Paper Information (Reference)

Title: Real Time Robust L1 Tracker Using Accelerated Proximal Gradient Approach

Author: Chenglong Bao, Yi Wu, Haibin Ling, and Hui Ji

Year: 2012

Journal: CVPR

Paper url: http://ai2-s2-pdfs.s3.amazonaws.com/f4a9/f457d85165ccb6a0b31c6dba125a38ca98c9.pdf

Code url: http://www.dabi.temple.edu/~hbling/code_data.htm (First paper from this url)

Dataset: https://sites.google.com/site/wuyi2018/benchmark

Paper Summary

This paper is based upon a development of a visual tracker called L1 tracker for tracking objects in real time which is upgraded model of a previously built in L1 tracker. The previous tracker was using the interior point method which happens to be slow, a new $l1$ norm related minimization model is presented in this paper which overcomes the problem of the previous model. The other problem in the previous tracker was that it losses tracking accuracy due to the inclusion of trivial templates (image occlusion and noise data) in the template dictionary. The second contribution of this paper is a numerical method is presented which solves the problem of $l1$ norm minimization very fast and gives a real time L1 tracker. Target and trivial templates are contained by the template dictionary whereas target template refers to object which is being tracked and it updated dynamically while trivial template refers to noise, background and occlusions. For the optimization of the previously L1 tracker, in this paper they have introduced energy coefficient contains have small value if occlusion can be neglected but if there are a large number of occlusion then energy coefficient contains too large values. Accelerated proximal gradient (APG) approach is used for solving proposed minimization problem.

Reproducibility Overview

Code

The code is written in matlab and it is available. To run the code in the matlab one of its dependency needs to be installed that is “Installmex” and then first compile the "IMGaffine_c.c" code. This is also described in the documentation file of the code (Readme.txt). I installed “MinGW-w64 compiler 4.9.2 version 16.1.0” this on my 64 bit system to run c language code. You can install it by using matlab add-ons feature.

Data

In this paper dataset is in the form of sequence of images and the paper has used 8 types of datasets (jump,car4,singer,woman,pole,sylv,deer and face) and out of those datasets only one dataset (“car4”) is provided with the code. For other datasets I emailed all the 4 authors and of which 2 emails were not working. While out of 2 successful emails I got reply from one author that dataset

can be found from the referred paper. I found the other datasets from the following url: https://sites.google.com/site/wuyi2018/benchmark.

VMs/Containers
No any virtual machine or container provided with the paper.

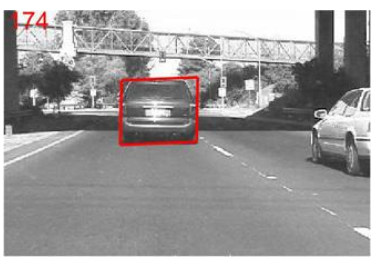
Workflows
There is no workflow provided and paper does not describe any workflow used in the study.

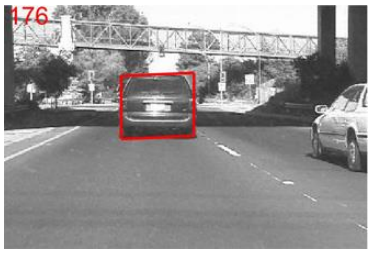
Provenance
The L1 tracker developed in this research study is an enhancement of the previously developed L1 tracker and authors have described the limitations of previous system and significant changes and updated the previous L1 tracker.

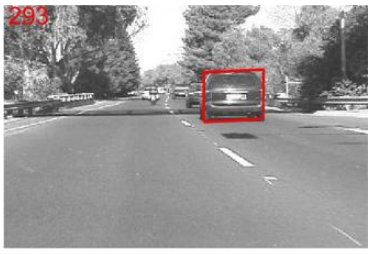
Other Reproducibility Information
In this paper authors have compared their work with 5 other papers but they have not provided the code of other papers. They have made a table 1 which is the result of the comparison but for reproducing that tablet data is not provided and the method they followed to produce that table is also not provided in detail, only some detail is given in table caption. Similarly in figure 4, pictorially efficiency of algorithm is shown but it cannot be fully reproduced similarly as it is provided because for the tracking of an image sequence one needs to define the initial frame size of the target in order to track the object through the images but author has not provided the frame size of the datasets he used in his paper except the "car4" dataset and it becomes difficult to give the exact frame size as author has given, to get the similar results.

Reproducibility Rating
I would rate this paper 3 out of 5 because the code can be ran, dataset can be acquired and results can be seen but for reproducing the table, figures and graph of the paper one need to have the corresponding details and dataset which is not provided and authors are also not responding on the emails so I am unable to communicate with the authors and get the dataset or information to reproduce the paper.

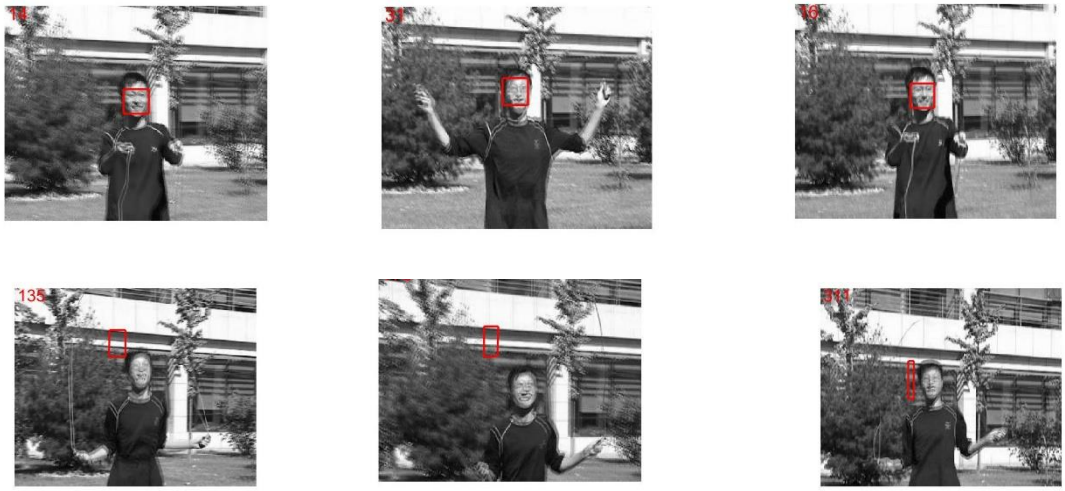
Experimental Work
Initial frame size of only "car4" dataset is given, when I ran the code I got the following results. I have taken the 6 images which author have shown in figure 3 in the paper. Tracker tracked the car similarly as it is shown in the paper.



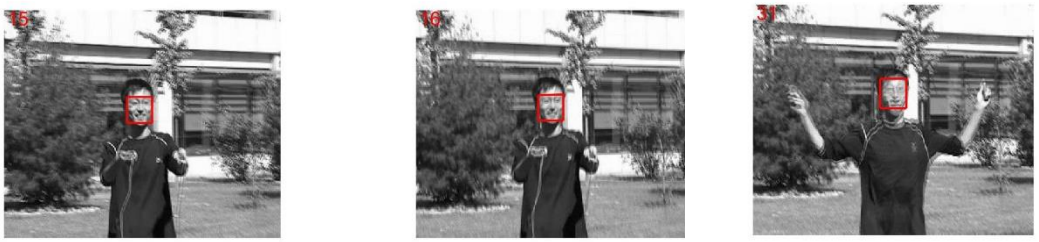




For the other datasets I used the initial frame size for the target which was given in the ground truth file. I tried different frame size when results were not coming similar to figure 3 of the paper but still except ‘car4’ results were coming different. I have taken frame size of image 1 and image 13 of each dataset from the ground truth file for tracking the object. I took image 13 because in the ‘car4’ dataset author have started tracking from image 13 and there is no details given about that. So for testing, first I took frame size of image 13 and then image 1.
In the jump dataset I got the following results which were not matching accurately with the given results in the paper. First started tracking from image number 13 and it can be seen that tracker was not able to track the 135,176,311 images.

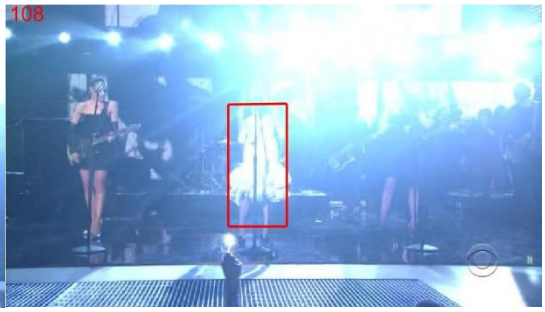


Then I started tracking from first image but results were similar as tracker was not able to track the 135,176,311 images.

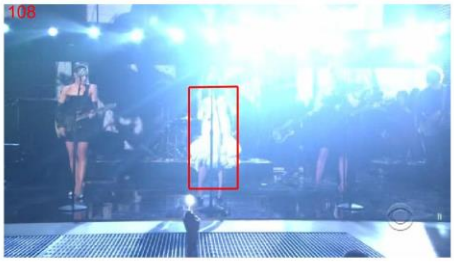


c) singer

Similarly I ran the code for 'singer' dataset and got the following results. First tracking started from image 13. In this tracker was not able to track the 4(108) and 5(133) images.



Then I started tracking from first image and as the result shows there is no much difference.



Similarly I tried on all 8 datasets but results were not reproduced accurately because actual frame size and original dataset was not provided. Below figures shows the results of all datasets whose tracking started from image 13.



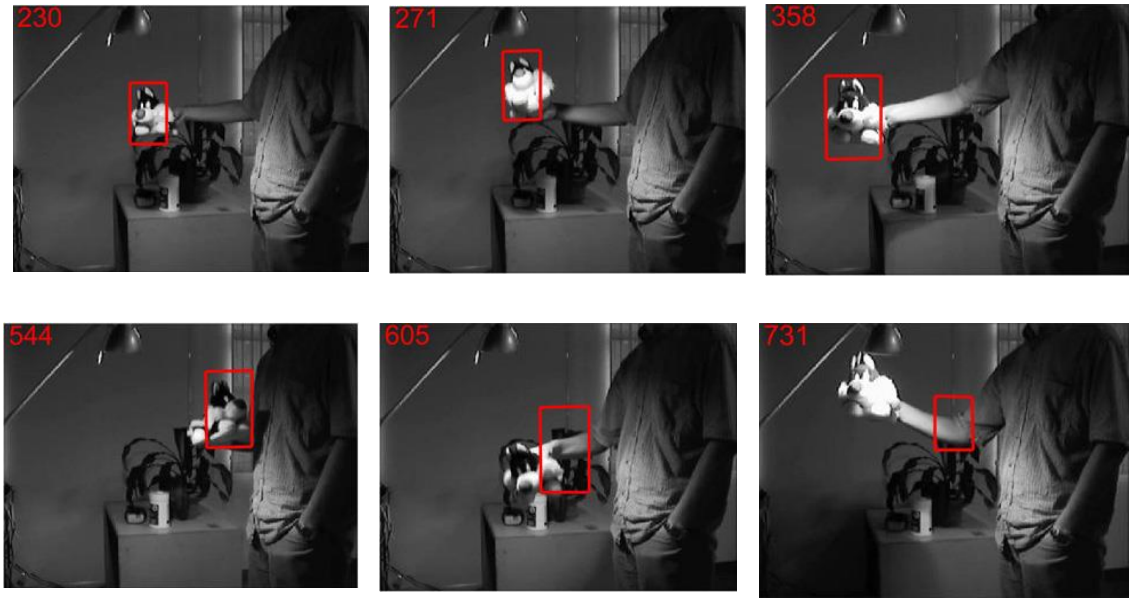



Woman

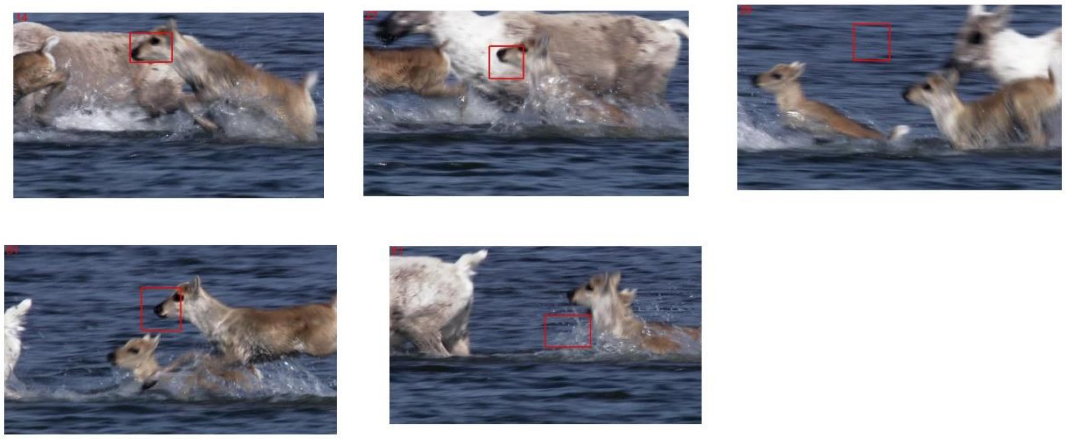




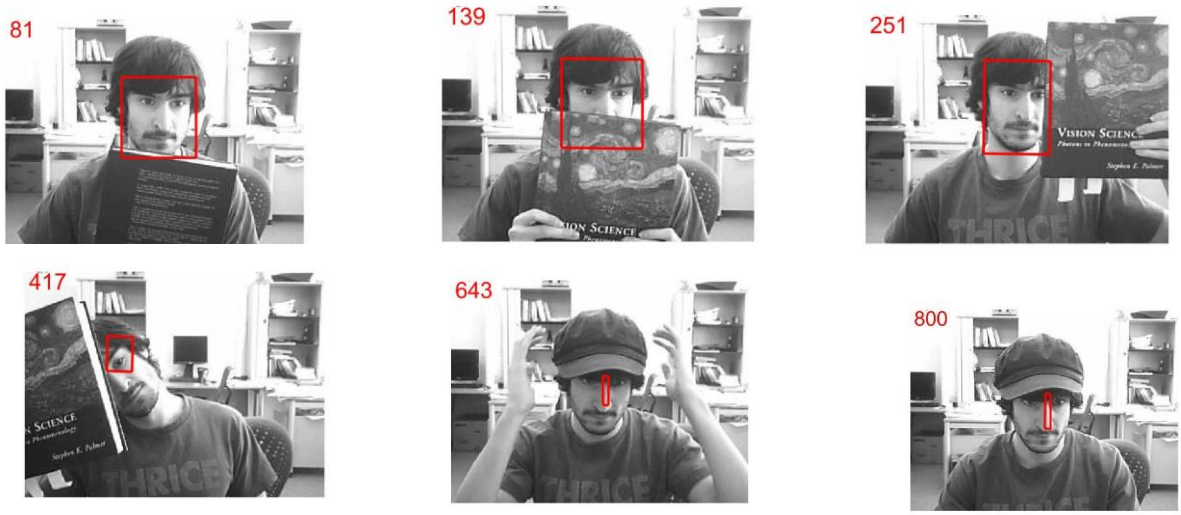

Pole



sylv



Deer



The results of this paper are compared with the following five paper:

1. Incremental Learning for robust visual tracking (IVT).

   Results obtained using IVT are shown below.

2. Visual tracking with online multiple instance learning (MIL).

   Code and data is not available for this paper.

3. Visual tracking decomposition (VTD).

   Application is available with the release version but certain dll files were missing. I found the cv097.dll, cxcore097.dll and highgui097.dll but mclmcrrrt7_16.dll was not found. For this I installed Matlab Compiler Runtime but the error continued. I tried the run on the specified Matlab version 2011b but still dll error was there.

4. Real-time Probabilistic Covariance Tracking with Efficient Model Update (ICTL).

   Code and dataset of this paper is not available.

5. Real-time tracking via online boosting (OAB).

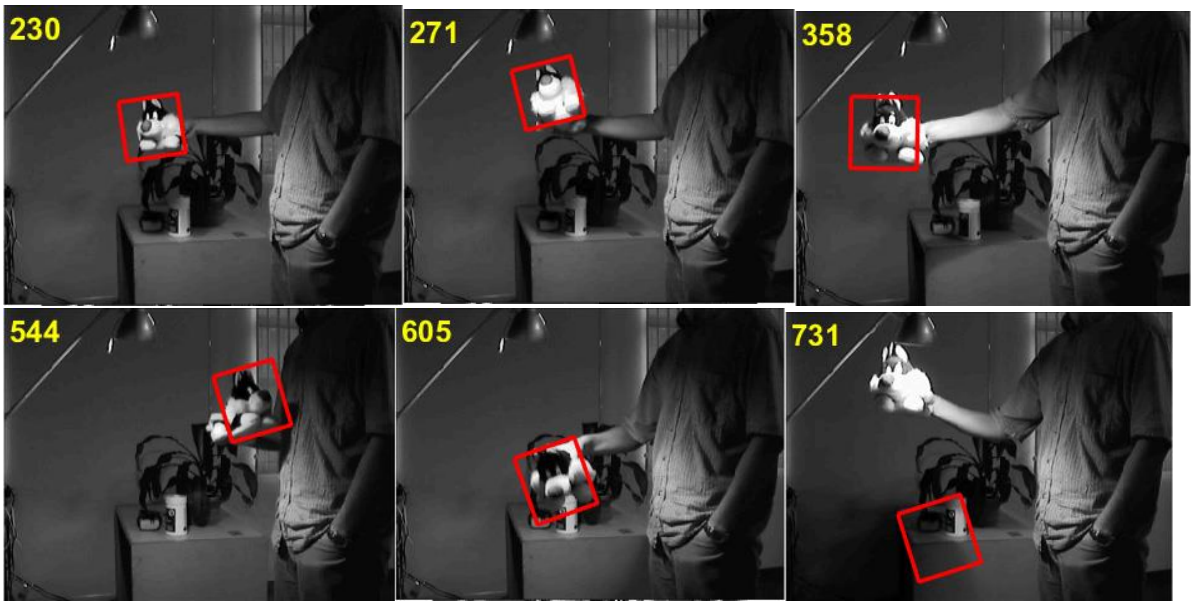
   Results obtained using OAB are shown below.

IVTL

So, for reproducing the comparison results I tried to reproduce these five papers and I am able to reproduce two of them successfully and code of remaining two papers was not available while code of one paper was not running. In IVT the code is given in Matlab and the datasets are provided in .mat format. Two datasets are common in IVT and original paper. I emailed author about providing other datasets in the required format or the details to build the format but I did not get the reply. I got the following results on the 'car4' and 'sylv' datasets. In the car4 results are approximately similar to given in the paper.

Car4

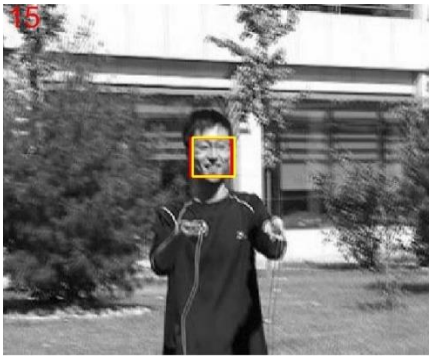
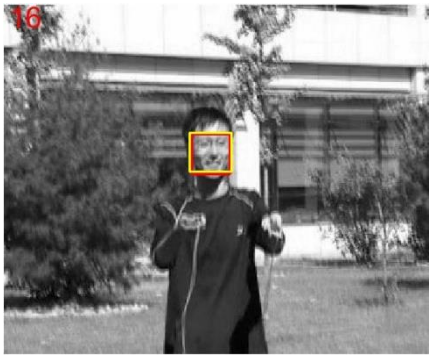
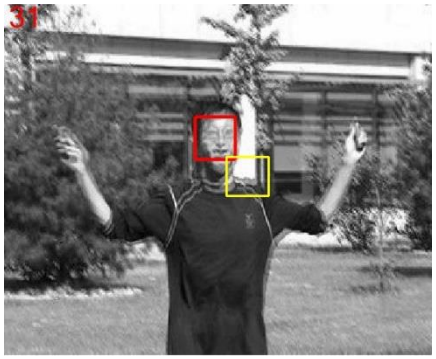
But in ‘sylv’ results are different because in paper it is shown that in image 605 IVT is not able to track the target but it is shown that it tracked the object.



OAB

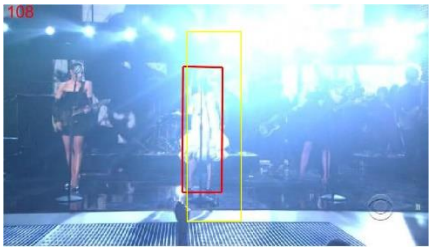
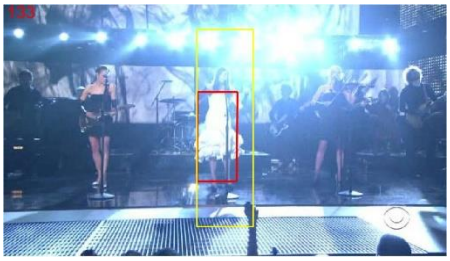
In OAB tracker you need to select the initial frame size by using the mouse and that start tracking that object. There is an option to define the coordinates of the first frame but it was not working and the proper documentation is not given to use it. I emailed all the authors of the paper but I did not get the reply. So for accurately providing the initial frame size of a target of an image I used the results of the original paper and got the following results.

In the paper it is shown OAB was not able to track the target in last three images and I got the same result but the coordinates are given different in the paper from the above results.

In the jump dataset OAB lost the tracker after image 35 but it is shown that it successfully tracked all the images.




In this results are approximately similar as it is shown in paper that frame size increases. These are 5 images because OAB automatically started tracking from 14 image so image 8 is not tracked.






The above results are not as shown in the paper. Similarly I used OAB to track images in all the datasets but there exists difference in the results from the paper.

Reproducing Table 1

In the table 1, the error is shown of each dataset of each of the tracker. Author has not mentioned any details about reproducing the table in the paper and not in any file of the code repository. In the paper below the figure in the caption it is written that error is calculated between two center points using Euclidean distance which is normalized by the size of the target using ground truth. Author has not provided any dataset to reproduce this table. So using the given information I tried to reproduce the table by getting the frame size of the resulted images through all the trackers using the Photoshop. It is time consuming task to manually point to each frame and get it x-axis, y-axis and then calculating its width and height. Afterwards it is required to calculate the center point of each target. To calculate the Euclidean distance it is also required to calculate center points of the ground truth data. There are more than 300 images in each dataset and there are 8 datasets and three trackers, so to manually calculating center point of each image may take too much, that's why I only used 8 images which author have shown in the figure 4. The error I calculated is not matching with the table 1 and it may be due to missing information such as mentioned "normalized by the size of the target" which have no any details and may be author have used any other method which he have not mentioned. However the error shown some accuracy while matching it with the ground truth data. When the target was tracked then error is small while when target is not tracked the error is big. The error is shown in below table.

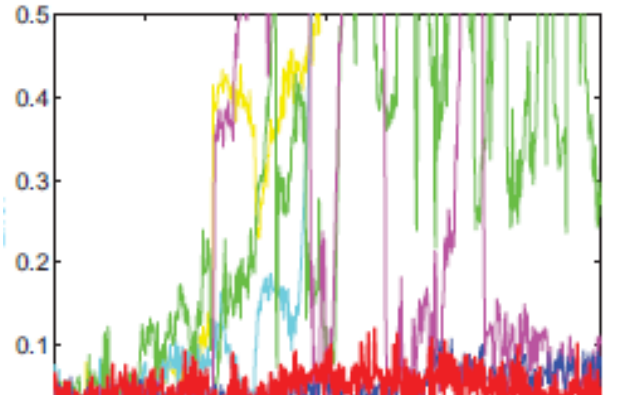
	OAB	IVT	Ours
jump	137.14		87.14
car	98.51	18.56	16.32
singer	207.772		180.83
woman	112.36		86.08
sylv	16.01	33.05	19.6
Average	114.3584	18.56	77.994

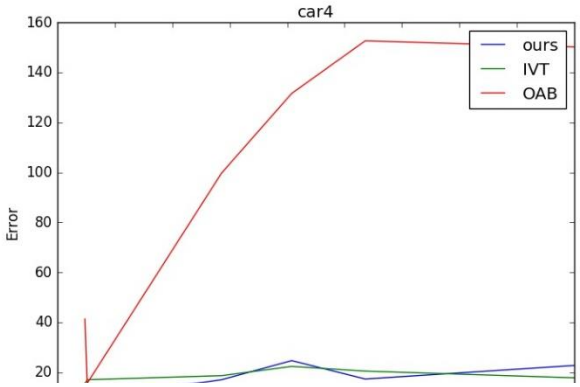
The average error for these trackers shows that IVT tracks better than all but it may not by accurate due to missing datasets and Ours (represents to taken paper) gives less error than OAB which shows its efficiency.

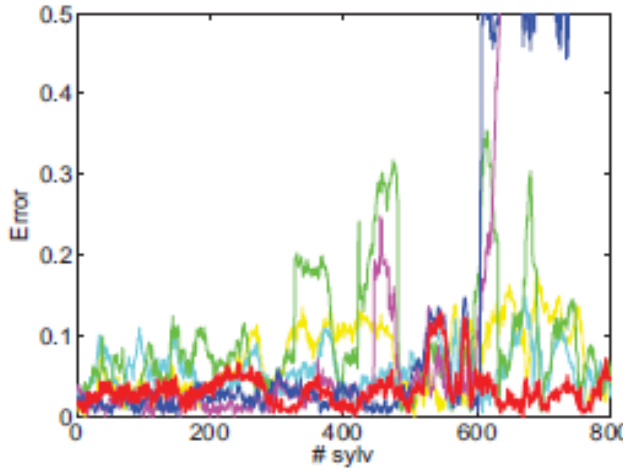
Reproducing figure 4

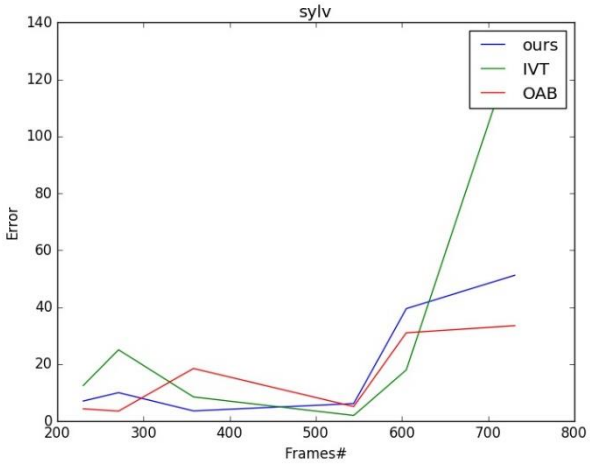
Author have just mentioned that graphs in figure 4 are drawn using table 1 and figure 3. So I wrote python script to draw the figures using the errors calculated in table 1. I was not able to match these graphs with the graphs given in the paper because details are not provided in the paper but if compare the car4 graph with the graph given in the paper then it can be seen that OAB error increases in both the graphs while IVT and L1 tracker (ours) error remain small. The results may be came more accurate if error have been calculated of the each frame. Similarly if we can see in sylv graph that ours and IVT errors are more similar as in both the graphs error increase between 400 to 600 frames and then it decreases but OAB error is not matching. While in jump dataset OAB error is at high at first in both the graphs but afterwards tracker lost the tracking so error was not measured for the remaining frames. In the singer dataset error for the OAB tracker was shown

in the increasing order in my calculated graph and the paper graph while L1 tracker (ours) error in paper is shown to be in decreasing order while when I calculated it is shown in increasing order. In the woman dataset error for OAB is approximately coming similar that at frame number 200 it increased highly but after frame number 400 it decreased gradually.

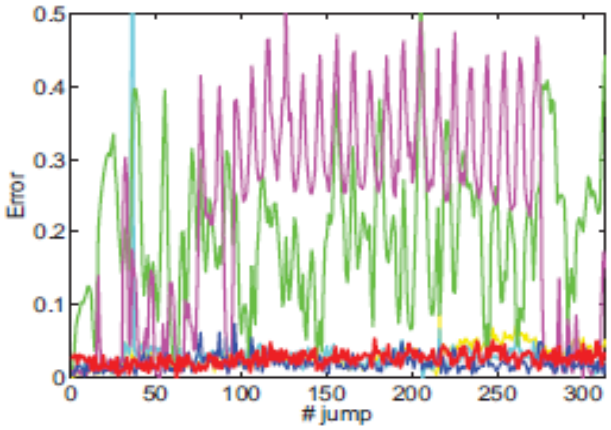


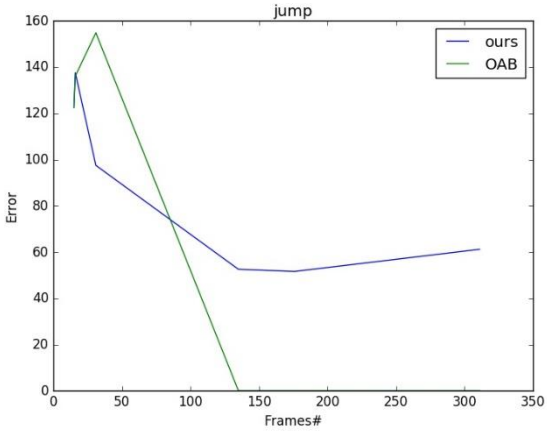




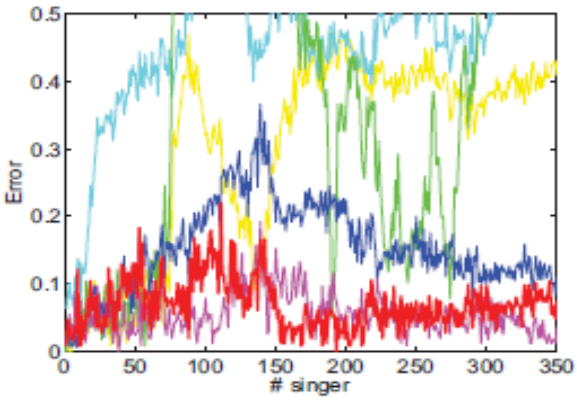


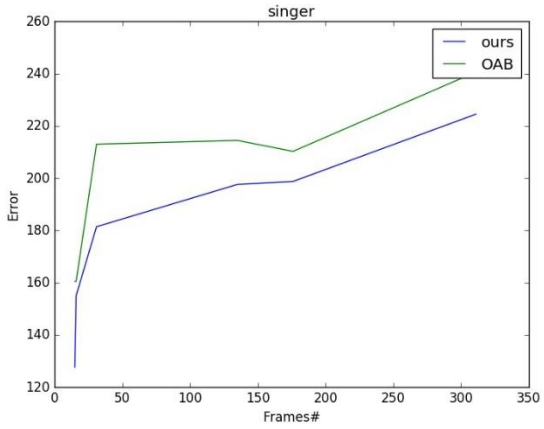
sylv



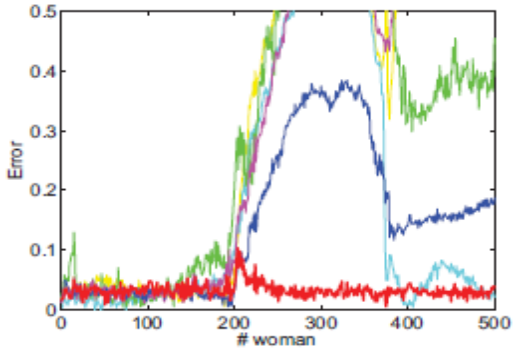


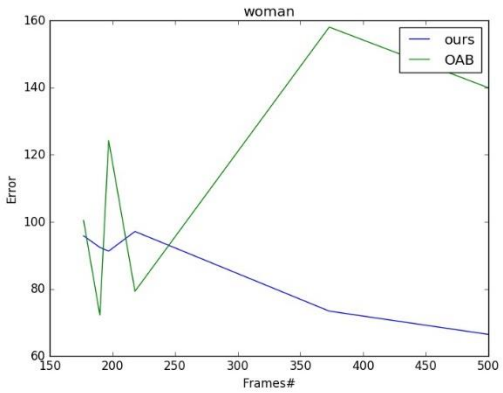
Jump





Singer





Woman

Conclusion

The results mentioned in the paper are not achieved even by changing the frame sizes, it may be due to not getting the actual dataset which is used in this paper because I used similar dataset got from other paper. This paper may be better reproduced if authors have provided data to generate the table and graph and if they would have provided description about making the table and figure. Furthermore the error calculated are approximately similar which presented in the paper, which shows that there exists a relationship between the ground truth values and the frame size used which author have not provided. These error graph can be further improved if size of all the frames is fetched and error is calculated, then these graphs may show a better relation with the graphs given in the paper.

References

[1] Chenglong Bao, Yi Wu, Haibin Ling, and Hui Ji. Real Time Robust L1 Tracker Using Accelerated Proximal Gradient Approach, 2012.

[2] Helmut Grabner, Michael Grabner, Horst Bischof. Real-Time Tracking via On-line Boosting.

[3] David A. Ross, Jongwoo Lim, Ruei-Sung Lin, Ming-Hsuan Yang. Incremental Learning for Robust Visual Tracking, 2008.