

UNIVERSITY OF TECHNOLOGY SYDNEY
Faculty of Engineering and Information Technology

Research on Object Tracking Technology for Orderless and
Blurred Movement under Complex Scenes

by

Manna Dai

A Thesis Submitted
in Fulfillment of the
Requirements for the Degree

Doctor of Philosophy

Sydney, Australia

2019

Certificate of Original Authorship

I, Manna Dai declare that this thesis, is submitted in fulfilment of the requirements for the award of PhD, in the Faculty of Engineering and Information Technology at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This thesis is the result of a research candidature conducted with another University as part of a collaborative Doctoral degree.

This document has not been submitted for qualifications at any other academic institution. This research is supported by the Australian Government Research Training Program.

Signature:

Production Note:
Signature removed prior to publication.

Date: 7 July 2019

© Copyright 2019 Manna Dai

ABSTRACT

Research on Object Tracking Technology for Orderless and Blurred Movement
under Complex Scenes

by

Manna Dai

Visual tracking is widely found in anomaly behaviour detection, self-driving, virtual reality. Recent researches reported that classic methods, including the Tracking-Learning-Detection method, the Particle Filter and the mean shift, were surpassed by deep learning in accuracy and correlation filtering in speed. However, correlation filtering can be affected by boundary effects. The conventional correlation filtering fixes the size of its detection window. When its detection window only captures partial target images due to large and sudden scale variations, the correlation filtering fails to locate the tracked target. When the target is undergoing violent shaking, motion blurs and orderless movements appear along with it. The conventional correlation filtering locks itself in the previous position of the target, and hence, the target is out of the sight of the correlation filtering. In this case, the correlation filtering drifts or fails to track. Therefore, this thesis topic is to track single-objects under complex scenes with attributes of motion blurs, orderless motions and scale variations. The main research innovation is listed as follows.

(1) An approach for searching orderless movements is designed in a generative-discriminative tracking model. To address the uncertain orderless movements, a coarse-to-fine tracking framework is adopted. A spatio-temporal correlation is learned for the detection in the subsequent frames. Experiments are conducted on public databases with orderless motion attributes to validate the robustness of the proposed approach.

(2) A template matching method is proposed for tracking objects with motion

blurs. An effective target motion model is designed to provide supplementary appearance features. A robust similarity measure is proposed to address the outliers caused by motion blurs. Our approach outperforms other approaches in a public benchmark database with motion blurs.

(3) An ensemble framework is designed to tackle scale variations. The scale of a target is estimated based on the Gaussian Particle Filtering. A high-confidence strategy is used to validate the reliability of tracking results. Our approach with hand-crafted or CNN features outperforms the methods based on correlation filtering and deep learning in databases with scale variations.

To sum up, this thesis addresses boundary effects, model drifts, fixed search windows and easily interfered hand-crafted features of objects. Different trackers are proposed for tracking single-objects with orderless movements, motion blurs and scale variations. As future work, our methods can be extended to using a neural network to further improve single-object tracking models.

Dissertation directed by Professors Xiangjian He, Dadong Wang and Shuying Cheng

Statement Indicating the Format of Thesis

The format of this thesis is Conventional thesis.

Dedication

I would like to dedicate this thesis to my husband, Gao Xiao. I want to thank him for being so patient with me and supportive through the these years of my study for the PhD degree. His love and wisdom have enabled me to face up to any difficulties and make me be the person who I am today. I would also like to thank my parents, Jingjuan Kang and Dexing Dai, for all of the financial support and encouragement throughout these years of my study abroad. All great wishes to you.

Acknowledgements

I am very grateful for my principal supervisor Professor Xiangjian He in the University of Technology Sydney (UTS) and co-supervisor Professor Dadong Wang in Data61, CSIRO. They have given me great support and help not only in my study but also in my life. I thank Professor He for helping me receive a UTS IRS Scholarship to fully cover my tuition fees and providing me Top-Up Scholarships from his research grants on several occasions. I thank Professor Wang for helping me receive a Data61 Top-Up Scholarship. I thank Professors He and Wang for enriching my academic experience.

I would also like to thank Prof Shuying Cheng, my supervisor in my home university - Fuzhou University, for sending me to UTS for this collaborative PhD degree.

Thank all students in the Computer Vision and Pattern Recognition Laboratory of UTS for having given me pleasant memories.

Thank my parents and my husband for supporting and encouraging me to pursue my PhD degree and keeping me going in the academic career.

Thanks all others who have helped me although I have not mentioned above!

Manna Dai
Sydney, Australia, July 2019.

List of Publications

Journal Papers

- J-1. Manna Dai, Shuying Cheng*, Xiangjian He*. “Hybrid generative-discriminative hash tracking with spatio-temporal contextual cues”, *Neural Computing and Applications*, vol. 29, no. 2, pp. 389-399, 2018. (SCI, ERA B, JCR IF 4.664)
- J-2. Manna Dai, Shuying Cheng*, Xiangjian He*, Dadong Wang. “Object tracking in the presence of shaking motions”, *Neural Computing and Applications*, pp.1-18, 2018. (SCI, ERA B, JCR IF 4.664)
- J-3. Manna Dai, Gao Xiao, Shuying Cheng*, Dadong Wang*, Xiangjian He*. “Structural Correlation Filters Combined with A Gaussian ParticleFilter for Hierarchical Visual Tracking”, *Neurocomputing*. (Under Review, SCI, ERA B, JCR IF 4.072)

Conference Papers

- C-1. Manna Dai, Peijie Lin, Lijun Wu, Zhicong Chen, Songlin Lai, Jie Zhang, Shuying Cheng*, Xiangjian He. “Orderless and Blurred Visual Tracking via Spatio-temporal Context”. *International Conference on Multimedia Modeling, MMM 2015*, pp. 25-36, 2015. (EI)

Contents

Certificate	ii
Abstract	iii
Dedication	vi
Acknowledgments	vii
List of Publications	viii
List of Figures	xiii
List of Tables	xx
Abbreviation	xxiv
Notation	xxvi
1 Introduction	1
1.1 Background and Significance	1
1.2 Review of Single-Object Tracking Methods	2
1.2.1 Generative Tracking Methods	3
1.2.2 Discriminative Tracking Methods	5
1.2.3 Mixed Tracking Methods	6
1.2.4 Research Difficulties of Single-Object Tracking	8
1.3 Technical Route of Single-Object Tracking Methods	12
1.4 Evaluation of Single-Object Tracking Methods	13
1.5 Thesis Organisation	17

2	Research on Single-Object Tracking in the Presence of Orderless Movement	19
2.1	Introduction	19
2.2	Analyses on Related Methods	21
2.2.1	Average Hash Method Based on Low-Frequency Image Information	21
2.2.2	Classical Tracking Methods Based on Particle Filtering	22
2.2.3	Tracker Based on Spatio-Temporal Context	23
2.2.4	HGDHT Tracking Framework Based on Bayesian Theory	25
2.3	Experimental Results and Discussion	30
2.3.1	Experimental Setting for Parameters	31
2.3.2	Quantitative Analysis	33
2.3.3	Qualitative Analysis	34
2.4	Conclusions and Future Work	39
3	Research on Single-Object Tracking under the Scenario of Motion Blur	41
3.1	Introduction	41
3.2	Analyses on Experimental Related Methods	43
3.2.1	Template-Matching Tracking Method Based on Motion Model	43
3.2.2	Sampling Strategy Based on Motion Model	44
3.2.3	Double-Template Strategy Based on Online Updating	46
3.2.4	Similarity Measure Based on a Gaussian-Uniform Kernel	47
3.2.5	Tracking Optimisation Based on Prior Probability Updating	50
3.2.6	Framework of the Proposed Method	52

3.3	Experiments	55
3.3.1	Experimental Environment and Parameter Settings	55
3.3.2	Tracking Evaluation in Scenarios of Motion Blur	57
3.3.3	Evaluation of Key Parts of SMT	65
3.3.4	Analysis of Experimental Results of Complete Databases	70
3.4	Conclusions	75
4	Research on Single-Object Tracking under the Scenario of Scale Variation	77
4.1	Introduction	77
4.2	Analyses on Experimental Related Methods	79
4.2.1	Structural Correlation Filters	79
4.2.2	Gaussian Particle Filter	80
4.2.3	Ensemble Tracking Method	81
4.3	Tracking Method Based on Correlation Filters and the Gaussian Particle Filter	82
4.3.1	Weak Classifiers Based on Structural Correlation Filters	82
4.3.2	Strong Classification via a Homogeneous Ensemble Layer	85
4.3.3	Gaussian Particle Filter Using CNN Features	87
4.3.4	KCF-GPF Tracker	89
4.4	Experiments	91
4.4.1	Experimental Setups	91
4.4.2	Quantitative Comparison	93
4.4.3	Qualitative Comparison	99
4.5	Conclusions	102

5	Research on Single-Object Tracking under the Scenario of Orderless Motion and Motion Blur	104
5.1	Introduction	104
5.2	Analyses on Experimental Related Methods	106
5.2.1	Strategy of Image Shrinking	106
5.2.2	Similarity Based on Euclidean Distance	107
5.2.3	Construction of Spatio-Temporal Context	109
5.2.4	Flowchart of the Designed Tracker	110
5.3	Experiments	112
5.3.1	Experimental Setups	112
5.3.2	Quantitative Comparison	112
5.3.3	Qualitative Comparison	113
5.4	Conclusions	118
6	Conclusions and Prospects	119
6.1	Research Summary	119
6.2	Research Innovations	122
6.3	Research Prospects	123
	Bibliography	125

List of Figures

1.1	Difficulties of Single-object Tracking.	9
1.2	A technical route of this thesis.	14
1.3	A schematic diagram of qualitative analysis. Results of different trackers are represented by bounding boxes with different colours. . .	14
1.4	The main contents and the organisational structure of this thesis. . .	18
2.1	The flowchart of STC method.	23
2.2	The flowchart of the generative part of HGDHT.	26
2.3	The flowchart of the discriminative part of HGDHT.	27
2.4	The means of average CLE plots of all tested sequences with various values of parameter $\alpha \in [1, 3]$	32
2.5	The means of average DP plots of all tested sequences with various values of parameter $\alpha \in [1, 3]$	33
2.6	The means of average DP plots of all tested sequences with various values of parameter $\alpha \in [1, 3]$	34
2.7	The error plots of all tested sequences for different tracking methods.	37
2.8	The comparison of our approach with state-of-the-art trackers on videos Bike, Body, Car2 and Car4.	39
2.9	The comparison of our approach with state-of-the-art trackers on videos David, Deer, Face and Shaking.	40

3.1	The search scope definition and the acquirement of the centre locations of candidate patches in the $(t + 1)$ -th frame. The blue box denotes the object in the $(t - 1)$ -th frame and the purple box denotes an object in the t -th frame. Then, we get the decomposition values $\mathbf{x}_t^{(1)}$ and $\mathbf{x}_t^{(2)}$ and lengthen them a times in order to obtain our search scope in the $(t + 1)$ -th frame. The right chart shows that the search scope is segmented into equal blocks and each point represents the centre of a candidate patch. The candidates are extracted by the sliding window.	45
3.2	The illustration of double-template strategy.	46
3.3	The effect of outliers. The sequence Clifbar contains the outlier caused by motion blur.	47
3.4	The illustration of the used Uniform kernel and the used Gaussian kernel and the Square distance.	48
3.5	The illustration of the prior probability model of context.	52
3.6	The flowchart of the SMT method.	54
3.7	Average DP and average OP of SMT with different θ	57
3.8	Average DP, average OP and average FPS of SMT with different a.	57
3.9	Average DP, average OP and average FPS of SMT with different d.	58
3.10	Average DP and average OP of SMT with different κ	58
3.11	Average DP and average OP of SMT with different α	60
3.12	Comparison results based on precision plot for all 20 evaluated movies with motion blur in the OTB-50 database with location errors below a threshold ρ in the range of $[0, 50]$ (pixels). The mean distance precision of each tracker is reported (colour figure online).	62

3.13	Comparison results based on success plot for all 20 evaluated movies with motion blur in the OTB-50 database with overlap percentages over a threshold η in the range of $[0, 1]$. The mean overlap precision of each tracker is reported (colour figure online).	62
3.14	The partial tracking results of our tracker and 6 state-of-the-art trackers in sequences BlurBody, BlurCar2 and BlurFace from OTB-50 database. Three sequences have motion blur attribute and also include other attributes, such as scale variation, deformation, fast motion or in-plane rotation.	64
3.15	The comparison of the proposed approach with variations in sampling, templates and similarity measures with and without the update process tested on the 20 videos that have motion blur and are selected from the OTB-50 database. The figure of the success plot contains the mean overlap precision at a threshold q of 20 pixels for each method (colour figure online).	67
3.16	Precision plot for all 50 sequences in the OTB-50 database with location errors below a threshold in the range of $\rho \in [0, 50]$ (pixels). The mean distance precision of each tracker is reported (colour figure online).	71
3.17	Success plot for all 50 evaluated sequences in the OTB-50 database with overlap percentages over a threshold in the range of $\eta \in [0, 1]$. The mean overlap precision of each tracker is reported (colour figure online).	71
3.18	Ranking plot for the experiment baseline in the VOT 2015 databases. The accuracy and robustness ranks are plotted along the vertical and horizontal axis, respectively. Our method (denoted by the green triangle) achieves superior results in the accuracy-robustness experiments (colour figure online).	75

- 4.1 The diagram of computing the optical flow using the Lucas-Kanade method (LK) [14] between two sequential images. The first column shows original full images at Frame 27 and Frame 28 from Sequence BlurBody in the OTB-2013 database [147]. The second column denotes an x -axis difference image I_x , a y -axis difference image I_y and a time-axis difference image I_t . I_x and I_y are obtained using the Scharr gradients on the input image. I_t is obtained by computing the pixel value differences between two images. As shown in the third column, these three output images of difference are employed to obtain an optical flow image at Frame 28 via the Least Square method (LS) [14]. 83
- 4.2 The diagram of the homogeneous ensemble layer. A sample set is generated via Eq. 4.6 and Eq. 4.7. Multiple Structural correlation filters are regarded as same-type weak classifiers, which are also called homogeneous base classifiers. Then, all weak classifiers are assembled as a strong classifier via a facile weighted sum strategy based on reliability estimation in Eq. 4.11. 85
- 4.3 The diagram of tracking using the Gaussian Particle Filter based on CNN features. The first column is an object in the last frame, and the second column denotes M Gaussian random samples with different target locations and different target scales in the current frame. Then, CNN features are extracted from each sample by using the pre-trained VGG-Net. Finally, a weighted sum strategy is employed to obtain the location and the scale of a target in the current frame. 88

- 4.4 The diagram of the architecture of the proposed KCF-GPF method. In the homogeneous ensemble layer, we execute motion detection using the LK optical flow method to find the potential locations of a target, generate weak classifiers in this potential locations, and assemble the weak classifiers to construct a strong classifier to obtain a target location. In the CNN-based GPF layer, samples are generated in the target location and CNN features are extracted from each sample via the pre-trained VGG-Net. The weight of each sample is measured. Samples with weights are combined to predict the final location and the scale of a target. 89
- 4.5 Success plots over all 50 sequences using OPE evaluation in the OTB-2013 database. The evaluated trackers are LMCF, CFNet, CFN, CFN_, CNT, BIT, SINT, SCT and KCF-GPF. All 11 tracking challenges include scale variation, out of view, out-of-plane rotation, low resolution, in-plane rotation, illumination, motion blur, background clutter, occlusion, deformation, and fast motion. The numbers in the legend indicate the average AUC scores for success plots. Our KCF-GPF method performs favourably against the state-of-the-art trackers. 95
- 4.6 Success plots over all 50 sequences using OPE evaluation in the OTB-2013 database. The evaluated trackers are Staple, SiamFC, SRDCF, DSST, MEEM, KCF, TLD, Struck and KCF-GPF. All 11 tracking challenges include scale variation, out of view, out-of-plane rotation, low resolution, in-plane rotation, illumination, motion blur, background clutter, occlusion, deformation, and fast motion. The numbers in the legend indicate the average AUC scores for success plots. Our KCF-GPF method performs favourably against the state-of-the-art trackers. 96

4.7	Precision and success plots over all 50 sequences using OPE evaluation in the OTB-2013 database. The numbers in the legend indicate the average precision scores for precision plots and the average AUC scores for success plots. Our KCF-GPF method performs favourably against the state-of-the-art trackers.	97
4.8	Success plots over all 100 sequences using OPE evaluation in the OTB-2015 database. Our method performs favourably against the state-of-the-art trackers.	98
4.9	The comparisons of the proposed tracker with the state-of-the-art trackers (SCT [24], CFNet [133], KCF [57], [69] and Struck [54]) in our evaluation on 10 challenging sequences (from left to right and top to down are Shaking, Lemming, Skating1, Subway, Singer2, Suv, Liquor, Woman, Soccer, Dog1, respectively).	100
5.1	The flowchart of the proposed algorithm. A template and samples are shrunk into 2×2 pixels. Euclidean Distance is used to compute similarity scores between a template and samples. A sample with the greatest score is regarded as a primary location, which is used to define new detection scope in this frame. In this detection scope, a spatio-temporal context model is built up for obtaining the final location of a target. In each frame, the final location of a target is used to update a template used in the next frame.	107
5.2	The comparison of our approach with state-of-the-art trackers in the challenging attributes of motion blur, fast motion, illumination change, scale variation, occlusions, rotation, background clutter, and pose variation. Especially our tracker can tackle motion blur and fast motion.	115

5.3 The comparison of our approach with state-of-the-art trackers in challenging attributes of motion blur, fast motion, illumination change, scale variation, occlusions, rotation, background clutter, and pose variation. Especially our tracker can tackle motion blur and fast motion. 116

List of Tables

2.1	Evaluated video sequences. ‘√’ denotes that the sequence contains the corresponding challenge, and ‘×’ implies that the challenge is excluded.	31
2.2	Centre location error (CLE) (in pixels). The best results are shown in red while the second and third ones are shown in blue and green. .	35
2.3	Distance precision (DP) (in pixels). The best results are shown in red while the second and third ones are shown in blue and green. . .	36
2.4	Comparison with average frames per second (FPS). The best results are shown in red while the second and third ones are shown in blue and green.	38
3.1	The 11 tested sequences containing motion blur attribute. They are selected from the OTB-100 database. ‘√’ indicates that the corresponding sequence has the corresponding attribute, and ‘×’ implies that the corresponding sequence does not have the corresponding attribute. SV, DEF, MB, FM, IPR, OPR, BC, IV, OV, OCC and LR represent the attributes of scale variation, deformation, motion blur, fast motion, in-plane rotation, out-of-plane rotation, background clutters, illumination variation, out-of-view, occlusion and low resolution, respectively.	56

- 3.2 The 20 video sequences that are selected from the OTB-50 database and have the attribute of motion blur. ‘√’ indicates that the corresponding sequence has the corresponding challenge, and ‘×’ implies that the corresponding sequence does not have the corresponding attribute. SV, DEF, MB, FM, IPR, OPR, BC, IV, OV, OCC and LR represent the attributes of scale variation, deformation, motion blur, fast motion, in-plane rotation, out-of-plane rotation, background clutters, illumination variation, out-of-view, occlusion and low resolution, respectively. 59
- 3.3 The quantitative comparison of our trackers with 12 state-of-the-art methods on 20 challenging sequences with motion blur attribute in the OTB-50 database. The results are reported in distance precision (DP) (%). We also provide the average values of DP. Here, the DP values is set to a threshold ρ of 20 pixels (referring to [155]) and the OP values is set to a threshold η of 0.5 (referring to [155]). The best results are shown in red while the second and third ones are shown in blue and green, respectively. Note that the proposed approach achieves the best average performance in terms of average DP and average OP, and the second best in terms of FPS. 61
- 3.4 The quantitative comparison of our trackers with 12 state-of-the-art methods on 20 challenging sequences with motion blur attribute in the OTB-50 database. The results are reported in distance precision (DP) (%). We also provide the average values of DP. Here, the DP values is set to a threshold ρ of 20 pixels (referring to [155]) and the OP values is set to a threshold η of 0.5 (referring to [155]). The best results are shown in red while the second and third ones are shown in blue and green, respectively. Note that the proposed approach achieves the best average performance in terms of average DP and average OP, and the second best in terms of FPS. 61

3.5	The processing time (FPS) comparison between the WMIL sampling method and the proposed sampling method over the 20 videos with motion blur selected from the OTB-50 database.	68
3.6	The processing time (FPS) comparison between the proposed approach with five different templates including the proposed double-templates over the 20 videos with motion blur selected from the OTB-50 database. The best result is shown in red while the second and third ones are shown in blue and green, respectively.	69
3.7	The comparison of our tracker with 12 state-of-the-art trackers on the whole OTB-50 database. The results are reported in terms of average overlap precision (OP) (%), average distance precision (DP) (%) and frame-per-second (FPS). Here, DP values are obtained at a threshold q of 20 pixels and the OP values are obtained at a threshold g of 0.5. The best results are displayed in bold while the second and third best results are shown in italic and bold italic, respectively. The results of our tracker are among the top three.	72
3.8	The comparison of our tracker with 12 state-of-the-art trackers on the whole OTB-50 database. The results are reported in terms of average overlap precision (OP) (%), average distance precision (DP) (%) and frame-per-second (FPS). Here, DP values are obtained at a threshold q of 20 pixels and the OP values are obtained at a threshold g of 0.5. The best results are displayed in bold while the second and third best results are shown in italic and bold italic, respectively. The results of our tracker are among the top three.	73
3.9	The comparison of our tracker with 12 state-of-the-art trackers on the whole OTB-50 database.	74
4.1	The parameters of KCF-GPF	91

- 4.2 The tracking results of 17 evaluated trackers over all 50 sequences using OPE evaluation in the OTB-2013. The entries in **red** denote the best, while the ones in **blue** indicate the second best and the ones in **green** represent the third best. 93
- 4.3 The tracking results of 17 evaluated trackers over all 50 sequences using OPE evaluation in the OTB-2013. The entries in **red** denote the best, while the ones in **blue** indicate the second best and the ones in **green** represent the third best. 93
- 4.4 The tracking results of all 4 evaluated trackers over all 100 sequences using OPE evaluation in the OTB-2015. The entries in **red** denote the best results and the ones in **blue** indicate the second best. 99
- 5.1 Centre location error (CLE) (in pixels). The entries in **red** denote the best, while the ones in **blue** indicate the second best and the ones in **green** represent the third best. 113
- 5.2 The results of an average centre location error (CLE) (in pixels), an average distance precision (DP) (%), an average overlap precision (OP) (%), and an average frames per second (FPS). The entries in **red** denote the best, while the ones in **blue** indicate the second best and the ones in **green** represent the third best. 114

Abbreviation

BC - Background Clutters
CLE - Center Location Error
DEF - Deformation
DP - Distance Precision
FFT - Fast Fourier Transformation
FM - Fast Motion
FPS - Frames Per Second
HOG - Histograms of Oriented Gradient
IPR - In-Plane Rotation
ISM - Incremental Similarity Matrices
IV - Illumination Variation
LBP - Local Binary Patterns
LR - Low Resolution
MIL - Multiple Instance Learning
MAP - Maximum a Posteriori
MB - Motion Blur
OCC - Occlusion
OP - Overlap Precision
OPE - One Pass Evaluation
OPR - Out-of-Plane Rotation
OV - Out-of-View
PCA - Principal Components Analysis
PF - Particle Filter

RBF - Radial Basis Function

SSD - Sum of Squared Differences

SV - Scale Variation

SVM - Support Vector Machine

SVT - Support Vector Tracking

Nomenclature and Notation

Capital letters denote matrices.

Lower-case alphabets denote column vectors.

$(\cdot)^T$ denotes the transpose operation.

$\mathbf{G} = [G_{ij}]_{m \times n}$ is the identity matrix of dimension $m \times n$.

\sum represents the summation notation.

$\sqrt{\quad}$ denotes the cube root notation.

\cup represents the intersection operation.

\cap is the union operation.