SwATrack: A Swarm Intelligence-based Abrupt Motion Tracker

Mei Kuan Lim, Chee Seng Chan
University of Malaya, Center of Image and Signal Processing
50603 Kuala Lumpur, Malaysia
imeikuan@siswa.um.edu.my, cs.chan@um.edu.my

Dorothy Monekosso
University of West England, Eng. & Maths.
Bristol BS16 1QY, United Kingdom
dorothy.monekosso@uwe.ac.uk

Paolo Remagnino
Kingston University, Comp. & Info. Sys.
Surrey KT1 2EE, United Kingdom
p.remagnino@kingston.ac.uk

Abstract

Conventional tracking solutions are not feasible in handling abrupt motion as they are based on smooth motion assumption or constrained motion model; where the motion is often governed by a fixed Gaussian distribution. Abrupt motion however, is not subjected to motion continuity and smoothness. To assuage this, we propose a novel abrupt motion tracker that is based on swarm intelligence - the SwATrack. Unlike existing swarm-based filtering methods, we firstly introduce an optimised swarm-based sampling strategy to enrich the trade-off between the exploration and exploitation of the search space in search for the optimal proposal distribution. Secondly, we propose adaptive acceleration parameters to allow on the fly tuning of the best mean and variance of the distribution for sampling. The adaptive strategy requires no training stage thus allowing flexibility in the motion model, while relaxing the number of particles deployed. Experimental results in both the quantitative and qualitative measures demonstrate the effectiveness of the proposed method in tracking abrupt motions.

1 Introduction

Visual tracking is pertinent in the tasks of motion-based recognition, automated surveillance and human-computer interaction [1]. In general, tracking is often simplified by assuming that the prior knowledge about the motion is governed by Gaussian distribution based on the Brownian or constant-velocity motion models [1, 2]. These assumptions however, do not hold true for many real world scenarios that exhibit abrupt motions such as in low frame rate videos, camera switching and fast motion. In this study, the abrupt motion refers to motion (position) that is random and changes at irregular intervals with unknown pattern; and cannot be modelled simply as the Brownian or constant-velocity motion model.

While considerable research efforts exist in relation to visual tracking, only a handful correspond to abrupt motion [3, 4, 5, 6]. Amongst them, most have been focused on online learning techniques to handle abrupt changes in the appearance instead of motion. In what constitutes the closer work to ours, Wong and Dooley [3] proposed a template-matching algorithm to track table tennis ball. Their method applied an automated two-pass segmentation method to detect the ball, followed by a Block Matching Detection (BMD) to specify the position of the ball. This detection-based strategy however, is impractical as it is highly dependent on the search window size and is computationally expensive.

Kwon and Lee [4] recently proposed the integration of Wang Landau sampling strategy into the Markov Chain Monte Carlo framework (A-WLMC) to deal with abrupt motion. The current frame is divided into N equal size sub-regions, and then the density of each sub-region is learned by the proposed algorithm to guide the state transition. In another variation, a feature-driven motion model from accelerated segment test feature matching is integrated into the particle-filtering (PF) framework by Liu et al. [5]. In [6], another adaptive sampling strategy coupled with a density-grid-based predictive model (IA-MCMC) is introduced to cope with abrupt motion. While these works have shown good results, they are still based on Monte Carlo sampling and require training stages that consequently increase the computational requirement. In short, an accurate tracking algorithm with reduced computational requirement remains a key challenge.

In summary, the contributions of this paper includes: 1) The proposal of a swarm intelligence-based abrupt motion tracker. To the best of our knowledge, there is yet to be any swarm intelligence-based method that handles abrupt motion. Most of the abrupt motion trackers are based on Monte Carlo sampling or its variations that combines Monte Carlo sampling + swarm intelligence; which are computationally expensive. 2) No learning stage is required from specific training data, thus providing flexibility while reducing the computational cost.

The rest of this paper is organised as follows. In Section 2, the motivation of the proposed work is presented, followed by a detailed explanation of the proposed algorithm in Section 3. Experimental results and discussion are given in Section 4, while Section 5 concludes this study.

2 Motivation

Visual tracking using stochastic solution is popular and involves a searching process for inferring the motion of a target known as the state, x_t from uncertain and ambiguous observations, z_t at a given time, t. Given observations, $z_{1:t-1} = \{z_1, \ldots\ z_t - 1\}$ from time *t=0* to *t=1*, the prediction stage applies the probabilistic transition model $p(x_t\,|x_{t-1})$ to predict the posterior, $p(x_t|\ z_{1:t-1})$ as Eq. 1.

To facilitate efficient tracking, in general, it is very common to simply use Brownian or constant-velocity motion models governed by Gaussian distribution with fixed mean and variance. However, these assumptions fail when the motion of the target is abrupt. In such case, tracking tend to drift from the actual position even though the appearance model is flawless. This is due to the inefficient samples which are drawn from the incorrect state space. Conventional particle-based tracking are known to suffer from *degeneracy* and *sample impoverishment* phenomenon. The former implies that a large computational effort is devoted to updating particles whose contribution to the approximation to $p(x_t|z_{1:t})$ is almost zero, while the latter leads to a loss of diversity among the particles.

Tracking with particle swarm optimisation (PSO) however, assuage the *degeneracy* and *sample impoverishment* phenomenon by optimising the search for the optimal distribution without any assumptions or prior knowledge on the target's motion. In general, PSO tracking involves propagating a swarm of particles over the image at random with the aim of searching for the best-fit search window or proposal distribution of a target. The state and velocity of the ith particle are updated by

$$p(x_t \mid z_{1:t-1}) = \int p(x_t \mid x_{1:t-1})p(x_{t-1} \mid z_{1:t-1})dx_{t-1} \tag{1}$$

$$x^{i,j}_{t+1} = x^{i,j}_t + v^{i,j}_{t+1} \tag{2}$$

$$\begin{aligned} v^{i,j}_{t+1} = (w * v^{i,j}_t) + (c_1 * r_1 * (pBest^{i,j}_t - x^{i,j}_t)) \\ +(c_2 * r_2 * (gBest_t - x^{i,j}_t)) \end{aligned} \tag{3}$$

where v is the velocity of the particle, ω, r_1, r_2, c_1 and c_2 are the acceleration constants, $pBest$ is the best position from the particle's individual search history and $gBest$ is the swarm's best solution which is also the optimal position. The cooperative interaction and information exchange between particles in PSO offers an organised way to escape the local maxima and achieve the global maximum. A known drawback of the conventional PSO is its sensitivity to the parameters settings and the lack of a reasonable mechanism to control the acceleration parameters. This, when is applied into tracking abrupt motion will constrain the exploration and exploitation of the search space. Thus, the full potential of the PSO algorithm has not been utilised.

3 Proposed Tracking - SwATrack

The proposed SwATrack is a variant of the conventional PSO to track target with abrupt motion. SwATrack is composed of two main components, i) appearance model that is usually the visual appearance cue and ii) velocity, v - we denote this as the motion model that describe the evolution of the state with time. In this paper, the appearance model is fixed - the HSV colour histogram and we focus on the study of a novel representation of the motion model that copes with abrupt motion efficiently.

3.1 Dynamic Model

The conventional PSO velocity estimation, v where we denote as the motion model in Eq. 3 is still subjected to a constraint search space and may fail to cope with some degree of abrupt motion. This is due to the fixed constant acceleration variables that require prior fine-tuning. In addition, we discovered that there is statistical relationship between these parameters. Thus, to ease self-tuned of these parameters by observing the quality of estimation, we introduced a novel motion model, v':

$$\begin{aligned} v'^{\,i,j}_{t+1} = E^{i,j}_t(w * v'^{\,i,j}_t) + (c_1 * r_1 * (pBest^{i,j}_t - x^{i,j}_t)) \\ +(c_2 * r_2 * (gBest_t - x^{i,j}_t)) \end{aligned} \tag{4}$$

where E is the exploration rate and c, r, ω are adaptive parameters with condition $p(w \cap c_1 \cap c_2) = 1$.

3.1.1 Exploration rate, E

We define the exploration rate, E as the parameters that adaptively i) increase the *exploration* with high variance and ii) increase the *exploitation* with low variance.

At every iteration, the quality of estimation for each particle is evaluated by its corresponding fitness value $f(x^i_t)$. $f(x^i_t) \to 1$ indicates high likelihood whereas $f(x^i_t) \to 0$ indicates low likelihood or no similarity between an estimation and target. Thus, if $f(x^i_t) \leq \mathrm{T}_{MinF}$, E is increased along with the maximum number of iterations, T by empirically determined stepsizes m *and* n respectively. This drives the swarm of particles to explore the region beyond the current local maxima, when the target cannot be tracked within a smaller region specified earlier (increase exploration). In contrary, when the particle search quality improves, $(f(x^i_t) \leq \mathrm{T}_{MinF})$, E is decreased alongside T, constraining the search around the current local maximum (exploitation). By utilising these exploitation and exploration abilities [7], our method is capable to track abrupt and smooth motions accurately and robustly; since there is no fixed assumption on the region for particles sampling.

3.1.2 Adaptive Acceleration Parameters

A drawback of the conventional PSO is the lack of a reasonable mechanism to control the acceleration parameters (ω, c and r); which are constant. This limits the search space and therefore could not cope with abrupt motion. To overcome this, we propose a mechanism to self-tune the acceleration parameters by utilising the velocity information of the particles. Firstly, we normalised the acceleration parameters so that they can be compared fairly with respect to the estimated velocity: $p(w \cap c_1 \cap c_2) = 1$.

The motion of the target throughout K number of frames is analysed to self-tune the settings of the inertia, w, cognitive, c_1, and social weight, c_2. When an object moves consistently in a particular direction, the inertia, w and cognitive weight, c_1 values are increased to allow resistance of any changes in its state of motion in the later frames. Otherwise, the social weight c_2 is increased by a stepsize to reduce its resistance to motion changes. The increase of the social weight allows global influence and exploration of the search space, which is relevant when the motion of a target is dynamic. Finally, the positions of the particles, $X^{i,j}_{t+1}(x^{i,j}_{t+1}, y^{i,j}_{t+1}, w^{i,j}_{t+1}, h^{i,j}_{t+1})$, are updated according to Eq. 2 until the swarm reaches convergence.

4 Experimental Results

This section experimentally validates our approach, showing that SwATrack can consistently track a target that moves with abrupt motion. Here, the tracking implementation and evaluation criteria are discussed. We evaluate the performance of the proposed method using three public table tennis dataset. Two separate comparisons are performed; i) comparison with the two state-of-the-art methods, a variation of template-matching method, BMD [3] and PF as proposed in [8]; and (ii) comparison with conventional PSO tracking (where the acceleration parameters are of fixed constants). Note that the parameters in BDM and PFs are optimised to obtain the best results for comparison. In addition, we validate our tracking results qualitatively using three videos obtained from Youtube and the benchmarked *Tennis* sequence [4, 6].

We initialise manually the 2D position of the target to be tracked in the first frame, and define a 15x15 pixels patch around the 2D position of the target as the reference frame. We employ colour histogram as our appearance model. The quality of the estimation is measured by its fitness value, which is represented by Bhattacharyya coefficient. Here, the initial values for SwATrack are $E{=}25$, $T_{MinF}{=}0.6$, $w{=}0.4$, $c_1{=}0.3$, $c_2{=}0.3$, $T{=}10$, $m{=}0.1$ and $n{=}0.1$, respectively. Note that the choices of the initial values are not as crucial, given the adaptive tuning of these parameters in the proposed method.

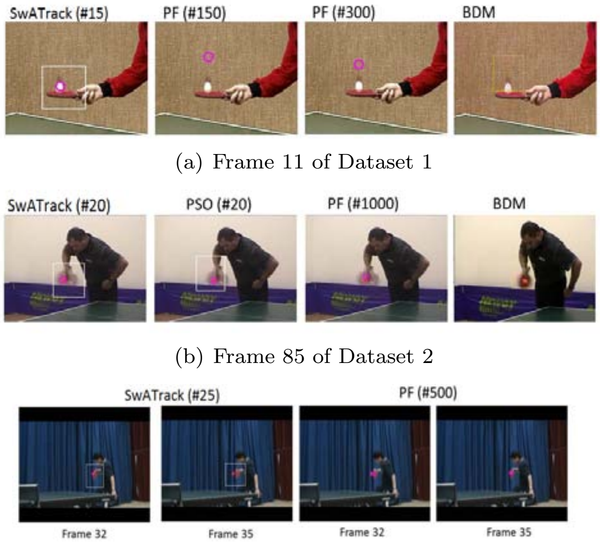
4.1 Performance Evaluation

We quantify the accuracy using detection rate. A detection is deemed correct if both of the criteria are met: Firstly, the same identity should be given to the corresponding object and secondly, the $F-measure$, F_j of jth object is > 0.5. Otherwise, tracking is classified as incorrect. The detection rate indicates the ratio between the number of correctly detected frames and the total number of frames in which the object appears in the scene.

4.2 Quantitative Analysis

A set of 3 test sequences that are obtained from table tennis game [1],[2],[3] were used for evaluation. Besides the abrupt motion of the ball, these sequences also exhibit (i) highly textured background; (ii) occlusions; (iii) small target size - about 8x8 to 15x15 pixels for an image resolution of 352x240 and (iv) low frame rate.

As shown in Table 1, the proposed algorithm outperforms the conventional PSO, PF and BDM even though the scale of ball becomes very small with minor occlusion. This is because existing trackers often fail to track the table tennis ball accurately when it exhibits



(a) Frame 11 of Dataset 1

(b) Frame 85 of Dataset 2

(c) Frame 32 and 35 of Dataset 3

Figure 1. A comparison between our proposed method - SwATrack, PF, PSO and BDM on three publicly available datasets.

	SwATrack	PSO	PF [8]	BDM [3]
Dataset 1[1]				
Accuracy(%)	**92.1**	86.4	82.0	81.3
Time (ms)	**20.9**	42.2	92.6	100.0
Dataset 2[2]				
Accuracy(%)	**98.8**	95.0	86.3	62.5
Time (ms)	**22.8**	30.6	90.8	105.0
Dataset 3[3]				
# particles	15	15	1000	N\A
Accuracy(%)	**98.4**	95.5	80.3	**98.4**
Time (ms)	**19.5**	35.8	62.8	100.0

Table 1. Comparison of the tracking results between SwATrack, PSO, PF and BDM

sudden change in motion as shown in Fig. 1. Although an increase in the number of particles in PF and an increase in the search space size in BDM would often increase their tracking accuracy, this is however, infeasible in practice. This is due to the large search space of the object state, which will lead to extremely expensive computational cost. In another alternative, an accurate motion model can be used to estimate the search space. However, an accurate motion model needed to be learned from a set of training data and thus does not cope well with unknown motion. In comparison, our proposed SwATrack does not make assumption on the motion model. Instead, the motion model is optimised at each frame by utilising the interactions between particles in a swarm for a more accurate tracking of abrupt motion. In addition, the proposed SwATrack requires minimal number of particles; about 10-20% of the samples used in PF. When the ball exhibits sudden change in its motion, the proposal distribution changes drastically and thus leads to inaccurate tracking when samples are drawn using the presumed Gaussian model. In PF, resampling from the incorrect distribution over a number of frames propagates the error and thus the

[1] SIF data from http://www.sfu.ca/ ibajic/datasets.html

[2] ITTF training dataset from `http://www.ittf.com/committees/umpires_referees/video/training/index.html`

[3] Xgmt open dataset from `http://xgmt.open.ac.uk/table_tennis`

Figure 2. Sample shots of SwATrack results.

(a) PF

(b) A-WLMC

(c) IA-MCMC

(d) SwATrack

Figure 3. A comparison between SwATrack, PF, A-WLCM[4] and IA-MCMC[6].

estimation is trapped in local optima. Meanwhile, the conventional PSO performs better tracking than the PF and BDM methods. However, the acceleration parameters are not fully utilised in PSO, making it more time-consuming and inconsistent as compared to the proposed method.

4.3 Qualitative Analysis

Further evaluation on SwATrack on videos obtained from Youtube is as depicted in Fig. 2. **(i) Abrupt motion:** The first and second sequences aim to track a synthetic ball which moves randomly, and a soccer ball which is being juggled in a free-style manner with highly textured background. It is observed that SwATrack is able to track the targets accurately. **(ii) Multiple targets:** The third sequence demonstrates the capability of the proposed system to track multiple targets; two simulated balls moving at random. Most of the existing solutions are focused on single target. Finally, **(iii) Low-frame-rate video:** The fourth sequence aims to track a tennis player in a low-frame rate video, which is down-sampled from a 700 frames sequence by keeping one frame in every 20 frames. Here, the target (player) exhibits frequent abrupt changes which violate the smooth motion and constant velocity assumptions. Thus, motion that is governed by Gaussian distribution based on the Brownian or constant-velocity motion models will not work in this case. Fig. 3 shows sample shots to compare the performance between conventional PF tracking (500 samples), A-WLMC (600 samples)[4], IA-MCMC (300 samples)[6] and SwATrack (50 samples). It is observed that the tracking accuracy of SwATrack is better than PF and A-WLMC even by using fewer samples. While the performance of SwATrack is comparable to IA-MCMC, SwATrack requires fewer samples and thus requires less processing requirement. These results further verify that the proposed method is able to track the moving targets accurately and effectively, regardless of the variety of change in the target's motion.

5 Conclusion

We presented a novel swarm intelligence-based tracker for visual tracking that copes with abrupt motion efficiently. The proposed SwATrack optimised the search for the optimal distribution without making assumptions or need to learn the motion model beforehand. In addition, we introduced an adaptive mechanism that detects and responds to changes in the search environment to allow on the fly tuning of the parameters for a more accurate and effective tracking. Experimental results show that the proposed algorithm improves the accuracy of tracking while significantly reduces the computational overheads, since it requires less than 20% of the samples used by PF. In future, we would like to further investigate the robustness of the proposed method as well as its behaviour change with the different parameter settings and sampling strategy.

References

[1] H. Yang, L. Shao, F. Zheng, L. Wang, and Z. Song, "Recent advances and trends in visual tracking: A review," *Neurocomp.*, vol. 74, pp. 3823–3831, 2011.

[2] C. García Cifuentes, M. Sturzel, F. Jurie, and G. J. Brostow, "Motion models that only work sometimes," in *BMVC*, 2012.

[3] K. C. P. Wong and L. S. Dooley, "Tracking table tennis balls in real match scenes for umpiring applications," *BJMCS*, vol. 1(4), pp. 228–241, 2011.

[4] J. Kwon and K. Lee, "Wang-landau monte carlo-based tracking methods for abrupt motions." *PAMI*, 2012.

[5] Y. Liu, S. Lai, B. Wang, M. Zhang, and W. Wang, "Feature-driven motion model-based particle-filter tracking method with abrupt motion handling," *Opt. Eng.*, vol. 51(4), 2012.

[6] X. Zhou, Y. Lu, J. Lu, and J. Zhou, "Abrupt motion tracking via intensively adaptive markov-chain monte carlo sampling," *TIP*, vol. 21, pp. 789 –801, 2012.

[7] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: a survey," *JAIR*, vol. 4, pp. 237–285, 1996.

[8] F. Yan, W. Christmas, and J. Kittler, "A tennis ball tracking algorithm for automatic annotation of tennis match," *Sig. Proc.*, pp. 619–628, 2005.