Object Tracking in Video Using the TLD and CMT Fusion Model

Hai Tran
Department of Informatics Technology
University of Education
HCMC, Vietnam
Email: haits [AT] hcmue.edu.vn

Abstract— **Object tracking has been an attractive study topic in computer vision in recent years, thanks to the development of video monitoring systems. Tracking-Learning Detection (TLD), Compressive Tracking (CT), and Clustering of Static-Adaptive Correspondences for Deformable Object Tracking are some of the state-of-the-art methods for motion object tracking (CMT). We present a fusion model that combines TLD and CMT in this study. To restrict the calculation time of the CMT technique, the fusion TLD CMT model enhanced the TLD benefits of computation time and accuracy on t no deformable objects. The experimental results on the Vojir dataset for three techniques (TLD, CMT, and TLD CMT) demonstrated that our fusion proposal successfully trades off CMT accuracy for computing time.**

Keywords:* *Object Tracking, Tracking-Learning-Detection (TLD), Clustering of Static-Adaptive Correspondences for Deformable Object Tracking (CMT).

I. INTRODUCTION (HEADING 1)

Object tracking has been an attractive study topic in computer vision in recent years, thanks to the development of video monitoring systems. Some are single-camera tracking systems, while others are multi-camera tracking systems. In industrialized nations, almost all of these systems are based on high resolution or full HD. Many video surveillance systems exist in Vietnam, a growing country.

In the field of object tracking, a new line of study has recently been presented that focuses on improving tracing accuracy while reducing processing time. Structured Output Tracking (STR) [1], Tracking-Learning Detection (TLD) [2,10], Sparsity-based Collaborative Model (SCM) [3], Fragments-based Tracking (FT) [4], Compressive Tracking (CT) [5], and Clustering of Static-Adaptive Correspondences for Deformable Object Tracking (CMT) [6,9] are some of the state-of-the-art methods for motion object tracking.

We have a comparison to these state-of-the-art methods using the Vojir dataset, which consists of sequences of varied lengths for the assessment, according to Georg Nebehay [6] et al. On the Vojir dataset of start of art tracking techniques, the average number of processed frames per second is shown in the table below:

TABLE I. Processed frames/sec of start of art

Tracking Method	No processed frames/sec
CMT	10.47
STR	11.96
TLD	18.16
SCM	2.21
FT	961
CT	45.33

The goal of this study is to find an object tracking approach that is appropriate for the Vietnam situation. Due to the low configuration of many Vietnam cameras monitoring systems, the tradeoff of accuracy for longer processing time is acceptable. To use in a real-world application in Vietnam, such as item tracking in apartment building surveillance cameras or infant tracking at a preschool.

Trackers are simple to use, need no startup, and create smooth trajectories. They, on the other hand, acquire inaccuracy over time (drift) and generally fail if the item is deformed or vanishes from the camera view. CMT can overcome these weaknesses.

With typical quality processing cameras systems in Vietnam, the TLD technique has a high processing time, but the CMT method has a greater accuracy. In order to enhance the processing time of CMT by combining with the TLD technique, a combinational model comprising two tracking methods is proposed. TLD and CMT are both open source and support the C++ programming language.

The following is how the rest of the paper is organized: TLD and CMT are two related works studied in Section II. Section III delves into the details of our TLD CMT fusion proposal model based on [7, 11], which combines TLD and CMT tracking methods. The experiments and assessment of our proposed model TLD CMT are discussed in Section IV. The final section contains the conclusion and recommendations for further study.

II. BACKGROUND AND RELATED OF WORK

A. *Tracking-Learning-Detection (TLD)*

TLD is a common technique used in many tracking systems. It indicates that the detection and tracking processes are both active at the same time. When the item does not disappear in the frame scene, these approaches are suitable; nevertheless, they are difficult to use when the object is out of view. TLD will solve these issues since the detection and tracking of objects are done independently.

Furthermore, as compared to prior traditional approaches, TLD will improve the precision of the learning process.

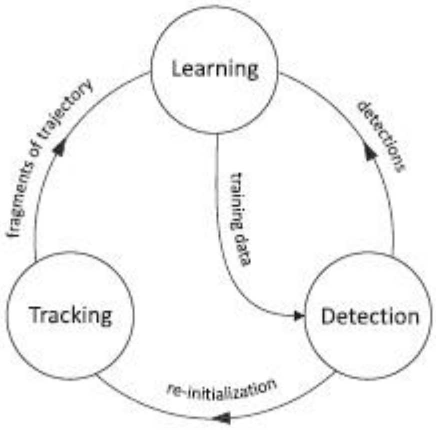


Figure 1. TLD overview process [7]

B. Clustering of Static-Adaptive Correspondences for Deformable Object Tracking (CMT)

CMT [6] use the cluster of correspondences for distinguishing the inlier and outlier key points. CMT is key point based tracking algorithm based on hierarchical clustering algorithm.

While TLD isn't great at tracking deformed things, CMT has a greater accuracy than TLD when it comes to deformable items. The clustering method, on the other hand, takes longer to process. Furthermore, the best cutoff threshold value should be determined during the procedure.

III. THE FUSION TLD AND CMT METHOD FOR MOTION OBJECT TRACKING

The aim of our TLD CMT model proposal is to integrate the TLD and CMT together. The three process stages in the fusion TLD CMT model are hierarchical clustering, identifying deformable/non-deformable objects, and running the tracking algorithm.

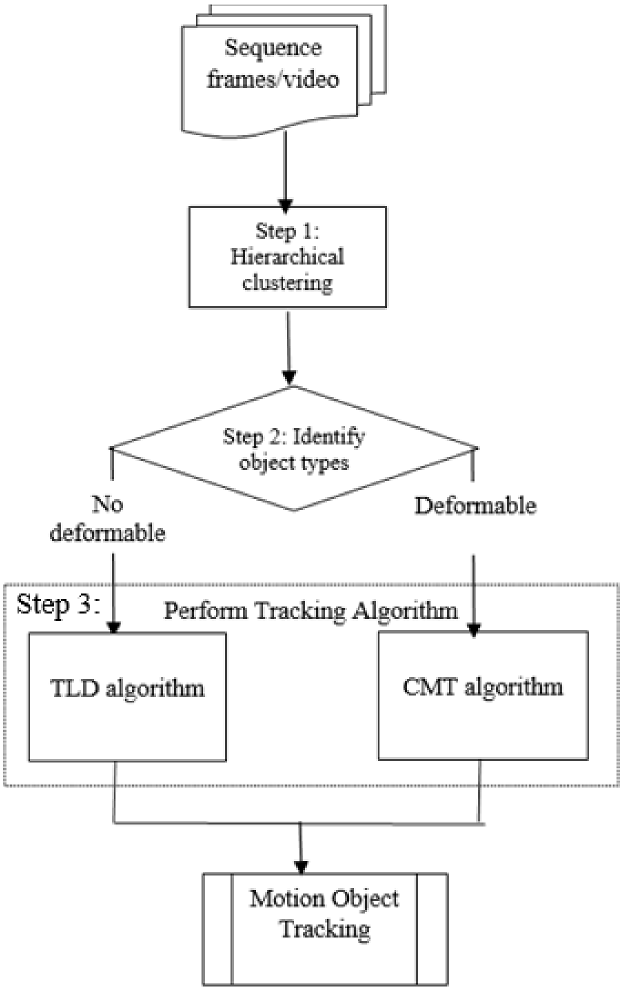


Figure 2. The Fusion TLD_CMT proposal model

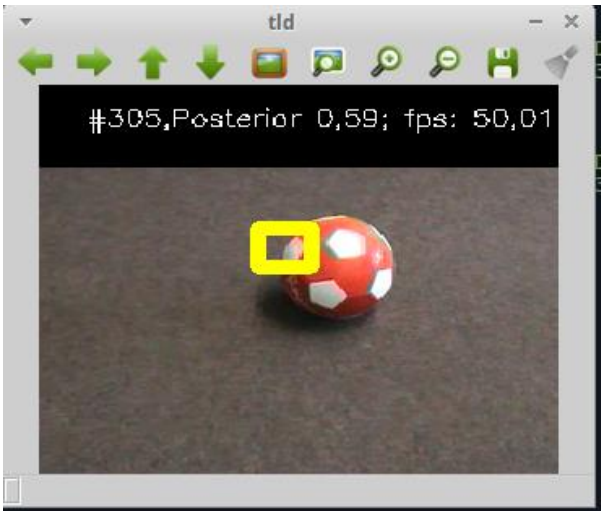
In step 1, the hierarchical clustering and key point matching of CMT will be performed to reduce the CMT algorithm's calculation time.

In step 2, identify object categories () in order to select the best tracking method for utilizing the benefits of CMT on deformable objects while keeping calculation time to a minimum.

In step 3, choose an appropriate tracking algorithm that balances time and accuracy.

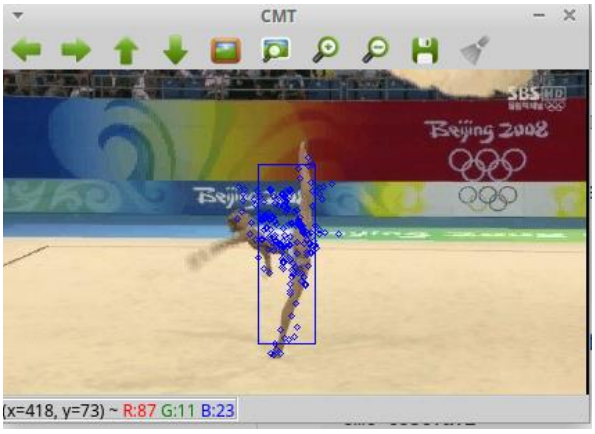
IV. EXPERIMENTAL RESULT AND DISCUSSION

We also created our proposal model on C++ programming languages in Ubuntu for our tests, based on TLD and CMT source code. TLD and CMT were also installed on our machine for comparison.



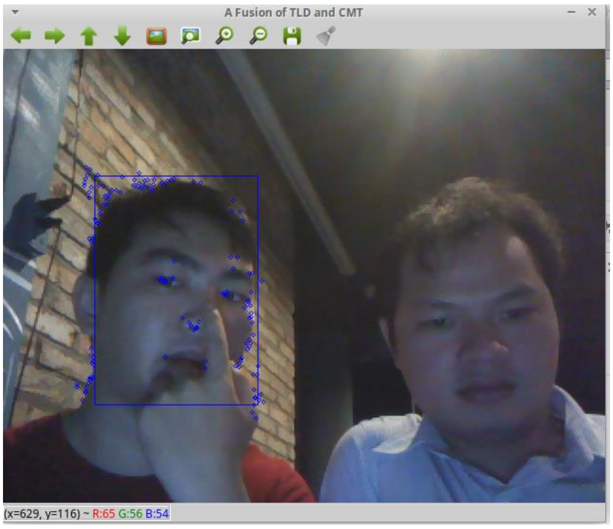
#305,Posterior 0,59; fps: 50,01

Figure 3. The TLD tracking system



(x=418, y=73) ~ R:87 G:11 B:23

Figure 4. The CMT tracking system



(x=629, y=116) ~ R:65 G:56 B:54

Figure 5. The Fusion TLD_CMT tracking system

All of the following tests were carried out on a standard laptop, a Dell Inspiron N5010 with a Core I5 460M processor running at 2,53 GHz, 6GB of RAM, a 240GB SSD, and Xubuntu 16.04 LTS installed. We utilize the Vojir tracking dataset [8], which consists of 77 sequences, for both tracking accuracy and calculation time.

TABLE II. Processed frames/sec of TLD, CMT, and Fusion TLD_CMT

#	Type	TLD	CMT	Fusion TLD_CMT
1	Ball	25	18	23
2	Board	7.7	0.5	2.7
3	Box	7.25	0.4	1.04
4	Cup on table	29	4	9.81
5	Dog 1	34	13	23.17
6	Gym	21	4.5	14.75
7	Lemming	10	0.44	1.08
8	Person partially occluded	34	5.1	25
9	Singer	20.9	0.54	1.71
	Average	**20.98**	**5.16**	**11.36**

The computation time for our proposed technique TLD CMT is always in the midway of TLD and CMT. Our TLD CMT technique achieves the same accuracy as CMT in deformable objects such as partially occluded people. Even if in certain cases in the live cameras system with actual data, our fusion TLD CMT can overcome the partially blindfolded object as shown in Fig. 5, our fusion TLD CMT can overcome the partly blinded object as shown in Fig. 5.

The result in Fig. 6 were computed on Vojir dataset for 3 methods (TLD, CMT, TLD_CMT) showing the success of tradeoff the accuracy of CMT to the computation time in our fusion proposal method in almost object types.

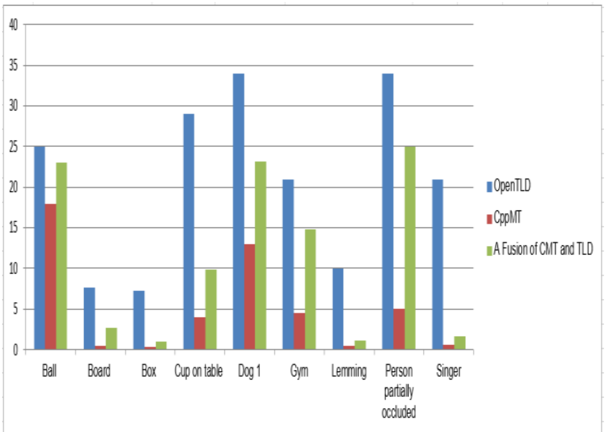


Figure 6. The Comparative of process time (frames/sec) one-by-one object types on Vojir dataset.

The average comparing of three methods is shown in Fig. 7.

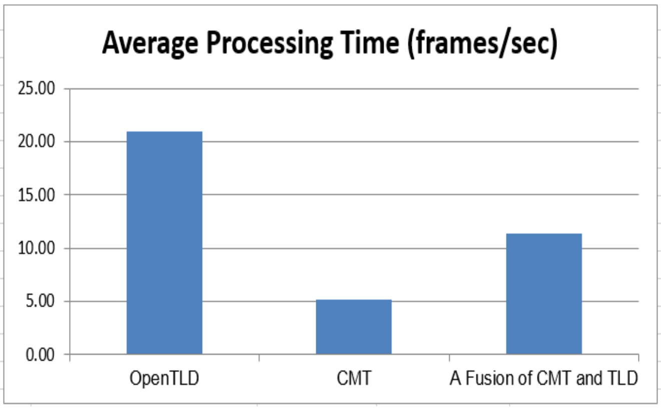


Figure 7. The Comparative of average process time (frames/sec) on Vojir dataset.

In general, our fusion approach takes more than twice as long as CMT, yet it maintains an accuracy of more than 80%, much like CMT (success plot compute all frames). It demonstrates the viability of using certain cameras for monitoring and tracking purposes.

V. CONCLUSION

We presented a fusion object tracking model that combined TLD and CMT in this study. In order to enhance CMT calculation speed, the proposed model featured TLD tradeoff partial accuracy CMT.

To restrict the calculation time of the CMT technique, the fusion TLD CMT model enhanced the TLD benefits of computation time and accuracy on t no deformable objects.

ACKNOWLEDGMENT (HEADING 5)

We thank all of our researcher team’s efforts and encouragements, as well as our kind colleagues at HCMUE in the project CS2019.19.02 and the anonymous reviewers whose comments and ideas were useful.

REFERENCES

[1] Hare, Sam, et al. "Struck: Structured Output Tracking with Kernels." (2014).

[2] Park, Eunae, et al. "Tracking-Learning-Detection Adopted Unsupervised Learning Algorithm." Knowledge and Systems Engineering (KSE), The Seventh International Conference on. IEEE (2015).

[3] Li, Xi, et al. "A survey of appearance models in visual object tracking." ACM transactions on Intelligent Systems and Technology (TIST) 4.4 (2013): 58.

[4] Erdem, Erkut, Séverine Dubuisson, and Isabelle Bloch. "Fragments based tracking with adaptive cue integration." Computer vision and image understanding 116.7 (2012): 827-841.

[5] Zhang, Kaihua, Lei Zhang, and Ming-Hsuan Yang. "Real-time compressive tracking." Computer Vision–ECCV 2012. Springer Berlin Heidelberg, 2012. 864-877.

[6] Nebehay, Georg, and Roman Pflugfelder. "Clustering of Static-Adaptive Correspondences for Deformable Object Tracking." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[7] A.T. Vo, T.Q. Le, H.S. Tran, and T.H. Le, “A Fusion TLD and CMT Model for Motion Object Tracking,” The International Conference on Information, System and Convergence Applications, July 13-16 in Chiang Mai, Thailand, ISSN 2383-479X, 2016, vol.3 no.1, pp. 60-63.

[8] Ciaparrone, Gioele, et al. "Deep learning in video multi-object tracking: A survey." Neurocomputing 381 (2020): 61-88.

[9] Lee, Sang Gu. "A Study on Utilizing Smartphone for CMT Object Tracking Method Adapting Face Detection." The Journal of the Convergence on Culture Technology 7.1 (2021): 588-594.

[10] Zhen, Xinxin, et al. "A visual object tracking algorithm based on improved TLD." Algorithms 13.1 (2020): 15.

[11] Li, Zhiyong, et al. "Learning a dynamic feature fusion tracker for object tracking." IEEE Transactions on Intelligent Transportation Systems (2020).

[12] Zdenek Kalal, Krystian Mikolajczyk, and Jiri Matas. "Tracking-Learning-Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 34, no 7, Page: 1409-1422, Jul. 2012.

[13] T. Vojir, and J. Matas, "The enhanced flock of tracker" In RRIV 2014.

[14] Guo, Jr-Hung, and Kuo-Lan Su. "Using Laser Range Finder and Multitarget Tracking-Learning-Detection Algorithm for Intelligent Mobile Robot." Sensors and Materials 27.8 (2015): 755-761.

[15] Staniszewski, Michał, et al. "Recent Developments in Tracking Objects in a Video Sequence." Intelligent Information and Database Systems. Springer Berlin Heidelberg, 2016. 427-436.

[16] Tran, Hai, et al. "Burn Image Classification Using One-Class Support Vector Machine." Context-Aware Systems and Applications. Springer International Publishing, 2015. 233-242.

[17] Li, Jiahe, Xu Gao, and Tingting Jiang. "Graph networks for multiple object tracking." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2020.

[18] Yu, Hongyang, et al. "Conditional GAN based individual and global motion fusion for multiple object tracking in UAV videos." Pattern Recognition Letters 131 (2020): 219-226.

[19] Abbass, Mohammed Y., et al. "A survey on online learning for visual tracking." The Visual Computer 37 (2021):

[20] Bouraya Jr, Sara, and Abdessamad Belangour. "Multi object tracking: a survey." Thirteenth International Conference on Digital Image Processing (ICDIP 2021). Vol. 11878. International Society for Optics and Photonics, 2021. 993-1014.

[21] Dutta, Anjan, et al. "Vision tracking: A survey of the state-of-the-art." SN Computer Science 1.1 (2020): 1-19.

[22] Karthik, Shyamgopal, Ameya Prabhu, and Vineet Gandhi. "Simple unsupervised multi-object tracking." arXiv preprint arXiv:2006.02609 (2020).

[23] Liang, Siyuan, et al. "Efficient adversarial attacks for visual object tracking." European Conference on Computer Vision. Springer, Cham, 2020.

[24] Qiu, Ji, et al. "Two motion models for improving video object tracking performance." Computer Vision and Image Understanding 195 (2020): 102951.

[25] Jiang, Shaokui, et al. "Faster and Simpler Siamese Network for Single Object Tracking." arXiv preprint arXiv:2105.03049 (2021).

[26] Ahmadyan, Adel, et al. "Objectron: A large scale dataset of object-centric videos in the wild with pose annotations." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.