A LITERATURE REVIEW ON VISUAL TRACKING WITH WEIGHTED ADAPTIVE LOCAL SPARSE APPEARANCE MODEL VIA SPATIO-TEMPORAL CONTEXT LEARNING

[1]N Santhosh Kumar, [2]Shaik Taj Mahaboob, [3]P G Varna Kumar Reddy

[1]PG Scholar, [2]Assistant Professor, [3]Assistant Professor (Adhoc)
[1]santhosh0436@gmail.com
[1]Electronics and Communication Engineering,
[1]JNTUA College of Engineering Pulivendula, Pulivendula, India

Abstract: We exploited to develop an effective appearance model for object tracking due to its well discriminative capability in distinguishing the target from its surrounding background. Most of these methods only consider either the holistic representation or the local one for each patch with equal importance, and hence may fail when the target suffers from severe occlusion or large scale pose variation. And the methods are Sparse Representation are using weighted local sparse appearance model, Integration of Spatio-Temporal Context into Appearance model using intensity levels. At last we conclude that, the explored the spatio-temporal context information to enhance the robustness of the appearance model, in which the global temporal context is learned via incremental subspace and sparse representation learning with a novel dynamic template update strategy, while the local spatial context considers the correlation between the target and its surrounding background via measuring the similarity among their sparse coefficients.

Index Terms - **Visual tracking; sparse representation; template update; spatio-temporal context.**

I. INTRODUCTION

This literature survey on the Visual Tracking with Weighted Adaptive Local Sparse Appearance Model via Spatio-Temporal Context Learning shows the main concern on continuous tracking the objects. Visual tracking which is an essential research topic in computer vision has broad applications in various fields, which inculcate self-driving cars, security and surveillance systems, and vision-based controls. Visual tracking continually infers the states of an annotated (manually labeled or detected in the first frame) target object in a video sequence. Despite of being studied for several decades visual tracking has made much progress in recent years. It remains a challenging task to develop a robust tracking algorithm due to severe variations in the appearance of tracked target caused by significant pose variation, complicated background clutter, drastic illumination variation, etc.

An effective appearance model ensures the robustness of a tracking system; this fact has attracted tremendous attention in recent years. Numerous effective representations have been proposed in order to design the appearance models that can be categorized into either generative or discriminative models. Generative models typically learn an appearance model to represent the target appearance and then use the model to search for the image region with maximal similarity. Generally, the representations for constructing generative appearance models include GMMs, color histograms, subspace representation and sparse representation.

In Jepson et al. proposed a GMM based representation with an online expectation maximization algorithm to overcome target appearance variations during tracking. In Adam et al., a set of local image patch histograms was utilized to represent a target object. In Ross et al. an incremental subspace learning method to learn a subspace representation that can adapt to the target appearance changes was proposed. In Kwon and Lee an observation model was decomposed into multiple basic observation models that are constructed using the sparse principal component analysis. In Wang and Yeung, a deep learning-based tracker that uses stacked de-noising auto-encoder to learn target presentations from a large number of in discriminative models tracking problem is cast as a binary classification task, which employ different discriminative features to train a classifier to separate the target from its surrounding background.

A visual tracking as a binary classification problem, which integrated an off-line SVM based classifier into an optical flow-based tracker was first formulated by Avidon. Collins et al. proposed a feature selection method to learn the most discriminative features online to separate the target object from the background. Babenko et al. [7] proposed to employ positive and negative bags to learn a multiple instance learning classifier for visual tracking. Zhang et al. [4] proposed a multi-expert restoration scheme to address the drift problem in tracking. Recently, Henriques et al. [8] proposed a fast tracker which exploits the circulant structure of the kernel matrix for kernelized correlation filters (KCF) that can be efficiently solved using fast Fourier transform algorithm. Li and Zhu improved the KCF tracker by integrating a scale adaptive scheme and colour-naming features. Ma et al. [1] employed features from hierarchical layers of convolution neural networks (CNNs) to learn an effective KCF representation for robust visual tracking.

Recently, sparse representations have been widely exploited in visual tracking which can be categorized into holistic and local sparse representation appearance models. In Mei and Ling exploited a holistic sparse representation of the target appearance for visual tracking that is learned through optimizing a `1 minimization problem. This work was extended by Li et al. using the orthogonal matching pursuit algorithm to solve the optimization problem efficiently and further improved in the efficiency by Bao et al. via the accelerated proximal gradient approach. However, these sparse representation-based trackers take into account the holistic templates of the targets, which are sensitive to severe partial occlusion and pose variations. In Liu et al., a local sparse appearance model was proposed that is integrated into the mean shift algorithm to enhance tracking robustness. However, this tracker is based on a static local dictionary obtained from the first frame and has a high probability of failing in dynamic scenes. Jia et al. presented a local sparse appearance model that employs an alignment-pooling method to combine the histograms of local sparse codling's of each patch, where the dictionary is updated in an online manner to handle target appearance variations, thus achieving favorable performance on some challenging scenes.

Although demonstrated success of the trackers based on local sparse appearance model their performance on the recently tracking benchmark is not favorable. For example, as reported by the benchmark, the AUC score of success plots of OPE for the ASLA tracker is just 0.434, which is much lower than the KCF based trackers, e.g., 0.514 for KCF and 0.567 for SAMF. We note that the performance of ASLA tracker still has a long way to go in improving if we take into account the importance of different local patches and attempt to integrate the spatio-temporal context information.

In this paper, we propose a simple yet effective method by combining the weighted local sparse model and spatio-temporal context information. The proposed method is motivated by the ASLA tracker, but takes into account the patch importance to measure the reliability of each patch with a reconstruction-error-based weight function of reconstruction error. Further, we employ the spatio-temporal context information, incorporating the global temporal context via an incremental subspace and sparse representation learning with a novel online template update strategy and the local spatial context considering the correlation between the target and its surrounding background.

II. SYSTEM OVERVIEW

The process of sparse representation involves with the video tracking with the help of adaptive weight. This paper consists of three main steps. In that, first is pre-processed the input image to obtain image that have a normalized size or intensity, second is sparse representation using adaptive weights and the last stage is spatio-temporal context learning. In this paper we track the object in the above three methods

A. Pre-processing of object

The object detection is the process of extracting the object from the background. It means to determine the position of object the in the image. This step we require the in images in different scales. In the input image having various illumines condition and more back ground modes, while tracking these modes will be in confusion state. Object recognition tends to fail if the test image has a different lighting condition than that of the training images.

B. Sparse Representation

The pre-processing of a video that changes the complex background and moves to the stage called sparse representation. In that model the image of the prescribed video is divided in to several patches. The patches of the image are sparsed with the help of local spatial context. For every sparse representation we have to give adaptive weights for better performance of the object. We can decrease the occlusion percentage in that object. for every frame in the object adaptive weights is given, for the every patch numbering is given, for every adaptive weight to the each frame the noise is deceased with the local sparse model patches.

C. Spatio-Temporal Context Integration

This method is mostly used for robust tracking .The target appearance changes between two consecutive high rate frames (about 25 frames), and it is also used for historical appearance variations. In the spatio-temporal context integration mainly are of two methods, one is global temporal context and the second method local temporal context learning. In the global temporal is used for background of an image by varying the intensity level and the local temporal is used for foreground or tracking object with the help of intensity levels. And finally we have an algorithm called tracking algorithm works on the principle of Bayesian theorem.

Table 1

A Summary of Visual Tracking Techniques

Sl. No.	Title	Author	Method/Algorithm Implemented	Advantages	Disadvantages
1	Real-time visual tracking using compressive sensing	H. Li, C. Shen, And Q. Shi. [6]	RTCST (Real time compressive tracking) by exploiting the signal recovery power of compressive sensing.	Results on various video sequences are verified by a new metric-Tracking Success Probability (TSP),	High computational complexity.
2	Robust tracking via multiple experts using entropy minimization	J. Zhang, S. Ma and S. Sclarof [2]	SVM on a budget algorithm and benchmark dataset of 50 video sequences	significantly improves the robustness of our base tracker, especially in scenarios with frequent occlusions	SVM approach lies in choice of the kernel
3	Robust object tracking via sparsity based collaborative model,"	W. Zhong,H. Lu, and M.-H. Yang [3]	Sparsity-based discriminative classifier (SDC) and a sparsity-based generative model (SGM).	an effective method to compute the confidence value that assigns more weights to the foreground than the background	Cost of explicitly multiplying each input by zero and adding up all those zeros
4	Robust object tracking by online fisher discrimination boosting feature selection	J. Yang. Zhang and Q. Liu [4]	online Fisher discrimination boosting feature selection mechanism	enhancing the discriminative capability between the target and background	Small amount of occlusion is reduced.
5	Visual tracking via boolean map representations	K. Zhang, Q. Liu, J. Yang, and M.-H. Yang [5]	Feature map of the intersection kernel, logistic regression classifier and tracking methods on the OTB50 and VOT2016 benchmark datasets.	This representation scheme is computationally efficient and facilitates achieving favorable performance in terms of accuracy and robustness	VOT2016 benchmark datasets are used.
6	Robust object tracking with online multiple instance learning	B. Babenko, M.-H. Yang and S. Belongie [7]	We use Multiple Instance Learning (MIL) algorithm.	object tracking that achieves superior results with real-time performance	Intensity levels of background and foreground can't be varied.

III. CONCLUSION

In this paper, the review of framework of sparse representation has been highlighted and presents a literature survey on the various techniques involved in spatio temporal context learning. These methods are measured on the basis of area under curve for selection and precision plots. WALSA was the better algorithm in overall sparse representation   for visual tracking.

REFERENCES

[1] B. Ma, J. Shen, Y. Liu, H. Hu, L. Shao, and X. Li, "Visual tracking using strong classifier and structural local sparse descriptors," IEEE Transactions on Multimedia, vol. 17, no. 10, pp. 1818–1828, 2015.

[2] J. Zhang, S. Ma, and S. Sclaroff, "Meem: Robust tracking via multiple experts using entropy minimization," in Proceedings of European Conference on Computer Vision, pp. 188–203, 2014.

[3] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity based collaborative model," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1838–1845, 2012.

[4] J. Yang, K. Zhang, and Q. Liu, "Robust object tracking by online fisher discrimination boosting feature selection," Computer Vision and Image Understanding, vol. 153, pp. 100–108, 2016.

[5] K. Zhang, Q. Liu, J. Yang, and M.-H. Yang, "Visual tracking via Boolean map representations," Pattern Recognition, vol. 81, pp. 147 – 160, 2018.

[6] H. Li, C. Shen, and Q. Shi, "Real-time visual tracking using compressive sensing," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1305–1312, 2011.

[7] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 8, pp. 1619–1632, 2011.

[8] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High speed tracking with kernelized correlation filters," IEEE Trans.Pattern Anal.Mach. Intell., vol. 37, no. 3, pp. 583–596, Mar. 2015.