Discriminative Object Tracking Using Two-stage Classification

Kaushik C. M. [1], *Lasitha Mekkayil[2] and Hariharan Ramasangu[3]

Department of Electronics and Communication Engineering, Faculty of Engineering and Technology,

Ramaiah University of Applied Sciences, Bengaluru-560058

*Contact Author E-mail: lasitha.ec.et@msruas.ac.in

Abstract

An efficient online tracking algorithm to sketch the combined direction and orientation of an object from a video is an arduous task. Most of the discriminative trackers learn from the previous frame output to estimate object in given frame. Many use single stage classifier for classification and the output of the classifier will be the final output in most of the cases. The presence of misclassification by the classifier can ead to failure of the algorithm. To overcome this problem, Two-stage classification is proposed in this paper, where the framework of the algorithm is the novelty. In the proposed algorithm, a Linear SVM classifier is used as a first stage classifier and a Laplacian Regularized Least square learning in Bayesian learning framework is used as a second stage classifier. Using the first stage classification, a set of positive features are selected and by using the second stage classification the tracker location is selected. Hence by doing so, the burden on the classifier is shared and misclassification is reduced considerably. The proposed algorithm is an efficient online tracker, which can handle most of the challenges and performs well in different attributes of selected standard 17 Dataset. The proposed algorithm has improved precision rate (74.1%) and success rate (54.7%) compared with all the available diverse range of trackers.

Key Words: Discriminative Tracking; Two-Stage Classification; Visual Object Tracking

1. INTRODUCTION

Despite having dissimilarity like jumble background, the difference in object position, pose and light, human vision can effortlessly and rapidly identify huge quantity of different objects present in the scenario. Human vision can easily segment an object, examine its nature and find the trajectory of the same. But it is extremely hard to build a robust system to function exactly as human vision. In recent decade computer vision has become pervasive due to its ability to model human vision. Significant advances in this field are also been made, but a lot of problems still remains to be solved. Due to these reasons and wide range of applications in industries, many researchers are getting involved in the area of computer vision. Few applications of computer vision are automation, surveillance, human-robot interaction, video indexing, vehicle navigation, augmented reality, diagnostic biomedical imaging and video editing. One of the technical concept common in above mentioned applications is visual object tracking. In its elementary form, we can say that visual tracking is the sketch of the path of a target from all the frames of a video. In pragmatic scenarios, visual object tracking can be considered as a challenging task owing to factors such as jumble background, deviation in viewpoints, occlusion, illumination and scale changes.

Recently varied variety of effective representation schemes have been initiated for the robust tracking of the object of interest, such as real-time compressive tracking [22], incremental visual tracking [2], visual tracking decomposition [10], tracking via sparsity-based model [7], multiple instance learning [12], etc. Altogether uses wide variety of feature extraction and classification methods, but still fails to track the object effectively. These algorithms use features such as colour, pixel values, histogram, texture, haar-like features [23] and classifiers includes naïve bayes [23], adaboost [21], svm[1], regularised least squares [9] etc.

2. RELATED WORK

There is a rich literature available on tackling tracking problem. In this section, we discuss the literature review conducted. Interested readers may refer to [8] [18] and [19] for a comprehensive review.

All the tracking algorithms proposed falls into a pair of categories: generative and discriminative. "Generative methods are selected to trace an object based on the similarity index of the target object to be tracked or we can say it is based on an object model. Discriminative tracking is a binary based feature classification problem that discriminates the target from the scene with respect to the features to be extracted" [8] [18] [19].

Discriminative tracking has received wide attention for its adaptive ability to handle appearance changes. The essential component of discriminative trackers is the classifier learning. Classifier learning can be either supervised, multiple instances learning [12] or semi-supervised [9]. However in supervised learning one positive sample (i.e, tracking result in the current frame) and multiple negative samples are used to update classifier. If the object location is not perfectly detected by the current classifier, the appearance model would be updated with a suboptimal positive example. Over time the accumulation of errors can degrade the classifier, and can cause drift. In case of semi-supervised learning scheme proposed by gardner [6] classifiers is trained with labelled samples for the first frame and leaving subsequent samples of next frame unlabeled. Thus semi-supervised learning is effective against drift and other problems. Saffari et.al [11] proposed multi-view boosting algorithm which considers the given priors as a regularization component over the unlabeled data and validated its robustness for object tracking. Gao et al. [5] employed the cluster assumption to exploit data to encode discriminant information of their representation, due to which tracker performance is improved. Kalal et al.[21] developed a p-n learning method to training a binary classifier with structured unlabeled data.

An online structured SVM based [4] robust tracking is proposed to reduce the effect of wrong labelling of samples. Recently a semi-supervised active learning with Laplacian Regularized Least Square (LapRLS) classification with Bayesian inference framework [9] is proposed and it performs well during various attributes. The algorithm considers object's motion model with Histogram Oriented Gradient (HOG) [4] feature extraction technique. The LapRLS classifier learning in Bayesian framework includes an active example selection procedure for tracking. The problem with AEST [9] is that while selecting samples features for the active example selection previous frame tracker result is considered. So if there is a misclassification in the previous frame output, samples generated for current frame output will be slightly deviated from required samples. This will degrade trackers performance. To overcome this problem a two-stage classification is proposed. Wherein the first stage, using SVM classifier negative samples are completely neglected and set of positive samples are considered for second stage classification. In second stage classification LapRLS classifier learning in bayesian framework is used to select a best positive sample from a collection of positive samples. By incorporating proposed method comparatively better results are obtained and discussed in further sections.

2.1 Tracking based on active example selection

The AEST [9] algorithm explains tracking based on a two-stage strategy. During stage I, Histogram oriented features are selected and by using an active example selection strategy training samples are selected. In stage II, the algorithm uses a LapRLS classifier for classification and tracking is performed based on a Bayesian Inference Framework.

Fig 1 Proposed Discriminative Tracker with Two-Stage Classification

3. PROPOSED METHOD

The Proposed block diagram is shown in Fig 1. In most of the visual object tracking algorithms single stage classifiers are used, where the classifier output will be final tracker output. If there is a misclassification occurred due to classifier, tracker output will be affected. . To overcome this problem, a two stage classification process is proposed. Here SVM classifier will be the first stage and LapRLS classifier learning in Bayesian frame work will act as second stage classifier. Novelty of algorithm lies in the frame work of algorithm.

3.1 SVM Classifier

"SVM classifiers find a separating hyperplane that maximizes the margin between the two classes, where the margin is defined as the distance of the closest point, in each class, to the separating hyperplane. This is equivalent to performing structural risk minimization to achieve good generalization" [1]

"Given a data set '{Xi,Yi}' of 'l' examples 'Xi' with labels 'Yi ∈{+1,-1}', finding the optimal hyperplane implies solving a constrained optimization problem using quadratic programming, where the optimization criterion is the width of the margin between the classes. The separating hyper-plane can be represented as a linear combination of the training examples and classifying a new test pattern 'x' is done using the equation 1": [1]

$$F(x) = \sum_{i=1}^{l} \alpha_i y_i K(X, X_i) + b \qquad (1)$$

"In equation (1) 'y_i' is labels of samples, 'K(X, X_i)' is a kernel function and the sign of F(x) determines the class membership of 'X'. Constructing the optimal hyperplane is equivalent to finding the nonzero 'α_i'. Any data point to 'X_i' corresponding to nonzero 'α_i' is termed as support vector." Support vectors are the training patterns closest to the separating hyperplane and the kernel function extends SVM to handle nonlinear separating hyperplanes. Popular kernel function used is a Gaussian RBF kernal [1].

The SVM classifier will be trained from positive and negative samples of previous frame ((t-1)th) obtained for classifier learning. Using classifier un-labelled sample of present frame (tth) will be classified and labelled as positive and negative. By using an SVM classifier only positive labelled features are chosen and feed into LapRLS classifier learning in Bayesian frame work classifier. Here Bayesian frame work classification is trained with previous frame outputs, to classify one sample patch as output from multiple patches. By doing so misclassification has been reduced, which intern helps to boost accuracy and precision of visual object tracking algorithm.

4. EXPERIMENTAL ANALYSIS AND RESULT

In our experiments, the target is initialized based on the starting frame ground truth values. The optimal feature selection algorithm is evaluated with 11 trackers [13], where each algorithm is verified using 17 challenging videos (6003 frames). The trackers comprises MIL [12], VTD [7], TLD [21], AEST [9], Struck [11], SCM [14], CT [22], SPT [20], LSST [17], RET [3] and ONNDL [16]. The proposed tracker is implemented in MATLAB (2013a), which runs at 0.8 Frame per second on an Intel Core i3 processor 2.2 GHz PC. The Dataset includes CarDark, Crossing, Couple, David1, David2, David3, Deer, Football, Football1, Freeman1, MountainBike, Singer2, Skating, Subway, Sylvester, Trellis and Tiger1. Challenges include Variations in Illumination and Scale, light and heavy Occlusion, Deformation of the object, Motion Blur due to movement of the object, In and Out of Plane Rotation, Clutters in Background and Low Resolution.

4.1 Quantitative Evaluation

The quantitative evaluations measures include centre location error and the success rate. The centre location error is the distance between the midpoints of both the ground truth and the tracker output. The success rate is defined as:

$$s(R_T, R_G) = \frac{area(R_T \cap R_G)}{area(R_T \cup R_G)} \quad (2)$$

Where 'R_T' denotes the output of multi stage classification algorithm and 'R_G' denotes the ground truth. The optimal feature selection algorithm make use of precision plot and success plot [13] to assess the effectiveness of various algorithms.

The overall performances of all the algorithms on 17 sequences are discussed by the plots as shown in Fig. 2 and Fig. 3. From precision plot (Fig. 2) we can infer that the performance of the proposed algorithm (74.1%) is improved compared to selected algorithms such as AEST (71.5%), followed by SCM (66.9%) and ONNDLT (66.5%). From success plot (Fig. 3) we can infer that proposed multi-stage classifier technique (54.7%) gives better success rate followed by AEST (53.8%), Struck (52.7%) and SCM (52.4%).

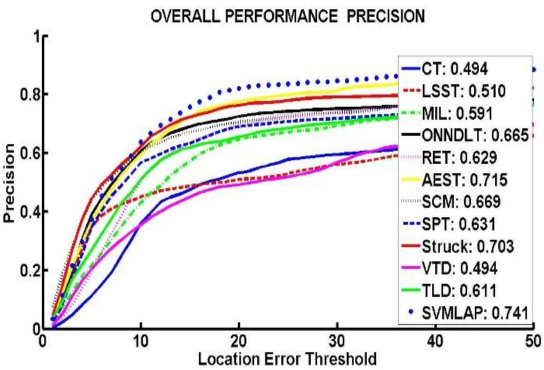

Fig. 2 Overall Precision plot showing the performance of the competing trackers on selected 17 video sequence.

4.2 Attribute based study

The dataset mentioned in the benchmark are equipped with wide range of attributes. These attributes are the elements that can alter the performance of the tracker. In our experiments, we employ the attribute based study to illustrate the effectiveness of the proposed algorithm compared with other state-of-the-art-trackers.

Precision: Out of 10 different attributes proposed two-stage classifier algorithm is outperforming AEST in 7

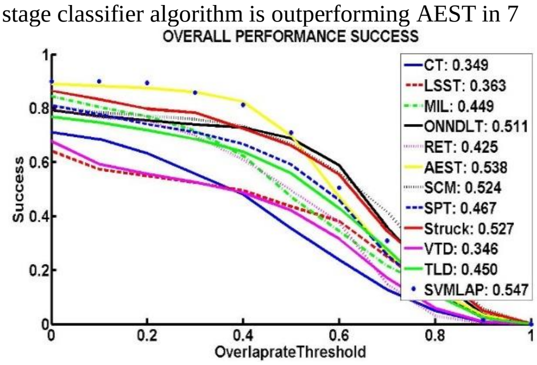

Fig. 3 Overall Success plot showing the performance of the competing trackers on selected 17 video sequence.

attributes such as Background clutter (proposed method-77.3% and AEST-76.3%), Deformation (proposed method-65% and AEST-61.4%), Illumination Variation (proposed method-64.3% and AEST-60.5%), In-plane rotation (proposed method-69.0% and AEST-65.6%), occlusion (proposed method-68% and AEST-67.4%), out of plane rotation (proposed method-73% and AEST-67.9%) and Scale variation (proposed method-79.3% and AEST-73.6%). In attributes like low resolution, Motion blur and abrupt motion AEST performs better than proposed algorithm. In attributes like Background clutter, Illumination variation, In-plane rotation, Scale variation, Deformation, Occlusion and Out of plane rotation proposed method outperforms all the other trackers in precision rate.

Success Rate: Out of 10 different attributes proposed algorithm perform better than AEST in four different attributes. In case of Illumination variation proposed algorithm and AEST give a success rate of 50.2% and 49.7% respectively. In case of In-plane rotation (proposed algorithm-51.5% and AEST-50%), out of plane rotation (proposed algorithm-54.4% and AEST-52.1%) and in Scale Variation (proposed algorithm-56.2% and AEST-55.5%) the proposed algorithm is able to provide better results compared to AEST algorithm. In other attributes, AEST gives better success rate compared to proposed method. In attributes like Out of plane rotation and In-plane rotation proposed algorithm outperforms most of the other trackers. In attributes like Background clutter, Illumination variation, Scale variation, Deformation, and Occlusion proposed method holds a position of second best compared with other trackers.

5. CONCLUSION

In this paper, we have proposed a two-stage classification algorithm with linear SVM as first stage classifier and LapRLS (laplacian regularized least square) classifier learning in Bayesian framework as second stage classifier. The success of the proposed algorithm is analysed using a wide variety of dataset. The results obtained shows the efficiency of the tracker in terms of accuracy and precision over the competing trackers. The algorithm is also evaluated quantitatively based on various attributes. The proposed algorithm is able to track the object perfectly and can also drastically reduce the error generated due to misclassification of classifiers.

REFERENCES

[1] Avidan, S. and Shamir, A. (2007) Seam carving for content-aware image resizing, *In ACM Transactions on graphics (TOG)*, Vol. 26, No. 3, p. 10.

[2] Babenko, B., Yang, M. H., & Belongie, S. (2011). Robust object tracking with online multiple instance learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8), 1619-1632.

[3] Bai Q., Wu Z., S. Sclaroff, M. Betke, and C. Monnier, (2013) Randomized ensemble tracking, *in ICCV*, pp. 2040–2047.

[4] Dalal N. and Triggs B., (2005) Histograms of oriented gradients for human detection, *in CVPR*, pp. 886–893.

[5] Gao, J., Xing, J., Hu, W., & Maybank, S. (2013). Discriminant tracking using tensor representation with semi-supervised improvement. *In Proceedings of the IEEE International Conference on Computer Vision* (pp. 1569-1576).

[6] H. Grabner, C. Leistner, and H. Bischof, (2008) Semi-supervised on-line boosting for robust tracking, *in Proc. Eur. Conf. Comput. Vis.*, pp. 234–247

[7] Kwon, J., & Lee, K. M. (2010, June). Visual tracking decomposition. *In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on* (pp. 1269-1276). IEEE.

[8] Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A. and Hengel, A.V.D., (2013). A survey of appearance models in visual object tracking. *ACM transactions on Intelligent Systems and Technology* (TIST), 4(4), p.58.

[9] Min Yang, Yuwei Wu, Mingtao Pei, Bo Ma and Yunde Jia, (02 July 2015). Online Discriminative Tracking with Active Example Selection. *Circuits and Systems for Video Technology, IEEE* (Volume: PP , Issue: 99)

[10] Ross, D. A., Lim, J., Lin, R. S., & Yang, M. H. (2008). Incremental learning for robust visual tracking. *International Journal of Computer Vision*, 77(1-3), 125-141.

[11] S. Hare, A. Saffari, and P. H. Torr, (2011) Struck: Structured output tracking with kernels, *in ICCV*, pp. 263–270.

[12] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, (2012) Robust visual tracking via multi-task sparse learning, *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2042–2049.

[13] Wu, Y., Lim, J. and Yang, M.H., (2013). Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2411-2418).

[14] W. Zhong, H. Lu, and M.-H. Yang, (2012) Robust object tracking via sparsitybased collaborative model, *in CVPR*, pp. 1838–1845.

[15] Wu, Y., Lim, J. and Yang, M.H., (2013). Online object tracking: A benchmark. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2411-2418).

[16] Wang N., J. Wang, and D.-Y. Yeung, (2013) Online robust non-negative dictionary learning for visual tracking, *in ICCV*, pp. 657–664.

[17] Wang, D., Lu, H., & Yang, M. H. (2013). Least soft-threshold squares tracking. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2371-2378).

[18] Yilmaz, A., Javed, O. and Shah, M., (2006). Object tracking: A survey. Acm computing surveys (CSUR), 38(4), p.13.

[19] Yang, H., Shao, L., Zheng, F., Wang, L. and Song, Z., (2011). Recent advances and trends in visual tracking: A review. Neurocomputing, 74(18), pp.3823-3831.

[20] Yao.R , Shi, Shen, Y. Zhang, and A. van den Hengel, (2013) Part-based visual tracking with online latent structural learning, *in CVPR*, pp. 2363–2370.

[21] Z. Kalal, J. Matas, and K. Mikolajczyk, (2010) Pn learning: Bootstrapping binary classifiers by structural constraints, *in Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 49–56

[22] Zhang K., L. Zhang, and M.-H. Yang, (2012) Real-time compressive tracking, *in ECCV*, pp. 864–877.

[23] Zhang, K., Zhang, L., & Yang, M. H. (2014). Fast compressive tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(10), 2002-2015.