

# Combating Tracking Drift: Developing Robust Object Tracking Methods



**Jiatong Li**

Faculty of Engineering and Information Technology  
University of Technology, Sydney

This thesis is submitted for the degree of  
*Doctor of Philosophy*

2017



## **Certificate of Original Authorship**

I certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This thesis is the result of the research candidature conducted jointly with Beijing Institute of Technology as part of a collaborative doctoral degree.

Signature of Student: \_\_\_\_\_

Date: \_\_\_\_\_

26 / 7 / 2017



I would like to dedicate this thesis to my loving parents and wife



## Acknowledgements

I would like to express my sincere appreciations to those who have supported me and contributed to this thesis during my PhD journey.

First of all, I would like to thank my supervisor, Prof. Richard Xu, for his valuable guidance and useful discussions during my research developing stage. He is who led me to the field of academic research, and who gave me a great opportunity to broaden my research view. His passion for academic knowledge always motivates me to move forward towards my research goals. Besides, I wish to show great gratitude to my co-supervisor Prof. Massimo Piccardi, and Prof. Min Xu, Prof. Qiang Wu for their great support during the course of my study. I would also like to thank Prof. Baojun Zhao and Prof. Chenwei Deng for their great help when I was in Beijing Institute of Technology. Very special thanks to Prof. Dacheng Tao for his invaluable insights and suggestions for the last stage of my PhD work.

Heartfelt thanks to my fellow labmates in Faculty of Engineering and Information Technology, UTS. Also great thanks to my friends and collaborators. We worked and enjoyed the life together for our intensive and joyful time. They are: Minqi Li, Xiang Feng, Shuai Jiang, Chang Liu, Tianrong Rao, Qingbo Xia, Lingxiang Wu, Tian Ding, Bin Yang, Peibo Duan, Felix, Sheng Wang, Zhichao Sheng, Maoying Qiao, Junyu Xuan, Wenjie Zha, Hongshu Chen, Fan Dong, Haimin Zhang, Baosheng Yu, Tongliang Liu, Zhe Chen, Zijing Chen. In particular, I am grateful to Jessica Anne Washington for her help in the thesis writing.

Finally, I would like to show my deeply gratitude to my family: my parents and my wife, for always supporting me selflessly in my life and study. It's their constant encouragement and determination that made all these possible ultimately.





## Abstract

Visual object tracking plays an important role in many computer vision applications, such as video surveillance, unmanned aerial vehicle image processing, human computer interaction and automatic control. This research aims to develop robust object tracking methods, which are capable of tracking general object without the prior knowledge of the target. Tracker drift is one of the most challenging issues in object tracking due to target deformations, illumination variations, abrupt motions, occlusions and background clutters. This thesis focuses on the tracking drift problem, and adopts three main solutions. These include: designing an efficient target shape feature extraction method, comparing target features with metric learning and using the ensemble tracking method to tackle the tracking drift during tracker online update. The main work and contributions are as follows:

1. We propose a Weber's Law Shape Descriptor (WLSD) for efficient object tracking under background clutters. Weber's Law indicates that the perceived change in stimuli is proportional to the initial stimuli, which means that the saliency variations is not only relative to the feature variations, but also the initial feature quantity. Motivated by Weber's Law, this thesis proposes a Weber's Law Shape Descriptor to describe the saliency variations of the shape contour. The proposed method first designs a Contour Angular Feature as the initial stimuli, and builds the WLSD according to Weber's Law. Then, WLSD is extended to multi-scale to enhance its shape discriminative strength. Finally, a feature selection scheme is used to extract the effective WLSD scales. The proposed WLSD is naturally invariant to the shape scale and rotation, and has low computation load as a whole shape descriptor. We further apply WLSD in thermal infrared object tracking, and propose a multi-feature integration method based on WLSD shape, target area and target trajectory. The experiments are first conducted on MPEG-7 and Tari shape dataset to test the shape discriminative capacity of WLSD, and then tested on the infrared tracking videos to validate the proposed multi-feature integration tracking framework.
2. We propose a Time Varying Metric Learning (TVML) object tracking method. Recently, tracking-by-detection (TBD) methods are very popular. Traditional TBD methods track the objects by training a binary classifier to discriminate the target and background, which

leads to two main issues: firstly, the classifier is unreliable when trained with insufficient data, secondly, comparing the object features with traditional Euclidean distance leans to tracker drift. To solve the above problems, we propose a time varying metric learning model and apply it to object tracking. The proposed TVML adopts Wishart Process to model the variation of the positive semi-definite (PSD) matrices, i.e. the metrics, and uses the Recursive Bayesian Estimation (RBE) framework to learn the metrics under the side information constraint. We introduce the side information constraint to omit the clustering of the negative samples, which is very suitable to the background cluster tracking scenarios. Furthermore, the RBE framework guarantees the proposed model is able to estimate the metrics with limited training data. The experimental results demonstrate the comparable performance of the TVML tracker compared to many state-of-the-art methods on OTB-50 dataset.

3. We propose a historical tracker snapshots based ensemble tracking method. There are frequent target appearance changes due to illumination variations, abrupt motions and target deformations, which needs the tracker to conduct online update to adapt to the target appearance variations. The online update of the tracker often leads to drift. This thesis proposes a historical tracker snapshots based ensemble tracking framework, and designs a Scale-adaptive Multi-Expert (SME) tracker according to the proposed method. The tracker ensemble is composed of the current tracker and its historical tracker snapshots. When the tracker drift is detected, the framework will select the suitable tracker snapshot to replace the current drift tracker according to their accumulated scores. Due to the fact that the tracker tends to be more confident to its own prediction, we propose to define the tracker score in a semi-supervised learning perspective to describe the consistency and the ambiguity of the tracker ensemble simultaneously. We use the regression correlation filter as the base tracker due to its high efficiency. Furthermore, we propose to establish the target scale pyramid to estimate the target scale accurately. The proposed SME tracker is tested on the OTB-50 and VOT2015 tracking dataset, which demonstrates the excellent performance of the proposed tracker with real time speed.
4. We propose a discrete graph based ensemble tracking method. The tracker ensemble is constituted by the current tracker and its historical snapshots as the multi-expert. This thesis proposes to introduce the discrete graph to model the tracker ensemble, where the graph node represents the expert hypothesis. After defining the unary and pair-wise score of the graph, the best expert is selected according to the graph path of the highest score. With the efficient solver of dynamic programming, the proposed method can implicitly analyze the reliability of the multi-expert trajectories by only computing their scores in the current frame, so as to correct the tracker drift. We integrate three base trackers into the proposed

tracking framework to validate its generality, including online support vector machine on a budget, hand-craft feature based correlation filter and convolutional neural network based correlation filter. The proposed three trackers are widely tested on the OTB-50, OTB-100 and VOT2015 dataset, which demonstrates the proposed trackers are superior to the compared state-of-the-art trackers in both the tracking accuracy and robustness measures.

**Keywords** — Object tracking, Weber’s Law, time varying metric learning, ensemble tracking, correlation filter, discrete graph.



# Table of contents

<b>List of figures</b>	<b>xvii</b>
<b>List of tables</b>	<b>xxi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Research Background . . . . .	1
1.1.1 Image Processing and Computer Vision . . . . .	1
1.1.2 Object Tracking . . . . .	2
1.2 Literature Review . . . . .	2
1.2.1 Generative Methods . . . . .	3
1.2.2 Discriminative Methods . . . . .	4
1.2.3 Ensemble Methods . . . . .	10
1.3 Tracking Drift . . . . .	11
1.4 Contributions . . . . .	12
1.5 Publications Related to the Thesis . . . . .	15
<b>2 Weber’s Law Shape Descriptor for Infrared Object Tracking</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Related Work for Shape Descriptors . . . . .	18
2.3 Weber’s Law Shape Descriptor . . . . .	20
2.3.1 Weber’s Law . . . . .	20
2.3.2 The Original Stimulus Feature - Contour Angle Feature . . . . .	20
2.3.3 Weber’s Law Shape Descriptor . . . . .	22
2.3.4 WLSD Histogram and WLSD Scale Selection . . . . .	25
2.4 WLSD for Object Tracking . . . . .	26
2.4.1 Infrared Target Representation . . . . .	26
2.4.2 Multi-feature Integration Tracking . . . . .	27
2.5 Experimental Results . . . . .	29

2.5.1	Evaluation on MPEG-7 Dataset . . . . .	29
2.5.2	Evaluation on Tari Dataset . . . . .	34
2.5.3	Experiment for WLSO based Tracking . . . . .	36
2.6	Conclusion . . . . .	39
<b>3</b>	<b>Time Varying Metric Learning for Object Tracking</b>	<b>41</b>
3.1	Introduction . . . . .	41
3.2	Related Work . . . . .	42
3.3	Time Varying Metric Learning Model . . . . .	43
3.3.1	Wishart Process . . . . .	44
3.3.2	Side Information Constraint . . . . .	44
3.3.3	Graphical Model for Time Varying Metric Learning . . . . .	45
3.3.4	The Sequential Monte Carlo Solution . . . . .	45
3.4	TVML for Tracking . . . . .	47
3.4.1	TVML Tracking Framework . . . . .	47
3.4.2	Target Representation and Model Update . . . . .	48
3.5	Experiments . . . . .	49
3.5.1	Model Validation . . . . .	50
3.5.2	Experiment for Object Tracking on OTB-50 . . . . .	51
3.6	Conclusion . . . . .	56
<b>4</b>	<b>Tracker Snapshot Based Multi-Expert Tracking</b>	<b>57</b>
4.1	Introduction . . . . .	57
4.2	Related Work . . . . .	59
4.3	The Proposed Tracker . . . . .	60
4.3.1	Tracker Snapshot Based Multi-Expert Tracking Framework . . . . .	60
4.3.2	Multi-Expert Selection Criteria . . . . .	61
4.3.3	Base tracker of Regression Correlation Filter . . . . .	63
4.3.4	Scale-adaptive Multi-Expert (SME) Tracker . . . . .	65
4.4	Implementation . . . . .	66
4.5	Experiments . . . . .	67
4.5.1	Experimental Setup and Metrics . . . . .	68
4.5.2	Evaluation on Object Tracking Benchmark-50 . . . . .	68
4.5.3	Evaluation on VOT2015 Challenge Dataset . . . . .	71
4.6	Conclusion . . . . .	72

---

<b>5</b>	<b>Discrete Graph Based Multi-Expert Tracking</b>	<b>73</b>
5.1	Introduction . . . . .	73
5.2	Related Work . . . . .	74
5.3	Proposed Method . . . . .	76
5.3.1	Relational Graph and Hypothesis Graph . . . . .	76
5.3.2	Multi-Expert Framework modeled by Discrete Graph . . . . .	78
5.3.3	Base Trackers for Multi-Expert Framework . . . . .	80
5.4	Implementations and Method Analysis . . . . .	81
5.4.1	Tracking by Multi-Expert Framework . . . . .	81
5.4.2	Multi-Expert Framework Efficiency . . . . .	83
5.4.3	Multi-Expert Framework Robustness . . . . .	84
5.5	Experiments . . . . .	87
5.5.1	Experimental Setup . . . . .	87
5.5.2	Evaluation on OTB-50 Dataset . . . . .	88
5.5.3	Evaluation on OTB-100 Dataset . . . . .	93
5.5.4	Evaluation on VOT2015 Challenge Dataset . . . . .	98
5.6	Conclusion . . . . .	100
<b>6</b>	<b>Conclusions and Future Work</b>	<b>101</b>
	<b>References</b>	<b>103</b>





# List of figures

1.1	Multiple Channel Correlation Filter Illustration. . . . .	7
1.2	Boundary effect of Correlation Filter. (a) Defines the example of fixed spatial support within the image from which the peak correlation output at the center of the bounding box. (b) A subset of patch examples used in a correlation filter where green denotes a non-zero correlation output, and red denotes a zero correlation output. . . . .	9
1.3	Thesis organization. . . . .	14
2.1	Illustration for $CAF_s$ . . . . .	21
2.2	$WLS D_{1 \rightarrow 1}$ (single scale WLS D) value linearly mapped to the colormap bar . . . . .	23
2.3	Illustration for $WLS D_{s \rightarrow w}$ . . . . .	24
2.4	$WLS D_{5 \rightarrow 5}$ visualization . . . . .	24
2.5	Target Searching flowchart . . . . .	28
2.6	MPEG-7 examples . . . . .	30
2.7	(a) The retrieval rate of single scale WLS D; (b) The front elevation of (a); (c) The left side elevation of (a); . . . . .	32
2.8	(a) The bulleye score variation with iteration increasing; (b) The first five selected scales of each training set in MPEG-7. . . . .	33
2.9	(a) The first ten selected scales accumulation of the training sets; (b) The variance of the first ten selected scales. . . . .	34
2.10	Tari dataset examples . . . . .	35
2.11	(a) The bulleye score variation with iteration increasing on Tari dataset; (b) The distribution of the selected scales of WLS D in Tari. . . . .	36
2.12	The variance of the first ten selected scales on Tari dataset. . . . .	37
2.13	Tracking snapshots for infrared video without occlusion. (Left columns with green rectangles: the proposed method, right column with red rectangles: mean-shift.) . . . . .	38

2.14	Tracking snapshots for infrared video with occlusion. (Lower two rows with green rectangles: the proposed method, upper two rows with red rectangles: mean-shift.) . . . . .	40
3.1	Time Varying Metric Learning Graphical Model . . . . .	46
3.2	TVML for tracking. There are three main steps: training (the flows in the first row), testing and model update (the flows in the second row). The training step collect the training samples and learn the metric. The testing step then use the learned metric to search the current target location followed by the step of online model update. . . . .	48
3.3	TVML on synthetic data: (a) metric determinant; (b) elliptical representation.	50
3.4	The success plot and precision plot over OTB-50 using one pass evaluation (OPE). The legend illustrates the area under curve (AUC) for the success plot, and the score of the threshold 20 for the precision plot. Only top-10 trackers are colored, and the others are shown in gray curves. Legend of the same color denotes the same rank. . . . .	52
3.5	The precision plots and success plots for four main attributes of the benchmark, i.e. background clutter, occlusion, deformation and scale variation. The figure in the title denotes the number of sequences belongs to specific attribute. . . .	54
3.6	Tracking snapshots of the top six algorithms among the overall performance including TVML, TLD [7], ASLA [14], VTS [23], VTD [13], CXT [110] over eight sequences. The illustration example videos from top-left to bottom-right are <i>Basketball</i> , <i>Bolt</i> , <i>David</i> , <i>Jogging</i> , <i>Jumping</i> , <i>Football</i> , <i>Singer2</i> and <i>SUV</i> . . .	55
4.1	Four typical sequences ( <i>Coke</i> , <i>Lemming</i> , <i>Tiger1</i> , <i>Shaking</i> ) show the importance of exploiting the historical tracker snapshots. The cyan bounding box is the tracking result of our base tracker, while the yellow box is the selected tracker historical snapshot by the multi-expert framework. The tracker snapshots are stored at the pre-defined interval, and the corresponding response maps are illustrated at the bottom-right of the image (listed in chronological order from left to right). The color brightness of the response map indicates the confidence degree of the tracker snapshot. The number at the top-left corner is the frame count. . . . .	58
4.2	Tracking snapshot based multi-expert framework illustration. . . . .	61

- 
- 4.3 The precision plot and success plot over 50 sequences benchmark using one pass evaluation (OPE). The legend illustrates the score of the threshold 20 for the precision plot, and the area under curve (AUC) for the success plot. Only top-10 trackers are colored, and the others are shown in gray curves. Legend of the same color denotes the same rank. . . . . 68
- 4.4 The success plots of eight attributes of the benchmark, i.e. scale variation, occlusion, out-of-plane rotation, background clutter, motion blur, deformation, illumination variation and out of view. The legend illustrates the AUC score for each tracker. . . . . 69
- 4.5 The AR ranking plot and the AR plot for VOT2015 Challenge Dataset. The tracker is better if its legend resides closer to the top-right corner of the plot.  $S$  is the data visualization parameter. . . . . 71
- 5.1 Graph illustration. (a) Relational Graph for single branch tree. (b) Hypothesis graph generated according to (a) assuming each entity of relational graph has four possible states. The brown balls denote the hypothesis nodes and the lines between balls represent the edges of the hypothesis graph. Note this is only the illustration and not all the edges are drawn for conciseness. . . . . 77
- 5.2 Illustration for the multi-expert framework modeled by discrete graph. The multi-expert ensemble includes the current tracker represented by the brown nodes and the historical snapshots colored by blue nodes. The historical tracker snapshot is stored at intervals of  $\Delta t$  frames, which is denoted by the black dash line, and the oldest expert is discarded if the expert number exceeds the maximum (the figure illustrates maximum 4 experts). Arrow direction denotes the same expert, and the graph edge exists in pair-wise adjacent nodes. . . . . 79
- 5.3 Illustrations for Drift Correction by the proposed graph based multi-expert framework. Four experts are used, and they are stored in chronological order whose trajectories are colored by yellow, red, blue and green respectively (The green one denotes the current tracker). The current tracker tends to drift, and the multi-expert framework corrects the drift by selecting the historical tracker hypothesis colored by yellow. The corresponding expert selection results are shown in the graph. Note the framework finds the disagreement in frame 368 and corrects the drift in frame 372. The response maps of the experts in frame 372 are shown at the bottom-right corner of that image. . . . . 83

5.4	The success/precision scores and FPS of HSME-deep, HSME-CF and HSME-SVM when the expert number varies. The bar plots show the success/precision scores of the three trackers with y-axis on the left, while the curve plots show the FPS of the three trackers with y-axis on the right. . . . .	84
5.5	The success/precision scores of HSME-deep, HSME-CF and HSME-SVM when the unary score parameter varies. . . . .	85
5.6	The success/precision scores of the proposed three trackers when the likelihood variance factor varies. Refer to Sec. 5.4.3 for details. . . . .	86
5.7	The success plots and precision plots over OTB-50 dataset using one pass evaluation (OPE). The legend illustrates the area under curve (AUC) for the success plot, and the score of the threshold 20 for the precision plot. Only scores of the top-15 trackers are shown, and the others are plotted in light gray curves. . . . .	89
5.8	The success plots and precision plots over OTB-50 dataset using temporal robustness evaluation (TRE) and spatial robustness evaluation (SRE). . . . .	90
5.9	Success plots for eight attributes of the top-10 performance trackers on OTB-50. 91	
5.10	The success plot and precision plot over OTB-100 dataset using one pass evaluation (OPE). Legends for tracker scores are located beside each plot. Only top-15 trackers are colored, and the others are shown in gray. . . . .	94
5.11	Long term sequences snapshots. Sequence name is at the lower left corner. . .	94
5.12	Tracking results of the top nine algorithms (HSME-deep, HSME-CF, HSME-SVM, STCT [47], Staple [126], HCT [41], MEEM [55], KCF [9], TGPR [135]) on TB-100 over twelve typical sequences. The video illustrations from top to bottom are <i>Trellis</i> , <i>Walking2</i> , <i>Dog1</i> , <i>Human8</i> , <i>Box</i> , <i>Skater2</i> , <i>Girl2</i> , <i>Human4</i> , <i>Jogging</i> , <i>Skiing</i> , <i>MotorRolling</i> and <i>Skating1</i> . . . . .	97
5.13	The AR ranking plot for VOT2015 Dataset. The tracker is better if its legend resides closer to the top-right corner of the plot. . . . .	98

# List of tables

2.1	Overall performance comparison . . . . .	31
2.2	Overall performance comparison on Tari dataset . . . . .	35
2.3	Bull'eye score on rotated and scaled Tari dataset . . . . .	36
3.1	Computation loads of the top-10 trackers in Fig. 3.4 are presented in three aspects, including frames per second (FPS), tracking method and the implementation. For method, S: Sampling based method, GS: Grid search based method. For implementation, M: Matlab, MC: Matlab + C, E: executable code.	53
4.1	Success and Precision Score of the SME Analysis . . . . .	70
5.1	Scores of the attribute plots in Fig. 5.9. . . . .	92
5.2	Success and precision scores for the proposed trackers and their base trackers. (Darker cells denote higher scores.) . . . . .	93
5.3	AUC \CLE scores for long term sequences shown in Fig. 5.11. . . . .	93
5.4	The results of VOT2015 Challenge Dataset. . . . .	99

