

Region Based Sampling Method for Tracking Abrupt Motion of an Object

E Ramya^{#1} (II ME), K Manivannan^{#2} (Assistant Professor)

[#]Department of Computer Science and Engineering, PSNA College of Engineering and Technology, Dindigul, TamilNadu, India

Abstract— The problem of tracking can be defined as identifying the motion of the target in consecutive frames in the video. Conventional tracking methods fail to track abrupt motions because the motion is not smooth and uniform. A novel tracking algorithm based on Monte Carlo sampling method to deal with abrupt motions is introduced. The algorithm efficiently handles smooth and abrupt motions with the help of marginal likelihood and density-of-state term. The N-fold way WLMC method captures abrupt motions more quickly with the rejection free property of the algorithm. It helps estimate the dos value with a less number of samples without obtaining samples from the whole state space. The target tracking is guided by considering the measurement of the adjacent targets. A context aware sampling method which considers the position of the objects which are adjacent to the target helps track the target in case of occlusions. Experimental results show that the algorithm tracks the target in scenes containing complex scenarios such as occlusions, fast and abrupt changes in position and scale.

Keywords— Abrupt motion, marginal likelihood, density of states

I. INTRODUCTION

Object tracking in video plays a major role in various surveillance systems, scene analysis, medical imaging and other intelligent systems. It aids in applications such as theft findings and illegal activities. Tracking involves associating the target in consecutive frames of the video. Tracking of target occurs in situations that contains either smooth motions or both smooth and abrupt motion. The identification of target in consecutive frames of the video

is relatively easy for scenarios where the target follows only smooth changes in its motion. The target can also face some abrupt changes in its motion during the lifetime of the video. Abrupt changes can be due to a sudden change in motion such as a twist, pull or a push. The scenarios become even more complex when the frame rate is low and the scene contains edited clips from different videos. Conventional tracking methods fail to address the problem of abrupt motion tracking. They rarely consider abrupt motions and fail to track the target when abrupt changes in motion and position occur. The objective of this work is to track the target accurately even though its position and scale severely change over time.

If a tracking method focuses on improving performance in tracking smooth motions, it may miss the abrupt motions of the target. On the other hand, if the method tries to track the abrupt motions robustly, the accuracy of tracking the smooth motions may decrease. The easiest way to handle abrupt motion is to guide the tracker to search the whole state space to capture the targets uncertainty in its behavior. But it is infeasible for practical applications because searching the whole state space will result in high computational cost. In addition, most MCMC algorithms are slow in sampling since they are collect samples from the state space by choosing random path for search.

The proposed tracking algorithm states that the smooth and abrupt motions can be effectively tracked at the same time by finding two sampling values, the marginal likelihood (ML) term and the density-of-states (DOS) term. The ML term gives information about where the target object is present initially in the frame, given the object to be tracked as the input. The state space is divided into subregions and samples are collected from each region based on the histogram calculation. The region that contains the object exhibit high value and hence the target can be identified. The ML term is then calculated around the regions near the current local maximum that corresponds to the object position. This

value captures the smooth motion changes of the target. This process is called exploitation.

In addition to the calculation of ML term value, density of state term is calculated in each sub region. This DOS term calculation involves periodic collection of samples from all sub regions. Once the target initial position is identified by ML term, the dos value is calculated from other regions. So if the target makes a abrupt change in its motion, it is captured by the dos term value which indicates a clear increase in its value. By this way abrupt motions of the target can be clearly explored and intelligently tracked. Using this two terms the tracker identifies both abrupt and smooth changes.

Our method consists of two parts: First, the target object is identified in the video and tracked by a NFWL sampling algorithm. Second, we include the dynamics of adjacent targets in the tracking strategy. It can be observed that adjacent objects may affect the motion of other objects. For this, measurements of the objects neighboring the target object are integrated into the sampling process. Our proposed method comes without additional computational cost since no extra measurements are made in the algorithm.

II. RELATED WORK

Although there are several methods developed for visual tracking, there are disadvantages in conventional methods. In sampling-based tracking methods, the particle filter is used as a means to collect samples from the state space. PF shows robustness for visual tracking by handling non-Gaussianity and multimodality of a target density, and by reflecting the uncertainty of the target motion. However, the particle filter as the state space increases shows poor results and complexity.

MCMC [15] reduces the computational cost in high-dimensional state space. These methods requires large amount of samples in order to track the abrupt motion. The number of samples exponentially increases as the motion becomes more abrupt and hence the computational cost. The Adaptive MCMC algorithm unlike conventional methods automatically changes the proposal variance of MCMC as the Markov Chain goes on. This adaptive scheme is highly helpful in tracking the abrupt motion and also it is not as time consuming as conventional methods. However, this algorithm also has drawbacks. The algorithm does not provide a systematic way of escaping the local maxima and has no efficient sampling strategy to deal with large state space.

Previously, abrupt motions are tracked by increasing the noise of the motion model [6]. So if there occurs a noise it is determined that there is an abrupt motion. Increasing the noise to handle abrupt motions may sometimes cause false alarms and the method becomes inefficient. When the state space becomes large performance degradation occurs. Hamiltonian [4] dynamics are introduced to find the object's trajectory. It incorporates gradient information of the dynamic trajectories and thus suppresses the random walk nature in traditional MCMC methods. It easily escapes local maxima and helps to reach global maximum. The disadvantages are it cannot accurately track the target

when the scale of the target changes. Calculation of proposal density is susceptible to errors.

Existing system and method either track smooth or abrupt motions. Most of the algorithms cannot efficiently track both abrupt and smooth motions. Some systems tracks abrupt motions but its performance degrades when either the state space increases or when there are large number of targets in the state space. Tracking fails in conditions when the object is totally or partially occluded. MCMC algorithms are slow in sampling and get trapped in local maxima. The proposed system overcomes all the problems and efficiently tracks the object motion in case of abrupt changes and even in a high state space and in occlusions. The algorithm helps to escape local maximum and reach global maximum.

III. PROPOSED SYSTEM FRAMEWORK

The objective of video tracking is to associate target objects in consecutive video frames. The association can be especially difficult when the objects are moving fast relative to the frame rate. Another situation that increases the complexity of the problem is when the tracked object changes orientation over time. For these situations video tracking systems must employ an efficient tracking algorithm that can effectively track the target in different possible motions of the target. The main aim is to efficiently track the target object in complex scenarios which contains abrupt changes. Our proposed method involves the following steps

A. Target Object Extraction

The first step of the tracking process is to take video as input. The video is made up of image frames. In order to track the motion of a target we need to convert the video into frames and the motion of the target is analyzed between intra and inters frames. It involves identification of various objects available in the context of the frame and then selection of specific target to be tracked. Usage of Computer vision algorithms like SURF to identify key points and descriptors in the target to be tracked. Based on the algorithm the particular target whose motion needs to be tracked is identified and fed us input to the tracking algorithm which then identifies the motion of the target in the consecutive frames.

B. Calculation of acceptance ratio

Once the target object is identified in the video using SURF, the location coordinates of the target is found and the state space is divided into disjoint regions. The likelihood term is calculated around that region where the object is identified. This region will have the high likelihood value. Now if the target makes any smooth motions around the region the likelihood estimation can capture the motions. Second is the calculation of density of states term. After the likelihood estimation the dos term for each region except the region where the object is actually found is set to 1. This is similar to histogram equalization. The equalization of dos term helps to track the target robustly when the search space contains similar objects. Periodically dos value for each sub region is calculated. If the ML term has high value it states there occurs a smooth motion. If the target makes abrupt changes the histogram value gets increased in that

particular sub region and the dos term estimation in the sub region can capture the abrupt changes in motion and position. By this way the algorithm easily escapes local maximum and reaches global maximum.

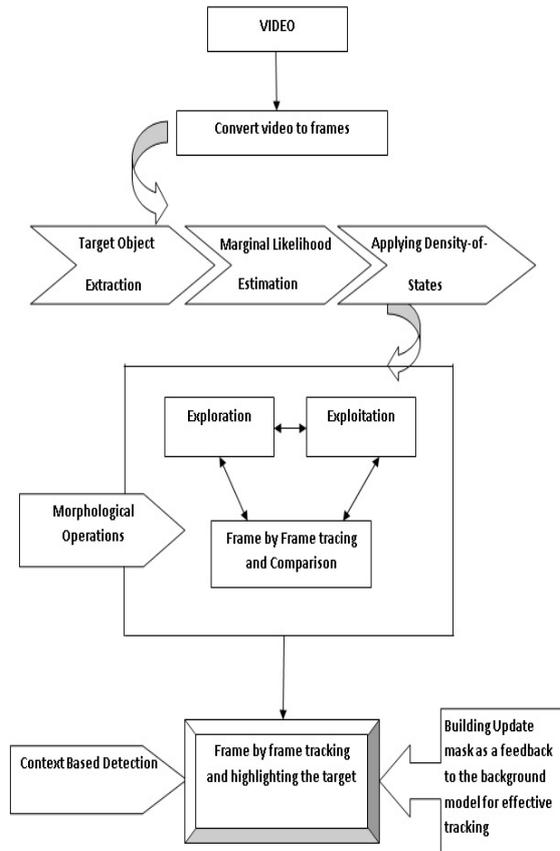


Fig. 1 System design of Region based sampling method

Frame by Frame comparison is done to see changes to the target object based on the algorithm. Algorithm for this step is discussed in the Section 4.

C. Morphological operations

Morphological operations such as noise removal are done to avoid false alarms in tracking. This includes tracked features, voted center positions with high probability, bounding boxes plus trajectories and final foreground labels. Noise removal involves repeated erosion and dilation techniques to filter the object to be tracked from the remaining objects in the context.

D. Context Based Detection

It involves identification of the key frame lists which acts as a representative of the video. The motion of one target normally influences the motion of adjacent targets. More precisely, the dynamics of the respective target are constrained by the position and movement of adjacent targets to avoid their collision. To incorporate this knowledge into tracking, we consider the measurements of objects neighboring the target during the sampling process of the current target. Thus when a target is tracked the position is the objects adjacent to the target are taken into account. So when the target undergoes partial or total occlusions, the target's future position in the frame can be guessed by the adjacent targets position. The objects

adjacent to the target changes as the target moves along its path.

E. Building Update Mask

Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. A common approach for object detection is to use information in a single frame. However, some object detection methods make use of the temporal information computed from a sequence of frames to reduce the number of false detections. This temporal information is usually in the form of frame differencing, which highlights changing regions in consecutive frames.

The knowledge about the previous frame which contains the target position in the state space before any change in motion helps us to efficiently track the algorithm. For this purpose an update mask which contains information about the previous frame is given as feedback to the tracking algorithm. To support tracking objects which stop still in the scene for a period, we can build an update mask as a feedback to the background model which stops the background model from updating the region covered by the objects that are reliably tracked.

An important element for the entire model is a reference point for calculating relative positions and matching alignment. The center point of the bounding box is vulnerable to the changes in object contour. We also propose an online sampling method to train classifiers for the position estimation and pixel labeling during occlusion. It involves construction of update mask using computer vision algorithm, sending feedback to the background model through update mask and highlighting the tracked objects.

IV. ALGORITHMS FOR THE PROPOSED MODEL

A. Tracking Method - Sampling

In order to deal with abrupt changes in both position and scale, the NFWL-based tracking method considers the whole state space S that is constructed by the Cartesian product of the state space of position S^P , and scale S^S , and divides S into d disjointed subregions. The efficiency of the algorithm comes from two properties in the sampling process. The first property is that it is a rejection-free algorithm. That is, it accepts all proposed states without the acceptance step. The second is that the method employs a more efficient scheme to estimate the DOS term in the estimation step. It tries to obtain the DOS term with a smaller number of samples as much as possible.

1) *Estimation Step*: The method does not have an acceptance step, so it should propose states with good quality. The states with good quality means the states that describe the target posterior distribution well enough. To describe it well, the states with high posterior probability should be more frequently sampled, and the states of low posterior probability should be less frequently sampled. For this, the method presents a new proposal density which consists of two parts. The first part is to choose a subregion of S , and the second is to propose a state that belongs to the subregion. When the previous state belongs to the i th sub region S_i , the method chooses the j th subregion S_j of the next state. The

method prefers to choose a new subregion that has higher ML score, lower DOS score, and lower distance from X_{t-1} compared with the previous subregion, respectively. This means that the method frequently makes a move to the subregion 1) where the target might be, 2) that is insufficiently explored, and 3) close to the subregion where the target existed at the previous frame. Using the calculation of $d(s_i, s_0) / d(s_j, s_0)$, the subregion which is close to the previous optimal state is proposed because the motion of the target is typically smooth. However, this assumption does not hold for abrupt motions. Therefore, the DOS term is additionally inserted by $g(s_i) / g(s_j)$. This term encourages proposal of the subregions that have not been explored sufficiently. Since these subregions are far from the previous optimal state, they typically contain the states of abrupt motions. Thus, due to the additional DOS term, our method can deal with the abrupt motions as well.

2) *Proposal step*: To update the DOS with a smaller number of moves as much as possible, the method calculates the lifetime of the previous sub region. Here $\tau(S_i)$ represents the lifetime value of the subregion S_i . Unlike WLMC in which the histogram of a subregion is increased by 1 whenever it obtains a sample belonging to the subregion, NFWL estimates how many times a move to other sub regions would be a rejected on average in the usual update scheme. This estimated value is the lifetime. The lifetime value reflects how many samples are continually obtained by the method at the previous subregion. By estimating this value, our method can update the histogram of the previous subregion not by one as in WLMC, but by the amount of the lifetime value in

$$h(S_i) \leftarrow h(S_i) + \tau(S_i)$$

where $h(S_i)$ indicates the histogram value of subregion S_i . The DOS score is then adaptively estimated by

$$g(S_i) \leftarrow g(S_i) * \tau(S_i) * f$$

To make the flat histogram, the sampling method should visit all subregions sufficiently by moving a subregion to other subregions. The problem of WLMC is that it needs many samples to move from a subregion to other subregions, which is called tunneling time. On the other hand, NFWL needs a small number of samples to move from a subregion to other subregions because it always accepts the moves. In this case, however, the histogram is biased due to the rejection-free property of NFWL. The lifetime value $\tau(S_i)$ compensates for the bias of the histogram by measuring how many times a move to other subregions would be rejected on average in the usual update scheme. Hence, by updating the amount of $\tau(S_i)$ at a time, NFWL quickly produces the flat histogram using a small number of samples.

3) *Annealing Step*: The method starts the process over the state space S_p and performs the WLMC-based tracking algorithm. If the histogram becomes semiflat, the method decreases S_p by selecting half the number of subregions in S_p . To choose the subregions, utilize the DOS information. S_p initially consists of d disjoint subregions. We choose the $d/2$ number of subregions according to the score in descending order because the DOS score indicates the probability of local maxima in a

subregion. Then, each chosen subregion is divided into two regions so that the total number of subregions reverts to d . If the subregions have been horizontally divided, they are then vertically divided at the next time. The annealing step is terminated when the number of iterations reaches a predefined value. The dividing strategy provides sufficient results for our tracking problem.

B. Tracking Algorithm

Algorithm Subregion sampling-based tracking method

Input: $\{X_{t-1}^i\}_{i=1}^N$
 Output: $X_t = \{ (X_t^y, X_t^x, X_t^s) \}_{i=1}^N$

Step 1: Bayesian Filtering step: The method randomly selects a sample X_{t-1}^i and moves the selected sample according to the second-order autoregressive function, and use the result as the initial state of the X_t Markov chain.

Step 2: for 1 to N do

Step 3: Proposal step: Propose a new state using

Step 4: Estimation step:

- 4: 1. Estimate the life-time .
- 4: 2. Update the histogram .
- 4: 3. Estimate the DOS.
- 4: 4. Estimate the marginal-likelihoods.

Step 5: Annealing step: Reduce the state space by the Process.

Step 6: Estimate the MAP state \hat{X}_t .

C. Context Based Sampling

The motion of a target normally influences the motion of objects adjacent to it in the state space. More precisely, the dynamics of the respective target are constrained by the position and movement of adjacent targets to avoid their collision. To incorporate this context knowledge into the particle filter, we consider the measurements of neighboring tracked targets during the sampling process of the current target. We assume that the spatial configuration of neighboring targets is constant between two consecutive video frames. Let $\hat{x}_{t-1,k}$ and $\hat{x}_{t-1,l}$ denote the estimated positions of the current target k and its neighbor l in the previous time step. This guide is determined as

$$g_{t,l}^{i+1} = x_{t,l}^i + (\hat{x}_{t-1,k} - \hat{x}_{t-1,l})$$

Propagating the current particle position by the motion model, the resulting position is defined as

$$b_{t,k}^i = f(x_{t-1,k}^i, u_{t-1,k})$$

For the final sampled particle position the guides from all neighboring targets and the motion model of the respective target are considered. The definition of neighboring targets typically includes the distance to the current target and the relative velocity in the last time step, if collision of the targets should be avoided.

V. EXPERIMENTAL RESULTS

Given the video as input and the target to be tracked, the target is identified in the frame using SURF. The video is converted into frames and the target motions are analyzed in each frame. For better understanding the target is represented by a circle around it.



Fig. 2 Input image with the target to be tracked

The tracking algorithm efficiently tracks the object when there are abrupt changes in position and scale of the target.



Fig. 3 Target tracked in low frame rate video

The algorithm incorporating the knowledge of neighboring object aids in tracking the object in case of partial and total occlusion.



Fig. 4 Target tracked in case of occlusions

Experimental results show that our tracking algorithm robustly tracks the object in complex scenarios such as abrupt changes and occlusions.

VI. CONCLUSION

The NFWL algorithm tracks the motion of the target efficiently when there are abrupt changes in the position and scale of the target. The algorithms effectively address the tracking of abrupt motions, as well as the accurate tracking of smooth motions. The algorithm tracks the target accurately even when the state space contains objects similar to the target. The dynamics of the respective target are constrained by the position and movement of adjacent targets. A context-aware sampling method, which considers the motion of adjacent targets during the particle sampling process of the current target

is introduced. This strategy improves the robustness of the tracker, and effectively tracks the target in partial and total occlusions. Accuracy is ensured with providing feedback to the system through update mask which contains information about the targets position in the previous frame.

REFERENCES

- [1] Bastian Leibe, Konrad Schindler, "Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles", IEEE transaction on Pattern Analysis and Machine Intelligence, 2008
- [2] Boris Babenko, Ming-Hsuan Yang, Serge Belongie, "Visual Tracking with Online Multiple Instance Learning", IEEE Conference on computer Vision and Pattern Recognition, 2009.
- [3] David J. Fleet, Allan D. Jepson, Thomas F. El- Maraghi, "Robust Online Appearance Models for Visual Tracking" IEEE transaction on Pattern Analysis and Machine Intelligence, Vol. 25, NO. 10, OCTOBER 2003.
- [4] Fasheng Wang, Mingyu Lu, "Hamiltonian Monte Carlo Estimator for Abrupt Motion Tracking", International Conference on Pattern Recognition, NOV 2012.
- [5] F.C. Park and J. Kwon, K.M. Lee, "Visual Tracking via Geometric Particle Filtering on the Affine Group with Optimal Importance Functions", IEEE Conf. Computer Vision and Computer Recognition, 2009.
- [6] Isabella Szotka and Matthias Butenuth, "An Adaptive Particle Filter Method for Tracking Multiple Interacting Targets", MVA2011 IAPR Conference on Machine Vision Applications, June 13-15, 2011.
- [7] Junseok Kwon and Kyoung Mu Lee, "Wang-Landau Monte Carlo-Based Tracking Methods for Abrupt Motions", IEEE Transactions on Pattern Analysis and Machine Intelligence, April 2013.
- [8] K. Smith, D.G. Perez, and J. Odobez, "Using Particles to Track Varying Numbers of Interacting People," Proc. IEEE Conf. Computer Vision and Computer Recognition, 2005.
- [9] Kyoung Mu Lee and Junseok Kwon, "Tracking of Abrupt Motion Using Wang-Landau Monte Carlo method", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008.
- [10] Robert T. Collins and Yanxi Liu, "On-Line Selection of Discriminative Tracking Features", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 27, No: 10, Oct. 2005
- [11] Takayoshi Yamashita, Yuan Li, Haizhou Ai, Shihong Lao Masato Kawade, "Tracking in Low Frame Rate Video: A Cascade Particle Filter with Discriminative Observers of Different Lifespans", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 10, Oct 2008
- [12] W. Li, X. Zhang, and W. Hu, "Contour Tracking with Abrupt Motion," IEEE International Conf. Image Processing, 2009.
- [13] X. Zhou and Y. Lu, "Abrupt motion Tracking via Adaptive Stochastic Approximation Monte Carlo Sampling", IEEE Conf. Computer Vision and Computer Recognition, 2010.
- [14] Y. Wu and G. Hua, "Multi-Scale Visual Tracking by Sequential Belief Propagation," IEEE Conf. Computer Vision and Computer Recognition, 2004.
- [15] Zia Khan, Tucker Balch, and Frank Dellaert, "MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 27, No.11, NOV 2005.