SHAPE SPACE SAMPLING DISTRIBUTIONS AND THEIR IMPACT ON VISUAL TRACKING

Amit Kale and Christopher Jaynes

Ctr. for Visualization and Virtual Environments
and Dept. of Computer Science
University of Kentucky
$\{amit, jaynes\}@cs.uky.edu$

ABSTRACT

Object motions can be represented as a sequence of shape deformations and translations which can be interpretated as a sequence of points in N-dimensional shape space. These spaces range from simple 2D translations to more inclusive spaces such as the affine. In this case, tracking is the problem of inferring the most likely point in the space for the next frame given a current set of hypotheses. A robust method for achieving this is the particle filter. In this case, likely points within shape space are selected in a two step process. First, image measurements assign likelihoods to proposed points. Likely points are then propagated forward using an dynamical model to derive a set of new points that are perturbed according to some sampling distribution. These distributions play an important role in tracking performance because dynamical models are seldom known and a Gauss Markov model is often assumed for the model dynamics. This paper address the problems inherent in utilizing uninformed sampling distributions for visual tracking. We introduce a principled adaptive sampling approach that takes into account constraints on each component of the shape vector. Further a more appropriate sampling distribution that takes place in a linear subspace representing the predominant motion in the shape space. Results demonstrate improved tracking performance in challenging conditions where targets exhibit changing motion models.

1. INTRODUCTION

Visual tracking for rigid and simple non-rigid objects is the inference of the current n-dimensional shape vector for a given frame that corresponds to a position and deformation of the template being tracked. Noisy observations, occlusion, unknown motion dynamics, and several other factors confound the tracking problem. In recent years, there has been a great deal of interest in applying particle filtering to to address some of these issues. Particle filter (PF) tracking, often referred to as condensation tracking or sequential importance sampling, can provide robust tracking for stochastic systems beyond that of a traditional Kalman filter that assumes linear dynamics and Gaussian noise [1, 2, 3].

Given a template of an object to be tracked, the PF tracker maintains a set of hypotheses (particles) that define a deformation of the template from the previous frame to its current location and shape. These deformations can be represented by a *shape space* where each hypothesis corresponds to a *shape vector*. A linear parameterization of this space such as affine or Euclidean similarity is typically used in order to encompass a range of expected image motions while remaining computationally tractable. These linearly parameterized image-based models work well for rigid objects and simple non rigid ones.

In PF trackers the number of particles depends on the accuracy of the estimated model dynamics. Reducing the number of particles by estimating a finely tuned dynamic model is not trivial [4] and is sometimes even impossible. Hence a Gauss Markov model is commonly used. Perturbation of the particle set when using a Gauss-Markov model translates to adding a Gaussian random vector ($\sim \mathcal{N}(0, V)$), where V is a covariance matrix whose entries represent the knowledge about expected motion in the video sequence. Furthermore V is usually chosen to be diagonal which implies that the elements of the shape vector are independent. For mixed mode motion sequences, choosing V that supports accurate tracking is not a trivial task, since sampling variances that are appropriate for some motions may not be valid as tracking proceeds.

This work addresses some of the limitations of the Gauss Markov model in the context of visual tracking. First, we introduce an adaptive sampling technique, that analyzes the current distribution of important particles in the shape space. Principal directions within this space are detected using linear subspace analysis. Shape vectors are projected to these subspaces and further perturbations occur within these subspace as they emerge. This eliminates subjective specification of sampling covariance. Secondly, we develop sampling distributions that are dependent on the nature of the components of the shape vector. Results show that these sampling distributions better represent motion constraints that are implied in higher dimensional shape space. In addition, they are more tolerant to initial errors in the estimated parameter variances, capable of adapting to motions that may emerge in the tracking sequence, and make more efficient use of a fixed particle budget.

1.1. Related Work

Visual tracking using particle filters [1, 2] has attracted considerable interest in the computer vision community. Here we discuss a few PF algorithms that also utilize aspects of the particle space itself and do not describe the great number of research efforts that are contributing to the PF tracking paradigm.

Isard et al.[3] presented an approach (ICondensation) to combine low-level and high-level information by importance resampling with a particle filter. Sullivan [5] et al. showed that random particles can be guided by a deterministic search based on cues about each particles potential for failed tracking. Although similar to our work in that the behavior of the PF is analyzed as tracking proceeds, this approach focuses on local behavior of a small subset of the particles. In contrast, the SPF approach analyzes the entire distribution of the particles after importance sampling to yield an improved sampling covariance that is consistent for all particles.

More recently Zhou [6] proposed an adaptive appearance and velocity model. The method augments the random walk model by an adaptive velocity term which is computed iteratively given the previous particle configuration. The method described here makes use of the information contained in the entire distribution of the particles at a given frame as well as the more appropriate sampling distributions that can be derived based on constraints on the shape vector.

2. TECHNICAL DETAILS

Given a sequence of images and a template, tracking generates hypotheses about the shape deformation and translation of the template in successive frames. Typically, the deformations of the object are restricted to a lower dimensional shape space. For computational simplicity, linear parameterizations are used, examples being the six-dimensional affine space or the Euclidean similarity group.

For the sake of generality, we assume an N-dimensional shape space $\mathcal{A} = \mathcal{L}(W, T_0)$ that maps a shape space vector $X \in \mathbb{R}^N$ to a deformed template, $T \in \mathbb{R}^{N_T}$:

$$T = WX + T_0 \tag{1}$$

where W denotes a $N_T \times N$ shape matrix. The constant offset T_0 denotes the template against which shape variations are measured. Typically templates take the form of either an edge map [1] or an intensity template [2].

Two important components of PF tracking include a state evolution model $p(X_t|X_{t-1})$ and an observation model $p(Y_t|X_t)$. Given the state transition and observation models, the tracker computes the posterior density, $p(X_t|Y_{1:t})$. Particle filtering approximates this posterior density by a set of weighted particles $\{X_t^j, w_t^j\}$ with $\sum_{j=1}^{M} w_t^j = 1$. It can be shown [7] that $\{X_t^j, w_t^j\}$ is properly weighted with respect to $p(X_t|Y_{1:t})$. Given the old sample set $\{X_{t-1}^j, w_{t-1}^j\}$ a new set of particles $\{X_t'^j\}$ is generated by sampling from $p(X_t|X_{t-1})$. The likelihood of each new particle, $w_t^j = p(Y_t|X_t'^j)$ is then computed by generating a point grid according to (1) for each particle and computing a similarity measure between the template and the region indicated by the grid. Finally an importance resampling step is carried out on $\{X_t'^j, w_t^j\}$ where in particles with greater weights may be selected several times and those with low weights may not be selected at all.

In the following sections we present an adaptive sampling algorithm based on the observation that oftentimes a subspace, $\mathcal{S}$, of the N-dimensional shape space, $\mathcal{A}$, can correctly describe the motion exhibited by the object being tracked (Section 2.1). Furthermore, new sampling distributions are derived that more efficiently sample each of the shape vector components independently (Section 2.2).

2.1. Sub-space Analysis of Important Particles

The distribution of the particles following the importance sampling stage (described above) can reveal information about the true nature of the local motion model. For example, if an object exhibits pure translation, after importance sampling, most of the "important" particles are likely to lie primarily in a subspace of $\mathcal{A}$. The canonical basis for $\mathcal{A}$ is not necessarily the best representation of the underlying motion at hand. Consequently, a search region implied by V that is defined in the canonical shape space may not correctly generate "useful" particles in the sense that they may lie in a part of the space which does not conform to the observed motion. A solution to this problem is to identify the subspace $\mathcal{S}$ of the shape space $\mathcal{A}$ which accounts for most of the variance of the important particles. In addition, samples for the next perturbation phase can be drawn from this subspace using distributions along its basis vectors that may in-fact represent correlated variables in the original space. Traditional particle filtering ignores such dependencies.

Assuming that the local motion model can be described by a linear subspace of $\mathcal{A}$, Principal Components Analysis [8] can be used to discover $\mathcal{S}$. Given the M shape vectors $\{X_t^1, \cdots, X_t^M\}$, for $X \in \mathbb{R}^N$, we use PCA to compute the principal directions $\{U_t^1 \cdots, U_t^N\}$ corresponding to the eigenvalues of the scatter matrix arranged in the descending order of their magnitude. Although not all motions that may be observed can be characterized by a linear subspace of $\mathcal{A}$ (i.e. rotations), search regions provided by the new basis are more appropriate nonetheless. In Section 2.2 we introduce a more appropriate sampling distribution that can facilitate the detection of non-linear principal surfaces.

The subspace $\mathcal{S} = \{U_1, \cdots, U_n\}$ spanned by the the top $n(\leq N)$ eigenvectors is computed in each frame of the tracking sequence

$$Z_t^j = [U_1, \cdots, U_n]' \, (X_t^j - \bar{X}) \tag{2}$$

where $\bar{X}_t = \frac{1}{M} \sum_{j=1}^{M} X_t^j$

Figure 1 depicts a two-dimensional subspace that is automatically discovered in a 3D shape space involving scale and translations. This subspace consistently predominates throughout the video sequence due to the fact the vehicle being tracked is exhibits translational motion only.

Given the principal eigen vectors $[U_1, \cdots, U_n]$, the corresponding eigenvalues can indicate the choice of sampling variance to achieve more useful search regions. In particular, particles within subspace $\mathcal{S}$ are perturbed

$$Z_t'^j = Z_t^j + \text{diag}\,[f(\lambda_1), \cdots, f(\lambda_n)]\, r(n, 1) \tag{3}$$

where $r(n, 1) \sim \mathcal{N}(0, 1)$ and λ_i is the ith eigenvalue discovered in the PCA step.

An important issue in the adaptive sampling procedure is the choice of n. When motion is restricted to a subspace viz. $n < N$ it is tempting to reduce the dimensionality of the shape space. However, the algorithm must be capable of responding to changes in motion dynamics. In particular, if subspace analysis reveals that motion is occurring in a 2D subspace for some number of frames, sampling variances must not preclude discovery of motions that may emerge in the higher dimensional space.

A small sampling variance along all axes is maintained via a function $f(x)$ that bounds the variances into reasonable values and retains a residual variance for axes whose variance may be close to zero

$$f(x) = \frac{a}{1 + e^{-1.85(x-4.7)}} + b \tag{4}$$

It should be pointed out that the choice of f is not critical to the behavior of the algorithm and for results shown here f was fixed to $a = 10, b = 0.05$ resulting in variances that are bounded to 0.05-10.

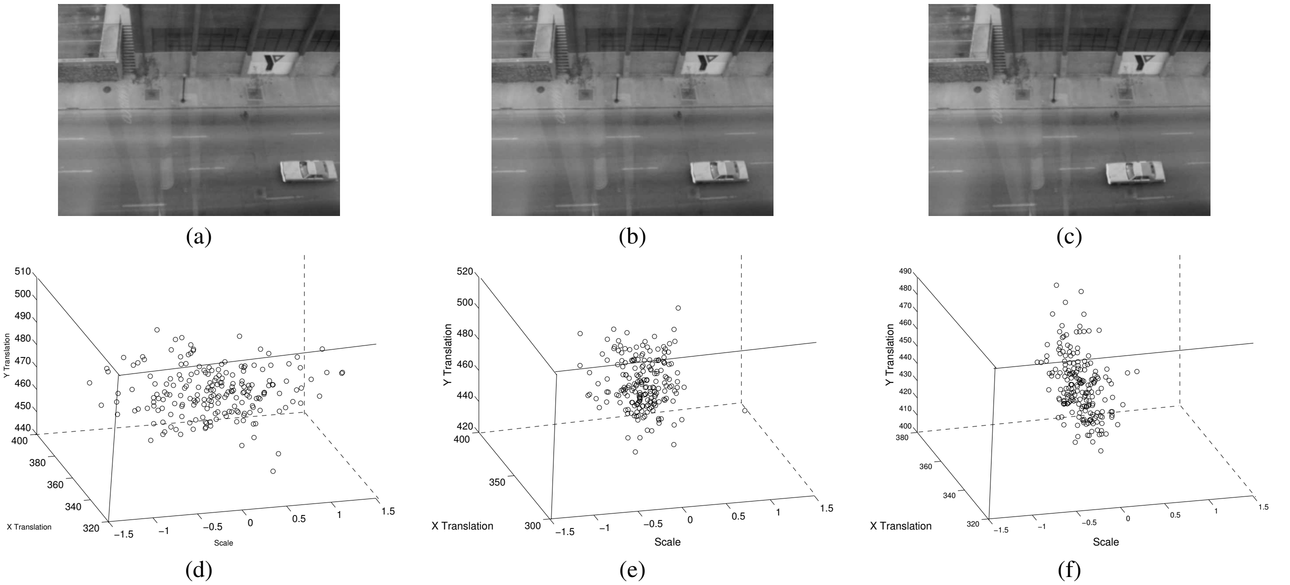


Fig. 1. Analyzing the important particle distribution for a 3D shape space for the case of pure translation. (a) (b) (c) Example frames from a vehicle tracking scenario. (d) the particle distribution utilizing the user specified covariance matrix for the frame 2 of the car sequence (e) the distribution of the particles after the importance resampling step. (f) the particle distribution using our adaptive sampling algorithm. Compare to the distribution in (e).

After perturbation in $\mathcal{S}$, the new particles, Z are re-projected to the N dimensional space

$$X_t''^j = [U_1, \cdots, U_n]\, Z_t'^j + \bar{X} \quad (5)$$

At this stage traditional likelihood computation for particles $\{X_t''^j\}$ is carried out. Following another importance sampling stage, the next iteration of the subspace algorithm is again applied. In the next section, we discuss explicit constraints on the sampling distributions that may be used in these alternative (sub)spaces.

2.2. Appropriate Sampling Distributions for Deformation Parameters

In most particle filters, an additive noise model with a spherically symmetric Gaussian distribution is used for ease of implementation. It is worthwhile examining this model in the context of visual tracking. For simplicity, let us consider a 4-D Euclidean similarity shape space that encompasses rotation, scale and translation. The shape vector in this case can be written as follows

$$X = [rcos\theta \quad rsin\theta \quad tx \quad ty]' \quad (6)$$

. While a Gaussian distribution is appropriate for translation terms this is not necessarily the case for the other shape vector components. Using functions of random variables it can be shown that the use of a Gaussian distribution for the first two components imply a Rayleigh density for r which reflects scale and a uniform density $\mathcal{U} \sim [-\pi, \pi]$ for θ. Clearly, such a choice for θ is overly general and uninformed. For example, given a particular value of θ in S this uniform density allows for perturbations that are arbitrarily far from the current θ.

A more appropriate sampling distribution should be derived from constraints on each component of the shape vector. Given the joint distribution $f_{R,\Theta}(r,\theta)$, the joint distribution of $z = rcos(\theta)$ and $w = rsin(\theta)$, can be shown to be:

$$f_{Z,W}(z,w) = \lambda \cdot \sum_{i=-\infty}^{\infty} f_{R,\Theta}\left(\sqrt{(z^2+w^2)}, tan^{-1}(\frac{w}{z}) + 2i\pi\right) \quad (7)$$

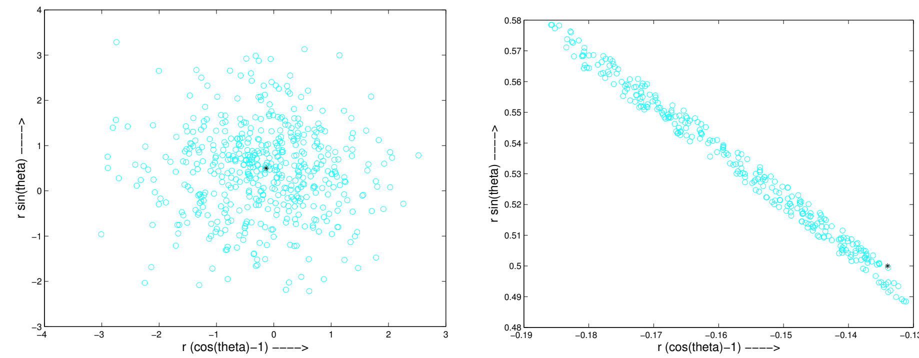
where $\lambda = \frac{1}{\sqrt{(z^2+w^2)}}$. A useful sampling density for (z,w) can be obtained by assuming that r and θ are independent. Furthermore assuming that $\theta \sim \mathcal{U}(a,b)$ where $-\pi < a < b < \pi$ the joint density may be simplified to

$$f_{Z,W}(z,w) = \frac{1}{\sqrt{(z^2+w^2)}} f_R\left(\sqrt{z^2+w^2}\right) I_{a,b}(z,w) \quad (8)$$

where $I_{a,b}(z,w)$ is an indicator function that is zero for values outside of $tan^{-1}(\frac{w}{z}) \in (a,b)$. The values a and b can be chosen to reflect how much rotation is expected to occur. Finally the remaining issue is that of $f_R(r)$. The only restriction on $f_R(r)$ is that its support should be the positive real line. For example, a simple choice would be $f_R(r) \sim \mathcal{U}(c,d)$ with $c \geq 0.$ and $d > 0$. Figure 2 (a) and (b) show the sampling distributions for the rotation components of a traditional Gaussian model versus the constrained distribution given above. The initial state is shown as a black crosshair in each image.

3. EXPERIMENTAL RESULTS

The new sampling distributions were derived as discussed in Section 2 and the dynamic subspace particle tracking framework was applied to a synthetic motion sequence in order to illuminate the difference between traditional particle filtering and the new method. Figure 3 shows tracking results for 30 frames of a template undergoing translation and rotation. The figure plots points the rotation



(a) (b)

Fig. 2. A comparison of 500 tracking particles drawn from two different distributions corresponding to rotation within shape space. (a) Traditional Gaussian distribution. Initial state shown at center of image. (b) Uniform distribution constrained to a range of $\theta \sim \mathcal{U}(-\pi/64, \pi/16)$ and $r \sim \mathcal{U}(0.9, 1)$ because this distribution is not symmetric, initial state is shown at bottom right.

components for the new method, traditional particle filtering, and ground truth.

As shown in the Figure 3, tracking performance using the same number of particles is improved. Much of the randomness contained in the MAP estimates generated by the traditional approach is eliminated. Because the new sampling distributions are more attuned to the shape vector components, a fixed particle budget can be used more judiciously. The result is both improved tracking performance and the ability to utilize a smaller number of particles. Subspace analysis is also capable of efficient sampling of the shape space. Several experiments in the vehicle tracking domain have shown significant tracking improvements including mixed mode motion sequences. Figure 1 is representative of the types of robust use of shape subspaces achieved.

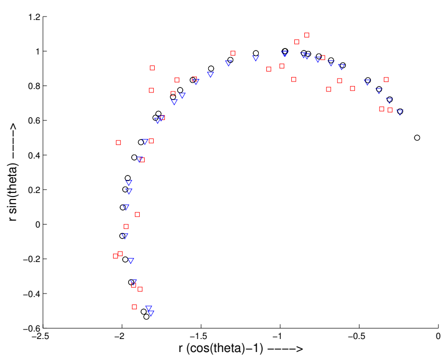


Fig. 3. Rotational tracking performance of traditional particle filter tracking (red squares) versus the new method (blue triangles) shown overlaid on ground truth rotation estimates (black circles). Points represent MAP estimate of important particles at each frame. (The shift of the origin to (-1,0) occurs on account of the particular choice of the Jacobian made by us)

4. CONCLUSIONS AND FUTURE WORK

In this paper we addressed limitations of traditional Gauss-Markov model which is commonly used in the absence of knowledge of model dynamics. First, we introduced an adaptive subspace approach which is able to utilize a limited particle budget more efficiently by sampling within linear subspaces which represent the predominant motion. Secondly, new sampling distributions based on motion constraints were introduced. These enable a more meaningful perturbation of each deformation parameter within the shape space, thereby providing more robust tracking given a limited particle budget.

In general, not all image motions can be modeled as subspaces of the shape space. For instance, rotation is more correctly represented as a principal curve in the affine shape space rather than linear subspace. In the future we wish to include principal curve analysis [9] in our adaptive sampling framework. Another open question is about the duration of observations are required before these principle manifolds can be discovered. We expect to investigate this in future work.

5. REFERENCES

[1] M. Isard and A. Blake, “Condensation - conditional density propogation for visual tracking,” *IJCV*, vol. 21, no. 1, pp. 695–709, 1998.

[2] B. Li and R. Chellappa, “A generic approach to simultaneous tracking and verification in video,” *IEEE Trans. on Image Processing*, vol. 11, no. 5, pp. 530–542, 2002.

[3] M. Isard and A. Blake, “Icondensation: Unifying lowlevel and high level tracking in a stochastic framework,” *Proc. of ECCV*, 1998.

[4] B. North, A. Blake, M. Isard, and J. Rittscher, “Learning and classification of complex dynamics,” *IEEE Trans. PAMI*, vol. 22, no. 9, pp. 1016–1034, 2000.

[5] J. Sullivan and J. Rittscher, “Guiding random particles by deterministic search,” *Proc. of ICCV*, 2001.

[6] S. Zhou, R. Chellappa, and B. Moghaddam, “Visual tracking and recognition using appearance-adaptive models in particle filters,” *IEEE Trans. on Image Processing*, November 2004.

[7] J.S Liu and R. Chen, “Sequential monte carlo methods for dynamic systems,” *JASA*, vol. 93, pp. 1032–1044, 1998.

[8] K.V Mardia, J.T. Kent, and J. M Bibby, *Multivariate Anaylsis*, Academic Press, 1979.

[9] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Verlag, 2001.