Joint Feature-Spatial-Measure Space: A New approach to Highly Efficient Probabilistic Object Tracking

Feng Chen [1] Xiaotong Yuan [2] Shutang Yang [1]
1 Department of Electronic Engineering, Shanghai Jiao Tong University
2 School of Information Security Engineering, Shanghai Jiao Tong University
1954 Huashan Road, Shanghai, 200030, P. R .China
E-mail: {ccff, yxt, styang }@sjtu.edu.cn

Abstract

In this paper we present a probabilistic framework for tracking objects based on local dynamic segmentation. We view the segmentation to be a Markov labeling process and abstract it as a MAP problem. In the Bayesian formulation, we exploit the Feature-Spatial-Measure distribution of local area as the conditional distribution. The Feature-Spatial vector is used to constrain the appearance of region while the Measure vector is used to constrain the label of the pixels in the region. One drive force to the introduction of FSM distribution is the HMMF model that makes it possible to estimate the Measure field by the minimization of a differentiable function. Mean-shift procedure and IFGT technique are used to further alleviate the computational costs. Very promising experimental results on synthetic and natural sequences are presented to illustrate the performance of the presented algorithm.

1. INTRODUCTION

Object tracking is to monitor an object's spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. This is done by solving the temporal correspondence problem. Among numerous algorithms, one popular method is appearance-based region tracking [1, 2, 5], which usually employs a statistical description of the region or the pixels to perform the tracking while ignoring the region structure. These approaches have great flexibility to track deformable and non-rigid objects as well as being robust to partial occlusion, but they all need some predefined motional models and are not self-adapted to scale change.

In this paper, we further investigate object tracking approaches based on appearance but without motion computation. Our goal is to present a new probabilistic model that permits the characterization of the solution for tracking in terms of dynamic segmentation, which is viewed to be a labeling process. The model presented in this paper is rigorously based on Bayesian estimation theory. In our formulation, we consider the feature, the feature location and the label measure vector to be probabilistic random variables. Given samples from regions representing objects and background, we estimate the Feature-Spatial-Measure (FSM) joint distribution servers as conditional distribution in Bayesian estimation. This joint distribution can be estimated using kernel density estimation. There are two problematic issues in applying the joint FSM distribution: First is the definition of label measure vector , which may determine the computational efficiency of the model; Secondly is that Feature-Spatial-Measure distribution are nonstandard in shape and can be high dimensional, therefore they require a general approaches to handle the density estimation. The method presented in this paper will address these two vital issues.

The structure of this paper is organized as follows: In section 2, we formulate the tracking problem in a probabilistic framework in the joint space and analysis its asymptotic behavior. In section 3, we adopt the Hidden Markov Measure Field to improve our model. In Section 4, we discuss the scheme for the minimization of the objective function and some computational tactics. In section 5, some experimental results and comparisons with other tracking algorithms are provided. Section 6 is the conclusions.

2. PROBABILITY TRACKING IN FSM SPACE

2.1 MAP Framework

Let (I^k) represents a sequence of images observed from the pixel lattice Ω and indexed by k. Assume that there are $M-1$ tracking regions $\{R_1^n,\ldots,R_{M-1}^n\}$ and one non-tracking region (background) R_M^n in image I^n, such that $\Omega=\bigcup_{l=1}^{M}R_l^n$; $R_i^n\cap R_j^n=\phi$, $i\neq j$. The tracking of $\{R_1^n,\ldots,R_{M-1}^n\}$ from time interval n to $n+1$ can be formulated as the problem of segmenting image I^{n+1} into $\{R_1^{n+1},\ldots,R_M^{n+1}\}$, given I^n, I^{n+1} and $\{R_1^n,\ldots,R_M^n\}$. The way to achieve this goal can naturally be regarded as a labeling process. Let l^{n+1} be the discrete label field associated with I^{n+1}. In the classical MRF model $l^{n+1}(r)\in\{1,\ldots,M\}$, denoting that pixel $r\in\Omega$ belongs to the region $R_{l^{n+1}(r)}^{n+1}$. We define the measure vector for pixels associated with l^{n+1} as $f^{n+1}(r)=(f_1^{n+1}(r),\ldots,f_M^{n+1}(r))$, where $f_k^{n+1}(r)=\delta(l^{n+1}(r)-k)$ and $\delta(x)$ is Kronecker delta function. Denote f^{n+1} the corresponding measure vector field. The goal of labeling is to maximize a posterior probability distribution, where $R^n=(R_1^n,\ldots,R_M^n)$. Through Bayesian rule, we get

$$\begin{aligned}P(f^{n+1}\mid I^{n+1},I^n,R^n)&=\frac{1}{Z}P(I^{n+1}\mid I^n,R^n,f^{n+1})P_f(f^{n+1})\\&=\frac{1}{Z}\prod_{r\in\Omega}P(I^{n+1}(r)\mid I^n,R^n,f^{n+1})P_f(f^{n+1})\end{aligned}\tag{1}$$

, where $P_f(f^{n+1})$ is the Gibbsian distribution [3], Z is normalized constants. The main challenge in the framework is the definition of conditional distribution

$$P(I^{n+1}(r)\mid I^n,R^n,f^{n+1},\theta)\tag{2}$$

2.2 FSM Distribution

We define the conditional distribution (2) in the joint Feature-Spatial-Measure space. In this space, we view the vector (u,x,f) as a multi-dimension probabilistic variable, here u is

the feature vector of a pixel (such as color, gradient, texture, et. cl.), x is the 2D coordinates and f is the discrete measure vector as defined above. We can view I^n as a "model image" that includes objects and background, while I^{n+1} as "target image" that needs to find the objects. The sample points in the model image are denoted by $I^n(r)=\{u^n(r),x^n(r),f^n(r)\}, r\in\Omega$. The sample points in the target image are denoted by $I^{n+1}(r)=\{u^{n+1}(r),x^{n+1}(r),f^{n+1}(r)\}, r\in\Omega$. The structure of the joint FSM space is generally complex and can be analysis only by nonparametric methods. We estimate (2) from the following joint FSM distribution:

$$\begin{aligned}&P(I^{n+1}(r)\,|\,I^n,R^n,f^{n+1})=P(u^{n+1}(r),x^{n+1}(r),f^{n+1}(r)\,|\,I^n,R^n,f^{n+1})\\&=\frac{1}{N}\sum_{s\in\Omega}\frac{1}{C_s}K_\sigma(u^{n+1}(r)-u^n(s))G_\tau(x^{n+1}(r)-x^n(s))T_\eta(f^{n+1}(r)-f^n(s))\end{aligned} \tag{3}$$

, where $K_\sigma(\bullet),G_\tau(\bullet),T_\eta(\bullet)$ are RBF kernel functions [2] with bandwidth parameters σ,τ,η separately, N is the total number of pixels in Ω, C_s is the number of pixels in region $R^n_{l^n(s)}$. We absorb the normalization constants into the kernels for convenience.

2.3 Asymptotic Behavior

To see the effect of changing the bandwidth of the kennel functions on the tracking formulation, we consider here the extreme case that $\sigma\to 0$, $\tau\to\infty$ and $\eta\to 0$. In this case, $G_\tau(\bullet)=c$ is a constant function and $K_\sigma(\bullet),T_\eta(\bullet)$ are Kronecker delta functions. Thus the joint probability estimate of equation (3) reduces to

$$\begin{aligned}&P(u^{n+1}(r),x^{n+1}(r),f^{n+1}(r)\,|\,I^n,R^n,f)\\&=\frac{c}{N}\sum_{s\in\Omega}\frac{1}{C_s}\delta(u^{n+1}(r)-u^n(s))\delta(f^{n+1}(r)-f^n(s))\end{aligned} \tag{4}$$

To maximize probability (4), for each pixel $r\in\Omega$, $f^{n+1}(r)$ should be set to be some $f^n(s)$, satisfying that $u^{n+1}(r)$ has the max histogram distribution in region $R^n_{l^n(s)}$. This conversed to the histogram tracking.

3 MODEL IMPROVED BY HMMF

We have defined the measure vector for pixel $r\in\Omega$ in a discrete space. Such definition is rather comprehensive but not easy for computation, since the discrete MRF optimization problem is typically solved by SA-like or EM-like algorithms with high complexity. To overcome this disadvantage, we adopt the Hidden Markov Measure Field (HMMF) model [3] to improve our model. HMMF constructs a doubly stochastic model with an additional hidden Markov random measure field. It has achieved great improvement over classical MRF model in both accuracy and computational complexity. In our tracking framework, we can use similar hidden measure vector in the FSM space. Let $p=(p_1,\ldots,p_M)$ $\sum_{i=1}^M p_i=1, p_i\ge 0$ be the hidden measure vector associated with the discrete label measure vector f. We view (u,x,p) as the probabilistic variable in FSM space. The conditional distribution (3) can be updated by just replacing f^{n+1},f^n with p^{n+1},p^n. The Gibbsian distribution $P_f(f^{n+1})$ in model (1) can be modified to be following as is discussed in [3]

$$p_p(p^{n+1})=\frac{1}{Z_p}e^{\sum_C W_C(p^{n+1})}$$

, where Z_p is the normalized constant and

$$W_C(p^{n+1})=W_{rs}(p^{n+1}(r),p^{n+1}(s))=\beta e^{-\left\|p^{n+1}(r)-p^{n+1}(s)\right\|^2} \tag{5}$$

, $<r,s>$ are neighboring sites in Ω and β is some positive constant. Here we choose the potential function $W_C(p^{n+1})$ in the form of (5), which is different from that in [3], for the purpose of applying mean-shift algorithm in the calculation. We will discuss the optimizing calculation in details in the following section.

4 TRACKING ALGORITHM

4.1 Mean-Shift Based Optimization

In this paper, we take Gaussian kernel as RBF kernel in joint FSM distribution. Take negative natural logarithm of the right hand of (1) (updated by HMMF), we obtain our energy function to be minimized:

$$\begin{aligned}E(p^{n+1})=&-\sum_{r\in\Omega}\ln\sum_{s\in\Omega}\frac{1}{C_s}e^{-\left(\left\|\frac{u^{n+1}(r)-u^n(s)}{\sigma}\right\|^2+\left\|\frac{x^{n+1}(r)-x^n(s)}{\tau}\right\|^2+\left\|\frac{p^{n+1}(r)-p^n(s)}{\eta}\right\|^2\right)}\\&-\sum_C W_C(p^{n+1})+const\end{aligned} \tag{6}$$

. To obtain the optimal estimator $(l^{n+1})^*$ for the label field, we use the following two-step procedure [3]:

Step1 Minimize the $E(p^{n+1})$ given by (5), subject to the constrains

$$\sum_{i=1}^M p_i^{n+1}=1, p_i^{n+1}{}_i\ge 0;$$

Step2 Find the mode for each measure $(p^{n+1}(r))^*$ in a decoupled way:

$$(l^{n+1})^*(r)=\arg\max_k(p_k^{n+1}(r))^*$$

We have tried Multi-scale gradient projection Newtonian descent (GPND) algorithm [3] to minimize (6). However, we found that this kind of iterative gradient descent algorithm is suffering from the parameters (such as the time interval of iteration) selection inconvenience. On the other hand, the number of iterative steps is always relatively large even if convergence is promised.

Since the energy function (6) is smooth and differentiable, and the displacement between the successive frames is small, the mean-shift algorithm [1] is a suitable candidate algorithm. The profile function of the kernels in (6) is convex and monotonic decreasing, hence the convergence can be promised according to the theorem 1 in [1]. In our framework, the mean-shift iteration procedure is as follows:

$${}^{m+1}p^{n+1}(r)=\frac{1}{2/\eta^2+2\beta\cdot\sum_{<r,s>}\rho^m}\left[\left(\frac{\frac{2}{\eta^2}\sum_{s\in\Omega}\frac{1}{C_s}p^n(s)\theta^m}{\sum_{s\in\Omega}\frac{1}{C_s}\theta^m}\right)+2\beta\sum_{<r,s>}p^n(s)\rho^m\right] \tag{7}$$

, where

$$\theta^m = e^{-\left(\left\|\frac{u^{n+1}(r)-u^n(s)}{\sigma}\right\|^2+\left\|\frac{x^{n+1}(r)-x^n(s)}{\tau}\right\|^2+\left\|\frac{{}^m p^{n+1}(r)-p^n(s)}{\eta}\right\|^2\right)}$$

, and

$$\rho^m = e^{-\left\|{}^m p^{n+1}(r)-{}^m p^{n+1}(s)\right\|^2}$$

. We initially set ${}^0 p^{n+1}(r) = p^n(r)$. As we know, the effect of the second term of (6) is to enforce the spatial coherence of $\{R_1^{n+1},\ldots,R_M^{n+1}\}$ separately. Through our experiments we have found out that such enforcement is not quite necessary since the given initiative regions $\{R_1^0,\ldots,R_M^0\}$ are always singly connected. Further more, the FSM distribution has already takes the spatial constrains into consideration. Thus, we eliminate this second term and simplifies the iteration procedure (7) to be:

$${}^{m+1} p^{n+1}(r) = \left(\frac{\sum_{s\in\Omega}\frac{1}{C_s}p^n(s)\theta^m}{\sum_{s\in\Omega}\frac{1}{C_s}\theta^m}\right) \quad (8)$$

This simplification can also be viewed to be a kind of greed algorithm and we call it as Greedy FSM (GFSM) tracking model. The procedure is rather quick, as will be shown in the experiments. To guarantee that the sum of the components of ${}^{m+1}p^{n+1}(r)$is one, we use the unification algorithm presented in [3]. In order to show how the Measure vectors evolve during the iteration, we give a simple demo that track a synthetic object in two successive frames. We focus on the pixel with coordinate (18, 29), which is highlighted in red for better visibility as shown in Fig.1. The Feature we used here is gray level. Since there is only one object for tracking, the Measure vector is 2-dimensional. Totally, the FSM variable is 5-dimensional. In the former frame, the focused pixel belongs to foreground and changes into background in the latter frame. We set the max iteration times to be 8, the Measure vector for this pixel in the latter frame changes from (1, 0) to (0.1204, 0.8796) during the iteration procedure. Fig.2 illustrates the evolvement of the 2D Measure vector.

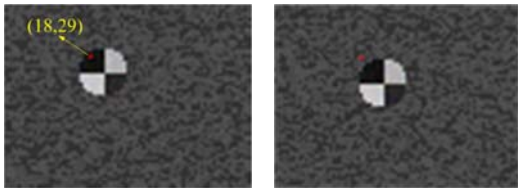


Fig.1 The red pixel is the one we focused on

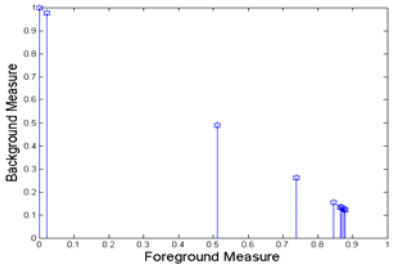


Fig.2 The 2D Measure vector evolves during the iteration

4.2 Accelerated by IFGT

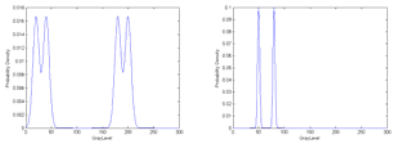
According to the expression (8) in the above algorithm, the computational complexity per frame is $O(P\cdot N^2)$, here P is the average number of iterations per frame. The algorithm will slow down in quadratic speed with the number of sample points. To further alleviate the computational cost, we apply the improved fast Gauss transform (IFGT) [4] to reduce its complexity from quadratic order to linear order $O(P\cdot N\cdot D^q)$, here D is the dimensionality of the FSM probabilistic variable the and q is the truncate order of Taylor expansions[4, 5].

4.3 Scale Adaptation

In natural tracking problems, the regions of the objects are always singly connected and the translation and deformation are small between successive frames. The GFSM model can be just applied on a rectangle area that surrounds the tracked region in current frame. These rectangles are called processing rectangles. One simple method is to generate a square with size h, centered at the centroid of the tracked foreground region. For each frame, h is updated by a certain fraction, which may be chosen as the ratio of areas of the tracked object and the processing square in the last frame, as we did in our experiments.

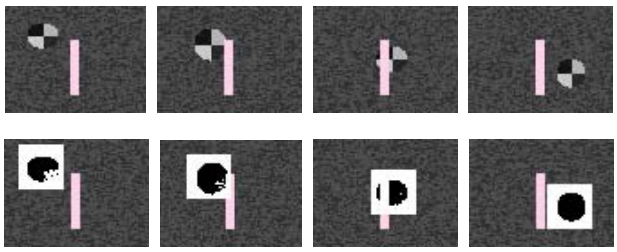
5 TRACKING EXPERIMENTS

In this section we present one synthetic and two real-world tracking examples to illustrate the performance of the proposed GFSM model and discuss some of the related issues. In the synthetic example, we generate an 80×60, 40 frames sequence. The probability distributions that generate background and objects are shown in Fig.3. To show that our model is adaptive to local deformation and scale change, we set the shape of the object changing from circle to ellipse alternatively and the radius of circles are unfixed. The partial occlusion is emulated by generating a pink bar centered at each synthetic image. The tracking results are shown in the form of local foreground/background binary segmentation image. The synthetic object has been tracked precisely despite the disturbing of partial occlusion, deformation and scale change, as is shown in Fig.4.



(a) (b)

Fig.3 (a) foreground distribution (b) background distribution



2 10 24 38

Fig 4 Track a synthetic objects with partial occlusion

The second experiment is to track a red ball in the sequence *"mobile"* (704×576, RGB, 31 frames). The highly complex scene leaves the appearance-based tracking a real challenge. In this experiment, we use a 7 dimension Feature-Spatial-Measure space (3D RGB, 2D location and 2D F/B measure vector). The constrains on the features, coordinates and measure vectors make our

tracker performs well under complex background, as is shown in Fig.5. The bandwidths are set to be (σ,τ,η)=(40,3,1). Different from the last experiment, we show the tracked foreground object as it looks like while the background all in white.

2 20 31

Fig.5 Tracking of a red ball in the sequence *"mobile"*

The third experiment is the tracking of a person in a real-world sequence captured in our laboratory（320×240 ,RGB, 323 frames）. We also adopt the 7D FSM probabilistic variable. This experiment shows the performance of our tracker in the situation that the region's appearance changed in the scene. We track the person by following his head region. As can be seen that during the first 200 frames, the man is leaving from the lens and the tracking region is the black hair. Then, after frame 200, the man turn around and moving towards the lens, the tracking region changed to be the man's face. The appearance of the tracking regions is different apparently. However, for our tracker, the FSM distribution is updated in each frame, thus the change in appearance can be learned in time. The tracking results are satisfying, as shown in Fig.6. The results are shown in the same form as in the second experiment. The bandwidths are set to be (σ,τ,η)=(60,26,1), and the average number of mean-shift iterations is 5. We use the IFGT to accelerate the calculation, the tracking speed is near real-time and the average processing rate is 14.3 fps. On the other hand, if we direct calculate the Gaussian kernel, the time cost will be unbearable (above 2 seconds for each frame). Fig. 7 gives the CPU time consumed by each frame during the tracking procedure.

122 140 160

215 236 254

280 300 310

Fig.6 Track the head of a moving person.

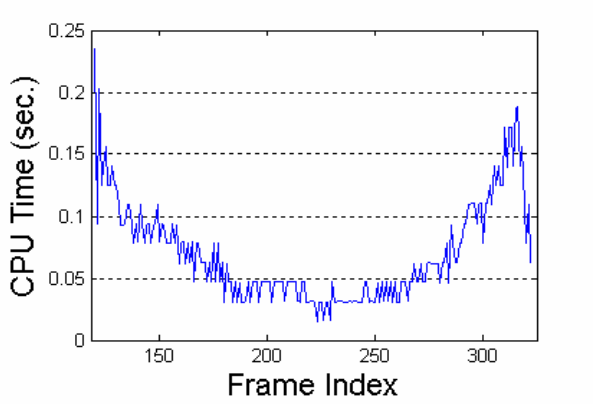


Fig.7. The CPU time for the tracking of each frame

6 CONCLUSIONS

The method presented in this paper is a general framework for tracking non-rigid objects in a sequence of images by local dynamic segmentation. In the proposed Feature-Spatial-Measure space, it is possible to track the objects that may undergoes almost any kind of movement and the tracking is robust to partial occlusion and self-adaptive to the deformation and scale change. One drive force to the FSM model is the HMMF model, which converts the discrete labeling problem to a continual optimization problem. The realization of the framework, the Greedy FSM, is highly efficient thanks to the mean-shift iteration and the IFGT technique.

7 ACKNOWLEDGEMENT

This research is supported by the the National High Technology Development 863 Program of China under Grant No. 2002AA145090. Dr. Changjiang Yang generously provides us the source code on IFGT.

REFERENCE

[1] D. Comaniciu, V. Ramesh, and P. Meer. Real-time tracking of non-rigid objets using mean shift. In Proceeding of IEEE Conference on Computer Vision and Pattern Recongition, vol. 2, pp. 142-149, Jun 2000.

[2] A. Elgammal, R. Duraiswami, L. Davis. Probabilistic tracking in joint feature-spatial spaces, Proceedings IEEE Conference of Computer Vision and Pattern Recognition, Wisconsin, Madison, Vol. 1, pp. 781 -788, Jun. 2003.

[3] J. L. Marroquin, E. A. Santana and S. Botello. Hidden Markov Measure Field Models for Image Segmentatio, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 25, no. 11, pp. 1380-1387, Nov., 2003.

[4] C. J. Yang, R. Duraiswami, N. Gumerov, and L. Davis. Improved fast Gauss transform and efficient kernel density estimation. In Proc. Int'l Conf. Computer Vision, pages 464–471, Nice, France, 2003.

[5] C. J. Yang, R. Duraiswami. A. Elgammal and L. Davis. Real-Time Kernel-Based Tracking in Joint Feature-Spatial Spaces. University of Maryland Department of Computer Science/UMIACS Technical Report CS-TR-4567, UMIACS-TR-2004-12.