STATISTICAL SHAPE THEORY FOR ACTIVITY MODELING

Namrata Vaswani, Amit Roy Chowdhury, Rama Chellappa

Center for Automation Research
Department of Electrical and Computer Engineering
University of Maryland, College Park
MD 20742, USA
{namrata,amitrc,rama}@cfar.umd.edu

ABSTRACT

Monitoring activities in a certain region from video data is an important surveillance problem today. The goal is to learn the pattern of normal activities and detect unusual ones by identifying activities that deviate appreciably from the typical ones. In this paper we propose an approach using statistical shape theory (based on Kendall's shape model) [3]. In a low resolution video each moving object is best represented as a moving point mass or particle. In this case, an activity can be defined by the interactions of all or some of these moving particles over time. We model this configuration of the particles by a polygonal shape formed from the locations of the points in a frame and the activity by the deformation of the polygons in time. These parameters are learnt for each typical activity. Given a test video sequence, an activity is classified as abnormal if the probability for the sequence (represented by the mean shape and the dynamics of the deviations), given the model is below a certain threshold. The approach gives very encouraging results in surveillance applications using a single camera and is able to identify various kinds of abnormal behaviors.

1. INTRODUCTION

Monitoring activities in a certain region from video data is an important surveillance problem. The goal is to learn the pattern of normal activities and detect unusual ones by identifying activities that deviate appreciably from the normal ones. In [1], the authors proposed building a tracking and monitoring system using a "forest of sensors" distributed around the site of interest. Their approach involved tracking objects in the site, learning typical motion and object representation parameters (e.g. size and shape) from extended observation periods and detecting unusual events in the site.

In this paper, we propose a different approach to the problem using Kendall's statistical shape theory [3]. In a low resolution video each moving object is best represented as a moving point mass or particle. In this case, an activity can be defined by the interactions of all or some of these moving particles over time. We model this configuration of the particles by a polygonal shape formed from the locations of the points in a frame and the activity by the deformations of this mean shape over time. It provides a global framework to model interactions between multiple moving objects over time. By contrast, a traditional activity modeling approach, like [1], would have to learn the motion pattern of each object and also its interaction with all other objects of interest.

Shape is defined as all the geometric information that remains when location, scale and rotational effects are filtered out [2]. Thus our shape based activity inference method would be invariant to scaling, translation or in-plane rotation of the configuration of objects (assuming orthographic projection which is valid when the camera is assumed to be at infinity).

Statistical shape theory has been an independent area of research by itself [2, 3] which began in the late 1970s and evolved into practical statistical approaches for analyzing objects using probability distributions of shape. Off late, it has been applied to some problems in image analysis. In [4], the author used statistical shape analysis for identifying landmarks on face images. Dryden and Mardia give examples of shape analysis in object recognition and image morphing [2]. All these examples, however, model the shape of a single object in static images.

Our work presents an approach for extending this method to modeling the shape formed by the locations of a group of moving objects in an image. We represent a video sequence of an 'activity' of moving objects by the average shape formed by the moving objects in any frame and deformations from it over time. In describing the motion of a deforming shape, we need to separate the effect of the global motion of the shape from its deformations. Extending Soatto's idea of static and dynamic deformable shapes [5], we define a "static shape activity" as one in which the shape formed by the moving particles remains almost constant with time (except for small deformations). In this case, there is not much information in the global motion parameters (translation, scale and rotation) and the activity can be represented by the mean "shape" and allowed range of deformations in different directions. The deformation process can be assumed to be stationary and ergodic. A "dynamic shape activity" on the other hand is represented by the pattern of global motion and/or deformation as a function of time.

Most of the kind of activities we are interested in can be modeled by the "static shape activity" description. Consider as an example, the video sequence of passengers getting out of a plane and moving towards the terminal (see figure 1 (a)). All passengers are supposed to follow the same path from the plane to the terminal. If one were to look at the shape formed by connecting the locations of all the passengers at any time instant it would look similar, except for deformations due to small variations in the path taken by each individual. We learn the mean shape of the polygon formed by the locations of the passengers in any frame. The deviations from the mean shape are projected into a linearized space about the mean shape and covariance of the deviations in this linear space is learnt.

Partially supported by the DARPA/ONR Grant N00014-02-1-0809

The dynamics of the deviations is learnt by fitting a Gauss Markov model. An abnormal activity is detected in a given test sequence of L frames if the probability for the sequence (represented by the mean shape and the dynamics of the deviations), given the model, is below a certain threshold. We are able to identify "spatial" abnormalities, e.g. deviations from the normal path, as well as "temporal" abnormalities, e.g. sudden stopping for prolonged periods of time when the normal activity should be continuous motion.

In addition to the above shape modeling technique (referred to as Kendall's shape theory), there exists a huge body of work in the vision community on shape tracking, analysis and similarity [6, 7, 8, 9, 10]. Many of these methods rely on identifying local geometric properties of the contours of the shapes.

2. SHAPE THEORY PRELIMINARIES

In this section we briefly explain the basic tools for statistical shape analysis as described by Kent and Mardia in [2] which we use in this paper. We use Kendall's representation of a shape configuration in m dimensional space as the $k \times m$ matrix formed by the locations of k landmark points on each specimen. For $m = 2$ dimensional shape a more convenient representation is a k dimensional complex vector with real and imaginary parts representing the x and y coordinates of the point.

2.1. Normalization

Pre-shape is the geometric information that remains after location and scaling information has been filtered out. Centered pre-shape (z) can be obtained by subtracting out the mean of the complex vector of landmark points (X) and scaling to norm one i.e.

$$z = \frac{CX}{||CX||}, \text{ where } C = I - \frac{1_k {1_k}^T}{k}, \tag{1}$$

I_k is a $k \times k$ identity matrix and 1_k is a k dimensional vector of ones.

2.2. Distance between shapes

A concept of distance between shapes is required to fully define the non-Euclidean shape metric space. The shape is non-Euclidean because of the scaling to norm one. The *full Procrustes distance* [2] of a complex configurations X from Y is given by the Euclidean distance between the full Procrustes fit of the preshape of X (z_X) onto the preshape of Y (z_Y).

Full Procrustes fit is chosen to minimize

$$d(Y, X) = ||z_Y - z_X \beta e^{j\theta} - (a + jb)1_k||, \tag{2}$$

where β is a scale, θ is the rotation and $(a + jb)$ is translation. Full Procrustes distance, $d_F(Y, X)$ is this minimum distance i.e. $d_F(Y, X) = \inf_{\beta,\theta,a,b} d(Y, X)$. Since the pre-shapes z_Y and z_X have already been normalized for translation and scale, the translation value that minimize $d(Y, X)$, $\hat{a} + j\hat{b} = 0$ and the scale, $\hat{\beta} = |z_X^* z_Y|$ is very close to one. The rotation angle, $\hat{\theta} = arg(z_X^* z_Y)$. (See chapter 3 of [2] for details)

For a population of similar shapes, a full Procrustes mean shape ($\hat{\mu}$) is obtained by minimizing (over μ) the sum of squares of full Procrustes distances from each shape Y_i in the population to the unknown mean shape, μ

$$\hat{\mu} = \arg\inf_{\mu} \sum_{i=1}^{n} {d_F}^2(Y_i, \mu). \tag{3}$$

It has been proved in [11] that the full Procrustes mean shape $\hat{\mu}$ can be found as the eigenvector corresponding to the largest eigenvalue of the matrix $S = \sum_{i=1}^{n} z_{Y_i} z_{Y_i}^*$. Obtaining the full Procrustes mean and aligning all shapes in the dataset to it (by finding their Procrustes fit to the mean) is known as *Generalized Procrustes Analysis*

2.3. Shape Variability in Tangent Space

To examine the structure of shape variability from the average shape, we define a linearized space (tangent space) about the mean shape and consider variance in the linearized space. The pre-shape formed by k points lies on a complex hypersphere of unit radius. The aligned pre-shapes (after generalized Procrustes analysis) of a dataset of similar shapes would lie close to each other and to their Procrustes mean on this hypersphere. Thus the tangent hyperplane at the mean is a approximate linear space to represent this dataset and so in this space, standard linear multivariate analysis techniques can be applied. In this paper, we define a linear Gauss Markov model on the time series of the tangent coordinates of the shape in consecutive frames.

The Procrustes tangent coordinates [2] of a pre-shape (z), taking the Procrustes mean as the pole for the tangent projection, are given by

$$\begin{aligned} v(z, \mu) &= [I - \mu\mu^*]\hat{\beta} e^{j\hat{\theta}} z \\ &= zz^*\mu - \mu|z^*\mu|^2. \end{aligned} \tag{4}$$

3. "SHAPE" ACTIVITY MODEL

Any activity in which all the particles are identical (have identical expected trajectories) will classify as a "static shape activity". In our experiments we have looked at the "activity" of passengers getting out of a plane and walking towards the terminal where this assumption is satisfied. Also under the identical particles assumption, the shape formed by p or $p+1$ moving particles does not look too different. Since Kendall's shape analysis methods (discussed above) are for a fixed number of points, we resample the curve connecting the points to represent it by a fixed number of points k without deforming the shape appreciably.

The x and y coordinates of each point are represented as a complex number and thus the positions of k particles form a k-dimensional complex vector. The complex vectors of shape configurations of particles at a time t are normalized as described in section 2.1. Generalized Procrustes analysis (discussed in section 2.2) on this sequence of normalized pre-shapes returns a mean shape for the sequence (using stationarity assumption here). Since deviations from mean shape are small, the normalized shapes can be projected into the tangent space (hyperplane) at the mean using equation (4).

Now, the vector of tangent coordinates is a complex k-dimensional vector. We string the real and imaginary parts of this vector to obtain a $2k$-dimensional real vector. Now since the pre-shape has been normalized for translation, it actually lies in a $2k - 2$ dimensional (real) space. Since the tangent coordinates are obtained by projecting the aligned preshapes perpendicular to the (complex) Procrustes mean shape, the dimensionality of the tangent space is actually $2k-4$ real dimensions. The rest of the analysis given below is performed in a $2k-4$ dimensional real space (which is equivalent to a $k-2$ dimensional complex space).

3.1. Dynamical Model in Tangent Space

Let the vector of tangent coordinates be represented by v_t. The origin of the tangent hyperplane is chosen to be the tangent coordinate of the mean and hence the data projected in tangent space has zero mean by construction. The time correlation between the tangent coordinates is learnt by fitting a one step *Gauss Markov model* , i.e.

$$\begin{aligned} E[v_t] &= 0 \\ v_t &= Av_{t-1} + n_t, \end{aligned} \tag{5}$$

where n_t is a zero mean i.i.d. Gaussian process and n_t is independent of v_{t-1}.

Since the activity is assumed to be stationary and ergodic, we can evaluate Σ_v for any time t as

$$\Sigma_v = E[v_t v_t^*] = \frac{1}{T}\sum_{t=1}^{T} v_t v_t^*. \tag{6}$$

Also a minimum mean square error (MMSE) estimate of A (using stationarity assumption) can be evaluated as

$$\begin{aligned} \hat{A} &= \Sigma_{v,1} * \Sigma_v^{-1} \quad \text{where} \\ \Sigma_{v,1} &= E[v_t v_{t-1}^*] = \frac{1}{T-1}\sum_{t=2}^{T} v_t v_{t-1}^*. \end{aligned} \tag{7}$$

Using $A = \hat{A}$ and ergodicity assumption, the noise covariance matrix can be calculated

$$\begin{aligned} \Sigma_n &= E[(v_t - Av_{t-1})(v_t - Av_{t-1})^*] \\ &= \frac{1}{T-1}\sum_{t=2}^{T}(v_t - Av_{t-1})(v_t - Av_{t-1})^*. \end{aligned} \tag{8}$$

Thus given a training sequence, we can use the above equations to estimate Σ_v, A and Σ_n. These parameters can then be used to test if any subsequence comes from the trained activity or not.

3.2. Testing

Using the stationary Gauss Markov model described above,

$$\begin{aligned} v_t &\sim \mathcal{N}(0, \Sigma_v), \ \forall t \\ v_{t+1}|v_t &\sim \mathcal{N}(Av_t, \Sigma_n). \end{aligned} \tag{9}$$

Thus any L length sequence (L arbitrary) of tangent projections of test data would have a jointly normal distribution with pdf

$$\begin{aligned} f(v_t, v_{t+1}, ...v_{t+L-1}) &\stackrel{(a)}{=} f(v_t)f(v_{t+1}|v_t)...f(v_{t+L-1}|v_{t+L-2}) \\ &\stackrel{(b)}{=} \frac{1}{\sqrt{(2\pi)^{2k-4}|\Sigma_v|}}\left(\frac{1}{\sqrt{(2\pi)^{2k-4}|\Sigma_n|}}\right)^{L-1} \times \\ &\exp\left(-\frac{v_t^*\Sigma_v^{-1}v_t + \sum_{\tau=t+1}^{t+L-1}(v_\tau - Av_{\tau-1})^*\Sigma_n^{-1}(v_\tau - Av_{\tau-1})}{2}\right), \end{aligned} \tag{10}$$

where (a) follows from the Markovian assumption and (b) follows from equations (9).

A given test sequence is said to be generated from the normal activity if the probability of occurrence of its tangent projections (in the tangent plane generated by the activity mean) using the above pdf is high. Thus the "distance to activity" metric for an L frame sequence d_L is

$$\begin{aligned} d_L(t+L-1) &= -log f(v_t, v_{t+1}, ...v_{t+L-1}) \\ &= v_t^*\Sigma_v^{-1}v_t + \textstyle\sum_{\tau=t+1}^{t+L-1}(v_\tau - Av_{\tau-1})^*\Sigma_n^{-1}(v_\tau - Av_{\tau-1}). \end{aligned}$$

We test for abnormality at any time t by evaluating d_L for the past L frames i.e. evaluate $d_L(t) = -log f(v_{t-L+1}, ...v_{t-1}, v_t)$.

Now, if one looks at the eigenvalues of Σ_v, there are 5-6 dimensions of "almost" zero variance (eigenvalues much smaller than the rest). One could choose these directions to represent the Approximate Null Space (ANS) of the data. If data from the same activity is projected in these dimensions it will be very close to the origin with very high probability (follows from Chebyshev's inequality [12]) while this will not happen in general for data from any outside the 'normal activity class' [13]. We use this idea to analyze tangent space data projected along the ANS using the same activity metric as defined above but applied only to the 6-dimensional ANS space data. The difference between the values of the activity metric for normal and abnormal activity is now more pronounced and computed at a reduced computational cost.

4. EXPERIMENTAL RESULTS

We use a video sequence of passengers getting out of a plane and walking towards the terminal as an example of a "static shape activity" to test our algorithm. We test the performance of the algorithm on simulated spatial and temporal abnormalities. Spatial abnormality (shown in figure 1(b)) is simulated by making one particle deviate from its original path and then move back. This simulates the case of a person deciding to not walk towards the terminal. Temporal abnormality is simulated by fixing the location of a particle thus simulating a stopped person (which can be a suspicious activity too).

Given a test sequence, at every time instant t we apply the activity metric to the past L frames with $L = 20$ i.e. $d_{20}(t) = -log f(v_{t-19}, v_{t-18}, ...v_t)$. Reducing L will detect abnormality faster but will reduce reliability.

In figure 2(a), the blue solid and dashed lines are the activity plots for two sequences of normal activity. The solid line is for the data we trained on and the dashed one is for a new 'normal' test sequence. In both these cases the metric remains below 200 (chosen as the 'normality' threshold) except between $t = 120$ to $t = 140$. The red circles plot is for the case of a temporal abnormality (one stopped particle). The particle is stopped at $t = 40$ but it takes a few frames before the contour starts deforming and some lag because $L = 20$. In figure 2(b) we compare the normal activities with the spatial abnormality. The green stars plot is for the spatial abnormality (one particle deviating from its expected path). This is also started at $t = 40$ but gets detected faster because the contour starts deforming almost immediately. Due to lack of space, we have shown the activity metric plots only for tangent vectors projected in the 6 dimensional approximate null space (ANS) of the normal activity.

(a) (b)

Fig. 1. (a): A 'normal activity' frame with 4 people, (b): Contour distorted by spatial abnormality

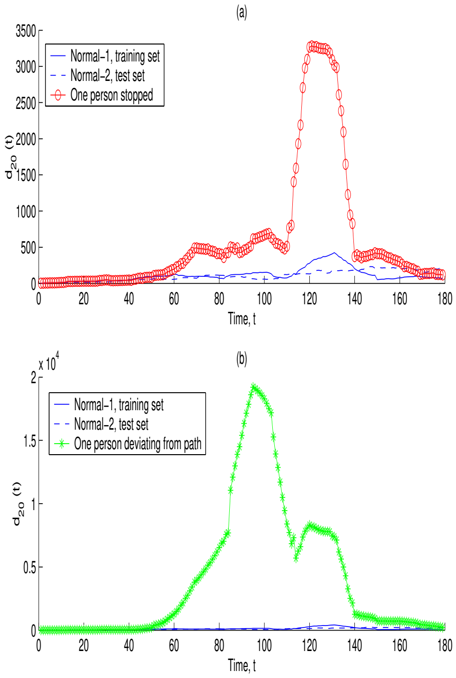

Fig. 2. Plots of the activity metric ($d_{20}(t)$) for normal and abnorm al activities: (a) compares normal activities with temporal abnormality and (b) shows the plot for spatial abnormality

5. CONCLUSION

In this paper, we have proposed a method for representing activity in low resolution video data where moving objects are modeled as point masses. We look at the shape formed by the configuration of the point objects at a given time instant and model activity by the mean shape and deformation of the mean shape over time. The deviations from mean shape are projected into a linearized space where we represent the dynamics of the activity using a stationary Gauss-Markov model. Experimental results of this method have been presented. As part of our future work, we will model the effect of observation noise and use a sequential Monte Carlo algorithm to solve the partially observed state problem.

6. REFERENCES

[1] W.E.L. Grimson, L. Lee, R. Romano, and C. Stauffer, "Using adaptive tracking to classify and monitor activities in a site," in *Conference on Computer Vision and Pattern Recognition*, 1998, pp. 22–31.

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

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

[4] I.L. Dryden, "Statistical shape analysis in high-level vision," in *IMA Workshop on Image Analysis and High Level Vision Modeling*, 2000.

[5] 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*, 2002, p. III: 32 ff.

[6] I. Cohen, N. Ayache, and P. Sulger, "Tracking points on deformable objects using curvature information," in *European Conference on Computer Vision*, 1992, pp. 458–466.

[7] D. Mumford, "Mathematical theories of shape: Do they model perception?," *SPIE*, vol. 1570, pp. 2–10, 1991.

[8] R. Basri, L. Costa, D. Geiger, and D.W. Jacobs, "Determining the similarity of deformable shapes," *Vision Research*, vol. 38, pp. 2364–2385, 1998.

[9] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models: Their training and application," *CVIU*, vol. 61, no. 1, pp. 38–59, January 1995.

[10] L. Torresani and C. Bregler, "Space-time tracking," in *ECCV02*, 2002.

[11] J.T. Kent, "The complex bingham distribution and shape analysis," in *Journal of the Royal Statistical Society, Series B*, 1994, pp. 56:285–299.

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

[13] Namrata Vaswani, "A linear classifier for gaussian class conditional distributions with unequal covariance matrices: Algorithm and analysis," in *Submitted to IEEE Trans. on Pattern Analysis and Machine Intelligence*.