http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (17): 4176-4179, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.4176.4179

Particle Probability-hypothesis-density Filter with Kernel Based State Extraction for Efficient Multi-target Visual Tracking

Wu Jing-Jing ,You Li-Hua and Cao Yi
School of Mechanical Engineering Jiangnan University, Wuxi, Jiangsu, 214122, China

Abstract: Particle probability hypothesis density (particle PHD) filter based visual trackers has attractive features of avoiding data association and the capability of solving nonlinear non-Gaussian models. But one main drawback of this approach is the unreliability of clustering technique for extracting state estimates, especially when target intersection and clutter make the distribution of particles complex multimodality. For improving the robustness and accuracy of state estimates, kernel based state extraction method is proposed for the tracker. Experimental results show the proposed method can efficiently track a variable number of objects in cluttered scene even when interactions of targets occur.

Key words: Probability hypothesis density, color histogram, clustering, kernel density estimation

INTRODUCTION

Probability Hypothesis Density (PHD) filter (Mahler, 2003) based trackers have enjoyed growing popularity in recent years. Due to the potential nonlinearity and non-Gaussiamity of visual target models, particle PHD filter (Vo *et al*., 2003) is used to implement the PHD filter. The efficiency of PHD based visual trackers relies on filtering algorithm and tracking features as most trackers, as well as observation model and state transition model etc. Among them, observation model is one of the most critical challenges for a robust tracker. The original particle PHD filter based visual tracker usually uses outputs of detectors like motion detector to establish the observation model, whose efficiency depends on the accuracy of the detections (Wang *et al*., 2008). Besides, another challenge is how to extract state estimates from the resampled particles representing the posterior intensity (i.e., the posterior PHD). Since, the intersections of multiple targets often lead to the complex multimodality distribution of the resampled particles, the classical k-means clustering algorithm may present serious degradation in state extraction performance.

In this study, to avoid inaccurate detections generating estimation errors in previous PHD based visual trackers, color histogram with position constraints (Comaniciu *et al*., 2003) is incorporated into PHD filtering framework, which incorporates the appearance model of the target with its temporal dynamics in a unifying framework. In addition, sequential kernel density approximation (Han and Davis, 2005) (SKDA) based clustering algorithm is proposed for accurate state extraction and we also refer to it as kernel based state extraction algorithm. Hence, a robust visual tracker based on particle PHD filter with robust state extraction is proposed.

PARTICLE PROBABILITY-HYPOTHESIS-DENSITY FILTER WITH KERNEL BASED STATE EXTRACTION FOR EFFICIENT MULTI-TARGET VISUAL TRACKING

In the proposed tracker, the target candidate region in an image is approximated with a w×h rectangle. Let the state of a target speed $ve_k = (ve_{x,k}, ve_{y,k})$. Assume that each target follows a linear Gaussian constant velocity model, i.e.:

$$\mathbf{x}_k = F\mathbf{x}_{k-1} + \mathbf{v}_k \quad (1)$$

where, F is the state transition matrix and v_k is the zero-mean Gaussian white process noise.

To incorporate the appearance information into the tracking framework, we incorporate the observation model designed by color histogram (Comaniciu *et al*., 2003) . Let $\{s_i\}_{i=1\ldots nh}$ be the pixel locations of the target centred at $p_k = (p_{x,k}, p_{y,k})$ and the radius of the target be h = (w, h). Define a function b: $R^2 \rightarrow \{1\ldots m\}$ associating the pixel at location s_i to the index b (s_i) of the histogram bin corresponding to the color of that pixel. The color histogram of a target candidate model $\hat{q}(p_k)$ and the probability of the feature u = 1, … ,m are defined by Eq. 2 and 3:

Corresponding Author: Wu Jing-Jing, School of Mechanical Engineering Jiangnan University, Wuxi, 214122, Jiangsu, China

$$\hat{q}(p_k)=\{\hat{q}^{(u)}\}_{u=1,\cdots,m},\ \sum_{u=1}^{m}\hat{q}^{(u)}=1 \tag{2}$$

$$\hat{q}^{(u)}(p_k)=C_h\sum_{i=1}^{n_h}k\left(\left\|\frac{p_k-s_i}{h}\right\|\right)\delta[b(s_i)-u] \tag{3}$$

where, u denotes the color histogram bins, k is a spatially weighting function and C_h is a normalisation term. Similarly, the reference target model can be represented by $\hat{q}_c=\{\hat{q}_c^{(u)}\}_{u=1,\cdots,m}$. Then the observation likelihood is defined by the similarity between a target candidate $\hat{q}(p_k)$ and the reference target model $\hat{q}_c$, i.e.:

$$p(z_k\mid x_k)=\frac{1}{\sqrt{2\pi}\sigma_c}\exp\left\{-\frac{d^2(\hat{q}(p_k),\hat{q}_c)}{2\sigma_c^2}\right\} \tag{4}$$

Where:

$$d(\hat{q}(p_k),\hat{q}_c)=\sqrt{1-\rho[\hat{q}^{(u)}(p_k),\hat{q}_c^{(u)}]}$$

is the similarity computed by Bhattacharyya coefficient:

$$\rho[\hat{q}(p_k),\hat{q}_c]=\sum_{u=1}^{m}\hat{q}^{(u)}(p_k)\hat{q}_c^{(u)}$$

and σ_c is the standard deviation of noise which is determined experimentally.

The multi-target visual tracking problem can be formulated as multi-target Bayes filter in a Random Finite Set (RFS) framework by propagating the multiple-target posterior in time. To alleviate the computation complexity in multi-target Bayes filter, Probability Hypothesis Density (PHD) filter(Mahler, 2003)is proposed to propagate the posterior intensity, i.e., a first-order statistical moment of the posterior multi-target state, by prediction and update steps. Due to nonlinear and non-Gaussian models involved in the visual tracking problem, Particle PHD filter (Vo *et al*., 2003) is proposed to implement PHD recursion by approximating the PHD with a set of random samples (weighted particles). Let posterior PHD at time k-1 $v_{k-1|k-1}(x)$ be approximated by $\{w_{k-1}^{(i)},x_{k-1}^{(i)}\}_{i=1}^{L_{k-1}}$ of L_{k-1} particles and their corresponding weights. The predicted PHD $v_{k|k-1}(x_k)$ approximated by weighted particles $\{\tilde{w}_{k|k-1}^{(i)},\tilde{x}_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ can be derived by applying importance sampling below:

$$v_{k|k-1}(x_k)=\sum_{i=1}^{L_{k-1}+J_k}\tilde{w}_{k|k-1}^{(i)}\delta_{\tilde{x}_k^{(i)}}(\tilde{x}_k) \tag{5}$$

Where:

$$\tilde{w}_{k|k-1}^{(i)}=\begin{cases}\dfrac{\phi_{k|k-1}(\tilde{x}_k^{(i)},x_{k-1}^{(i)})w_{k-1}^{(i)}}{q_k(\tilde{x}_k^{(i)}\mid x_{k-1}^{(i)},Z_k)}, & i=1,\cdots,L_{k-1}\\ \dfrac{\gamma_k(\tilde{x}_k^{(i)})}{J_k p_k(\tilde{x}_k^{(i)}\mid Z_k)}, & i=L_{k-1}+1,\cdots,L_{k-1}+J_k\end{cases} \tag{6}$$

Here, L_{k-1} particles and J_k particles are drawn from the importance function $q_k(\cdot\mid x_{k-1}^{(i)},Z_k)$ for targets at time k-1 and $p_k(\cdot\mid Z_k)$ for new targets respectively. Once the observation likelihood $p(z_k\mid x_k)$ is obtained and substitute $v_{k|k-1}(x_k)$ into Eq. 2, the weights are updated by:

$$\tilde{w}_k^{(i)}=\left[P_M(\tilde{x}^{(i)})+\sum_{z_k\in Z_k}\frac{P_D(\tilde{x}_k^{(i)})P(z_k\mid\tilde{x}_k^{(i)})}{\kappa_k(z)+C_k(z)}\right]\tilde{w}_{k|k-1}^{(i)} \tag{7}$$

where:

$$C_k(z)=\sum_{j=1}^{L_{k-1}+J_k}P_D(\tilde{x}_k^{(j)})P_k(z_k\mid\tilde{x})w_{k|k-1}^{(j)}$$

After updating, resample particles $\{w_k^{(i)},x_k^{(i)}\}_{i=1}^{L_k}$ from $\{(\tilde{w}_k^{(i)}/\hat{N}_{k|k}),\tilde{x}_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ using multinomial resampling algorithm. After the resampling step, to achieve accurate state extraction from the resampled particles $\{w_k^{(i)},x_k^{(i)}\}_{i=1}^{L_k}$, SKDA (Vo and Ma, 2006)based clustering method (i.e., kernel based method) is proposed. As a kernel density estimation method, SKDA is a fast and flexible non-parametric method to model multi-modal density, which has been successfully applied to background model and object tracking. The proposed SKDA based method first approximates the multi-modal density of the resampled particles by a weighted sum of Gaussians where each mode (i.e., a cluster of particles) is represented with a single Gaussian component and then a pruning method is employed for the component management of the Gaussian mixture, finally the peaks of the PHD defining candidate states of targets can be obtained from the final density. Here, all the parameters such as the number of modes, means, covariances and weights are automatically determined by a mean-shift algorithm.

Assume that at time step k-1, the underlying density of the resampled particles $\hat{f}_{k-1}(x)$ is a mixture of Gaussians and a Gaussian component $N(\lambda_{k-1}^l,x_{k-1}^l,P_{k-1}^l)$, with the parameter (i.e., weight, mean and covariance) and serial number l = 1,…, G_{k-1} represents a cluster. Assume that all resampled particles become part of the underlying density at time step k, first integrate the underlying density of one particle $N(\alpha,x_k^n,P_k^n)$, n = 1,..., L_k with a learning rate α and d-D state variable x_k^n into the density function $\hat{f}_{k-1}(x)$, then the initial density at time step k is:

$$\tilde{f}_k(x)=\frac{(1-\alpha)}{(2\pi)^{d/2}}\sum_{l=1}^{G_{k-1}}\frac{\lambda_{k-1}^l}{\left|P_{k-1}^l\right|^{l/2}}\exp\left(-\frac{1}{2}D^2(x,x_{k-1}^l,P_{k-1}^l)\right)+\frac{\alpha}{(2\pi)^{d/2}\left|P_k^n\right|^{l/2}}\exp\left(-\frac{1}{2}D^2(x,x_k^n,P_k^n)\right) \tag{8}$$

where, $D^2(x,x_{k-1}^l,P_{k-1}^l)=(x-x_{k-1}^l)^T(P_{k-1}^l)^{-1}(x-x_{k-1}^l)$ is the Mahalanobis distance between x and x_{k-1}^l. Then the variable bandwidth mean-shift is performed to find

convergence locations (i.e., new possible mode locations) in $\tilde{f}_k(x)$ and select the convergence locations $\{c_k^i\}_{i=1}^{N_c}$ at which no less than two points $\{x_c^j\}_{j=1}^{m}$ in $\{x_{k-1}^l\}_{l=1}^{G_{k-1}}$ and x_k^n with n = 1, …, L_k converged. If the Hessian $H(c_k^i)=(\nabla\nabla^T)\tilde{f}_k(c_k^i)$ is negative definite, then associate with the mode c_k^i a Gaussian component $N(\lambda(c_k^i),c_k^i,P(c_k^i))$ where $\lambda(c_k^i)$ is the sum of $\{x_c^j\}_{j=1}^{m}$'s weights and the covariance matrix $P(c_k^i)$ is computed by:

$$P(c_k^i)=\frac{\lambda_k^{\frac{2}{d+2}}}{|2\pi(-\hat{H}(c_k^i)^{-1})|^{\frac{1}{d+2}}}(-\hat{H}(c_k^i)^{-1}) \quad (9)$$

Now the density of $\{x_c^j\}_{j=1}^{m}$ can be represented by the new Gaussian component $N(\lambda(c_k^i),c_k^i,P(c_k^i))$. After integrations of all the resampled particles $\{w_k^{(i)},x_k^{(i)}\}_{i=1}^{L_k}$ using Eq. 8 one at each time step, updating their parameters by mean-shift and covariance estimation method in Eq. 9 and executing a component management procedure making the Gaussian components well separated, the underlying density of the resampled particles $\hat{f}_k(x)$ (i.e., the posterior intensity, PHD) will be obtained. In the Gaussian mixture representation of the posterior intensity $\hat{f}_k(x)$, extraction of multiple-target state estimates is straightforward since the means of the constituent Gaussian components are indeed the local maxima of $\hat{f}_k(x)$ and should be extracted as state estimates. From the discussed elements above, the detailed tracking process and state estimation method is shown below.

When tracking starts, target's initial state RFS is input into the proposed algorithm and extract reference models of targets using Eq. 3 at time k = 0. Then the tracking starts from time step k≥1 as follows:

- Prediction step: according to Eq. 6, for i = 1, …, L_{k-1}, draw particles $\tilde{x}_k \sim q_k(\cdot|x_{k-1}^{(i)},Z_k)$ and for i = L_{k-1}+1, …, L_{k-1}+J_k, draw particles $\tilde{x}_k \sim p_k(\cdot|Z_k)$ for new targets
- Compute observation likelihood: for i= 1, …, L_{k-1}+J_k, compute $P(z_k|x_k^i)$ using Eq. 4
- Update step: update weights $\{\tilde{w}_{k|k-1}^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ using $P(z_k|x_k^i)$ according to Eq. 7
- **Resampling:** Compute the total mass of targets:

$$\hat{N}_{k|k}=\sum_{i=1}^{L_{k-1}+J_k}\tilde{w}_k^{(i)}$$

resample $\{(\tilde{w}_k^{(i)}/\hat{N}_{k|k}),\tilde{x}_k^{(i)}\}_{i=1}^{L_{k-1}+J_k}$ to get $\{(w_k^{(i)}/\hat{N}_{k|k}),x_k^{(i)}\}_{i=1}^{L_k}$ using multinomial resampling algorithm and multiply the weights by $\hat{N}_{k|k}$ to get $\{w_k^{(i)},x_k^{(i)}\}_{i=1}^{L_k}$

- **State extraction:** Extract states from particles $\{w_k^{(i)},x_k^{(i)}\}_{i=1}^{L_k}$ using kernel based state extraction method
- Cluster the particles using SKDA based method

For n = 1, …, L_k, first integrate x_k^n into $\tilde{f}_k(x)$ using Eq. 8, then run Mean-shift algorithm to find convergence locations and select convergence locations $\{c_k^i\}_{i=1}^{N_c}$ where at least two points $\{x_c^j\}_{j=1}^{m}$ in $\{x_{k-1}^l\}_{l=1}^{G_{k-1}}$ and x_k^n converged as new mode locations.

For i = 1,…, Nc, compute the Hessian matrix H (c_k^i) and determine modes using method below

If H (c_k^i) is negative definite

Allocate a Gaussian component $N(\lambda(c_k^i),c_k^i,P(c_k^i))$ for the mode c_k^i where P (c_k^i) is computed by Eq. 9 and substitute Gaussian components located at $\{x_c^j\}_{j=1}^{m}$ with $N(\lambda(c_k^i),c_k^i,P(c_k^i))$

Else:

Left Gaussian components located at $\{x_c^j\}_{j=1}^{m}$ unchanged

End

- Combine the updated Gaussian components with the unchanged ones as the updated density $\bar{f}_k(x)$
- Remove the Gaussian components in $\bar{f}_k(x)$ with the weight $\bar{\lambda}_k^i < 0.2$ where 0.2 is set experimentally and merge similar clusters using pruning method in [6] to obtain the final density $\hat{f}_k(x)=\{N(\hat{\lambda}_k^i,\hat{x}_k^i,\hat{P}_k^i)\}_{i=1}^{N_c}$
- **State output:** Extract:

$$\hat{X}_k=\left\{\hat{x}_k^i \mid \hat{\lambda}_k^i>0.5\right\}_{i=1}^{N_c}$$

as the state estimates where 0.5 is set experimentally

EXPERIMENTAL RESULTS

The pedestrians sequence from BEHAVE dataset is used as test video. Figure 1 indicates that PHD filter based visual trackers can deal with a variable number of targets tracking problem without data association. Figure 1a presents the detections by a background subtraction detector. Figure 1b shows the particle PHD filter directly using detections as measurements (denoted as DPHD) and K-means clustering would like to generate false state estimates due to inaccurate detections such as a person detection splitting into several blobs. Figure 1c shows the particle PHD filter with observation likelihood based on color histogram and K-means clustering (denoted as KPHD) can avoid failures due to inaccurate detections but is inclined to output state estimates without satisfying accuracy for K-means clustering is incompetent for accurate assignment of particles in terms of Euclid distance especially in a scenario with occlusion. The tracking results of the proposed tracker are shown in Fig. 1d, which demonstrate that more accurate state estimates can be filtered and extracted effectively due to more accurate approximation of the density of the posterior PHD (i.e., the distribution of resampled particles).

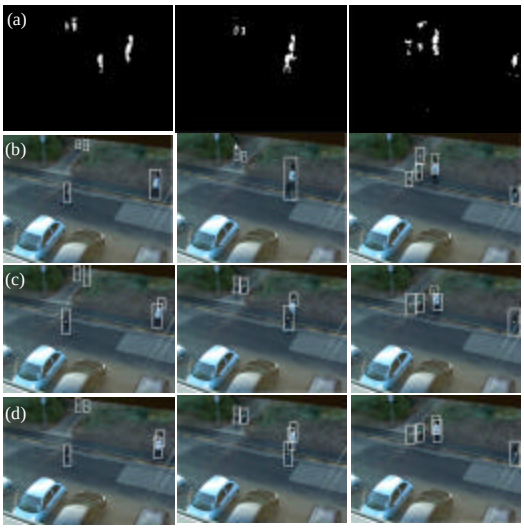

Fig. 1: Detecting and tracking results of frames23388, 23408, 23480: (a) Detections obtained by background subtraction, (b) Tracking by DPHD, (c) Tracking by KPHD, (d) Proposed method

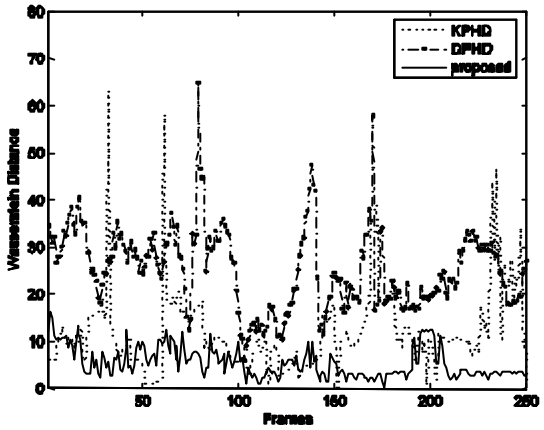

Fig. 2: Comparison results of Wasserstein distance for DPHD, KPHD and the proposed method

The Wasserstein distance (Vo and Ma, 2006) is introduced here to evaluate the tracking performance of the proposed algorithm, which aims at examining the quality of both the number and the state estimation by measuring the "distance" between two state finite sets representing the ground truth and estimates. In Figure 2, the comparison of Wasserstein distance of the three trackers in 250 frames is also provided, which demonstrates our tracker is the best.

CONCLUSION

In this study, we have presented a robust multi-target visual tracking framework based on PHD filter which stabilizes the tracker by incorporating color histograms of targets and their temporal dynamics in a unifying framework and improving the accuracy of state extraction using the proposed kernel based method. Experiments show the proposed framework can effectively track a varying number of targets with more accurate state estimates. Possible topics of future work include the incorporation of brightness data into the appearance model for more robust observation likelihood and the development of a more efficient state extraction method.

ACKNOWLEDGMENT

This study is jointly supported by the National Natural Science Foundation of China (61305016), Fundamental Research Funds for the Central Universities (Grant No. JUSRP1059) and Fundamental Research Funds for the Central Universities (Grant No. JUSRP51316B).

REFERENCES

Comaniciu, D., V. Ramesh and O. Meer, 2003. Kernel-based object tracking. IEEE Trans. Pattern Anal. Machine Intell., 25: 564-577.

Han, B. and L. Davis, 2005. On-Line density-based appearance modeling for object tracking. Proceedings of the IEEE International Conference on Computer Vision, Volume 2, October 17-21, 2005, Beijing, China, pp: 1492-1499.

Mahler, R., 2003. Multi-target Bayes filtering via first-order multi-target moments. IEEE Trans. Aerospace Electronic Syst., 39: 1152-1178.

Vo, B.N., S. Singh and A. Doucet, 2003. Sequential Monte Carlo implementation of the PHD filter for multi-target tracking. Proceedings of the 6th International Conference on Information Fusion, July 8-11, 2003, Queensland, Australia, pp: 792-799.

Vo, B.N. and W.K. Ma, 2006. The Gaussian mixture probability hypothesis density filter. IEEE Trans. Signal Process., 54: 4091-4094.

Wang, Y.D., J.K. Wu, A.A. Kassim and W. Huang, 2008. Data-driven probability hypothesis density filter for visual tracking. IEEE Trans. Circuits Syst. Video Technol., 18: 1085-1095.

Jingjing Wu received her M.S. degree in mechanical engineering from Jiangnan University, Wuxi, China, in 2007 and the Ph.D. degree in control science and engineering at Shanghai Jiao tong University in 2012, Shanghai, China. She is now with Jiangnan University. Her research interests are visual tracking, digital signal processing, Bayesian filtering, pattern recognition and multisensor data fusion.