CSE 252C Project Proposal
Re-implementation of Struct: a Structured Output Tracking with Kernels

Mike Liu, Yuhang Ming, and Jing Wang

May 15, 2017

1 Introduction

This project aims to re-implement Struck[1], an state of art tracking technique to track single object using single camera. Although the traditional approach to tracking, (adaptive) tracking-by-detection, has become particularly popular, there are multiple issues raised by the separation of the generation and labeling of samples and the updating of the classifier. The approach we are going to implement in our project, Struck, frames the whole tracking problem as a structured output prediction. It directly predicts the changes in object location between frames, thus avoiding the need for ad-hoc update strategies.

The rest of this document will list the milestones we hope to meeting in the next few weeks and some questions we hope to answer by the end of the project.

2 Milestones

In this section, we set 5 milestones that we will keep tracking when implementing the project.

1. Feature Extraction
   We will use Haar feature[2] to detect the initial position of the object. The Haar feature extraction code will be implemented no later than May 7th, 2017.

2. Gaussian Kernels
   By May 7th, 2017, we will implement the Gaussian kernel which corresponds to the Haar feature we use.

3. LaRank[3]
   We will use LaRank approach, which is an SVM solver based on the sequential minimal optimization(SMO)[4], to perform online optimization the the SVM. The LaRank algorithm will be implemented by May 17th, 2017.

4. Budget Maintenance[5]
   The budget maintenance mechanism is used to prevent the so-called "curse of kernelization", which is the unbounded growth of the number of support vectors. We will implement this mechanism no later than May 21st, 2017.

5. Multi-Kernels
   We will investigate into multi-kernel learning, as well as feature selection, to improve the tracker's performance. The investigation will be done by May 31st, 2017.

3 Questions

In this section, we will list 4 questions which will be answered during the implementation of this project.

1. What is Struck's time and space complexity, and scalability?
   Since Struck uses an online structured output SVM learning framework, one potential problem is the unbounded growth in the number of support vectors. The paper[1] claims to have introduced an budget maintenance heuristic by limit the number of support vectors to allow real-time tracking. Although a budget would upper bound the time and space complexity, heuristics come with trade-offs in accuracy, so we are interested to see the ratio between budget values and accuracy lost.

2. How does Struck perform compared to other tracking methods?
   As mentioned above, the Struck tries to improve the tracking performance by framing the problem as a structured output prediction. During the implementation, we want to compare the performance of Struck with other tracking methods, like MIForest[8], OMCLP[9], MIL[10], etc, to see whether it indeed improves the performance.

3. What are the observed limitations of Struck?
   Since the Struck is used to track single object using single camera, it may fail while trying to track multiple objects. We would like to investigate its limitation in multiple objects tracking task. Additionally, in single object tracking tasks, we also wonder its performance in some special situations like the absence of the object during the tracking.

4. What are potential ways to further improve Struck's performance?
   In [1], the authors give a short discussion about how multiple kernel learning can potentially result in tracking performance improvement. Therefore, during the implementation, we wonder how the various combination of kernels and features can improve the tracking performance. We also would like to investigate other methods which may lead to better performance.

4 Experiments

Our project will use OpenCV 2.4.13.2 and Eigen 3.2.8 to build our model. The core and highgui module of OpenCV to make use of the tracking sequences of images. The core module of Eigen is then used to speed up the calculation on image processing.

The sequences of images, which are available in the dataset of Visual Tracker Benchmark[6], is used to train and test the tracking model with different budget size.

To assess tracking performance, we will used Pascal VOC overlap criterion for tracking bounding boxes, the same evaluation criterion as the one presented in the paper. To be considered a correct detection, the area of overlap a_o between the predicted bounding box B_p and ground truth bounding box B_gt must exceed (50%) by the formula:

$$a_o = \frac{area(B_p \cap B_{gt})}{area(B_p \cup B_{gt})} \tag{1}$$

Where $B_p \cap B_{gt}$ denotes the intersection of the predicted and ground truth bounding boxes and $B_p \cup B_{gt}$ the union of the bounding boxes. [7]

We want to show that the performance of our implementation matches the results shown in the the original Struck paper. We also want to test and answer the questions we brought up in Section 3.

References

[1] Hare, Sam, Amir Saffari, and Philip HS Torr. "Struck: Structured output tracking with kernels." Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011.

[2] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137-154.

[3] Bordes, Antoine, et al. "Solving multiclass support vector machines with LaRank." Proceedings of the 24th international conference on Machine learning. ACM, 2007.

[4] Bordes, Antoine, Nicolas Usunier, and Léon Bottou. "Sequence labelling SVMs trained in one pass." Machine Learning and Knowledge Discovery in Databases (2008): 146-161.

[5] Crammer, Koby, Jaz S. Kandola, and Yoram Singer. "Online Classification on a Budget." NIPS. Vol. 2. 2003.

[6] Wu, Yi, Jongwoo Lim, and Ming-Hsuan Yang. "Online object tracking: A benchmark." Proceedings of the IEEE conference on computer vision and pattern recognition. 2013.

[7] Everingham, M., Van Gool, L., Williams, C.K.I. et al. Int J Comput Vis (2010) 88: 303. doi:10.1007/s11263-009-0275-4

[8] Leistner, Christian, Amir Saffari, and Horst Bischof. "MIForests: Multiple-instance learning with randomized trees." Computer Vision–ECCV 2010 (2010): 29-42.

[9] Saffari, Amir, et al. "Online multi-class lpboost." Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010.

[10] Babenko, Boris, Ming-Hsuan Yang, and Serge Belongie. "Visual tracking with online multiple instance learning." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.