatio-temporal analysis," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Madison, WI, June 2003.

[16] A. Bissacco, A. Chiuso, Y. Ma, and S. Soatto, "Recognition of human gaits," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.

[17] A. Srivastava and E. Klassen, "Bayesian and geometric subspace tracking," *Adv. in Appl. Probab.*, vol. 36, no. 1, pp. 43–56, 2004.

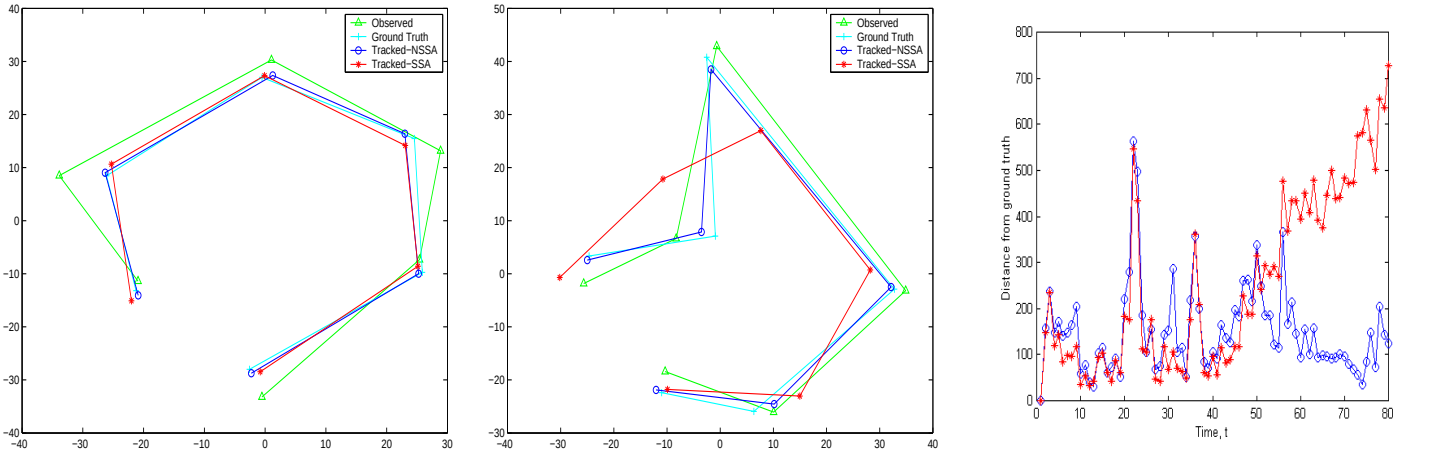
[18] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, Inc., 1991.

[19] Q Zheng and S. Der, "Moving target indication in lras3 sequences," in *5th Annual Fedlab Symposium College Park MD*, 2001.



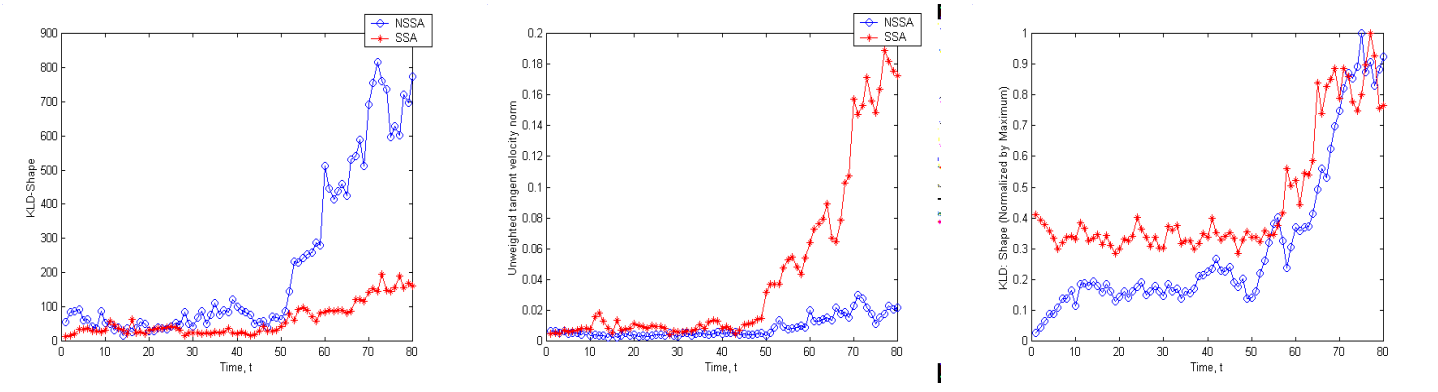
(a) Stationary Shape Activity (SSA) (b) Nonstationary Shape Activity (NSSA)

Fig. 1. SSA & NSSA depicted in $\Re^2$. $\mathcal{M}$ denotes the shape space. In (a), we show a sequence of shapes from a SSA; at all times the shapes are close to the mean shape & so the dynamics can be approximated in T_μ. In (b), we show a sequence of shapes from an NSSA, the shapes move on the shape manifold, $\mathcal{M}$, & so we need to define a new tangent space at every time instant.



(a) Normal, $\sigma^2_{obs} = 16$ (b) Abnormal, $\sigma^2_{obs} = 16$ (c) Distance from ground truth, $\sigma^2_{obs} = 81$

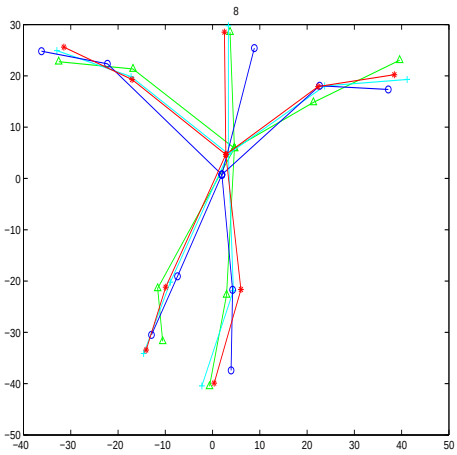
Fig. 2. Simulated shape: Abnormality introduced at $t = 40$. Tracks of normal and abnormal behavior using SSA, NSSA



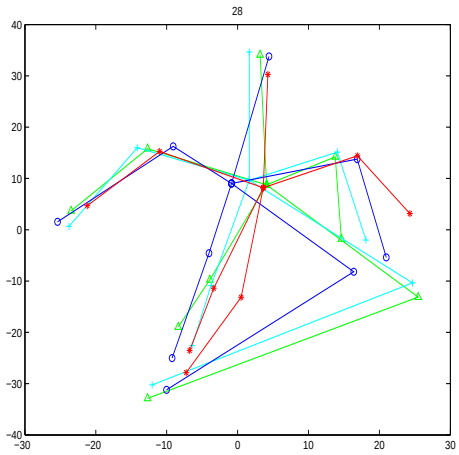
(a) ELL (Shape) (b) ELL (Shape Vel., Unweighted) (c) ELL (Shape), $\sigma^2_{obs} = 81$

Fig. 3. Simulated Shape Statistics: Abnormality introduced at $t = 40$. Note each ELL statistic plot in both (b) and (c) are normalized by their respective maximum values.

(a)The Figure Skater

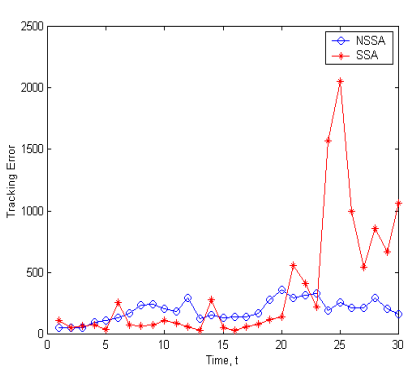


(b) Normal (SSA tracks it better)

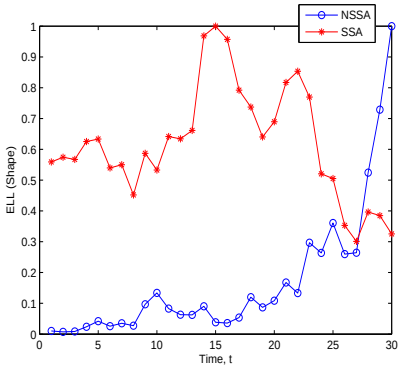


(c) Abnormal (SSA fails)

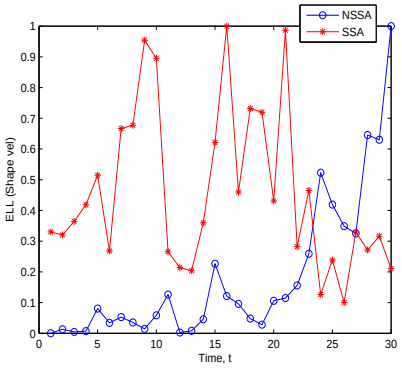
Fig. 4. Tracking the figure skater: Abnormality introduced at $t = 20$. SSA tracks the normal sequence better than NSSA. NSSA is able to track the abnormality (introduced at $t = 20$) better than SSA. Green triangles line is the observed (noisy) data, the cyan -+ line is the ground truth, the blue circles and red stars are filtered shape using NSSA and SSA respectively.



(a) Tracking Error

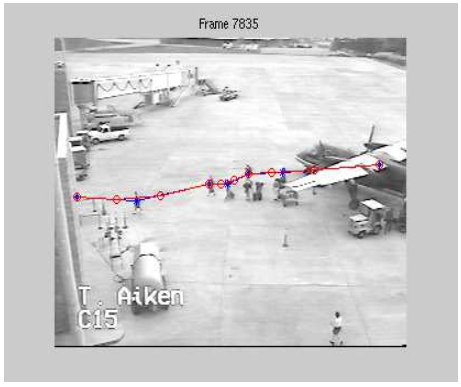


(b) ELL (Shape)

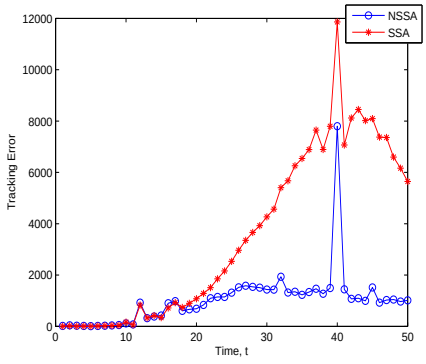


(c) ELL (Shape Velocity)

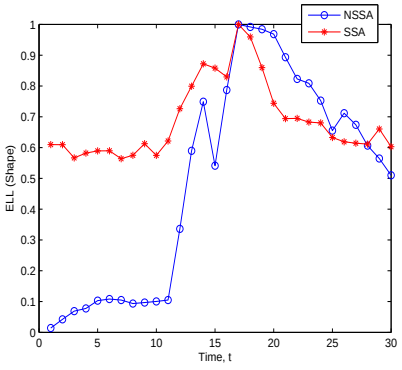
Fig. 5. Tracking the figure skater: Abnormality introduced at $t = 20$. NSSA remains in track and is able to detect using both ELL (Shape) and ELL (Shape Velocity). SSA loses track and hence is able to detect using only tracking error. Note each ELL statistic plot in both (b) and (c) are normalized by their respective maximum values.



(a) Group Activity



(b) Tracking Error, Abnormal vel.=4 (sudden)



(c) ELL (Shape), Abnormal vel.=1 (slow)

Fig. 6. Tracking activity performed by a group of people: Abnormality introduced at $t = 5$. NSSA is able to track the sudden abnormality (tracking error using NSSA shown in (b)) and also detect the slow change using ELL (shown in (c)). Here again each ELL plot is normalized by its maximum value.