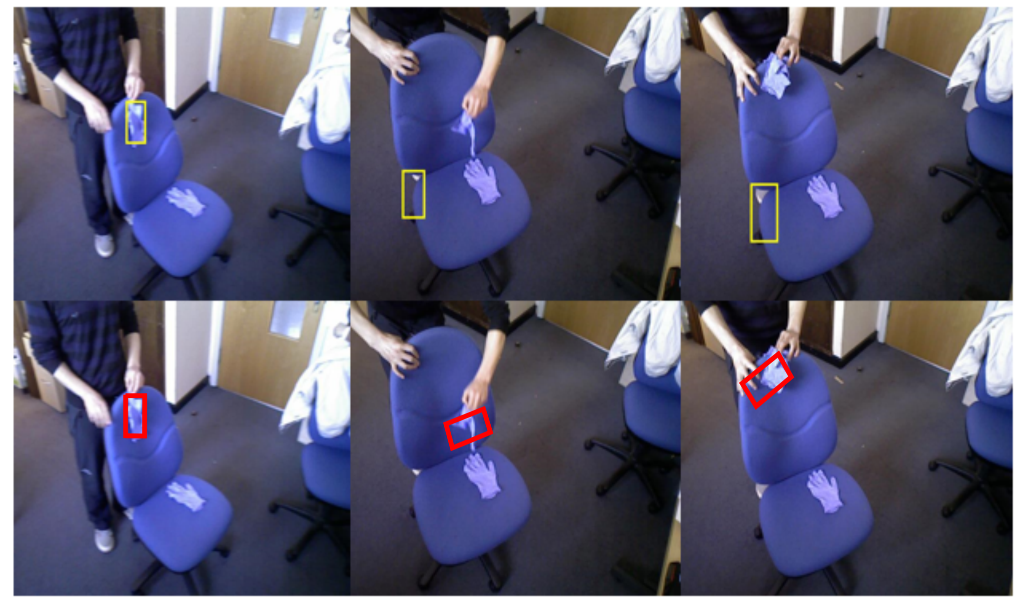
Rotation Adaptive Visual Object Tracking with Motion Consistency

Litu Rout[1], Siddhartha[1], Deepak Mishra[1] and Rama Krishna Sai Subrahmanyam Gorthi[2]

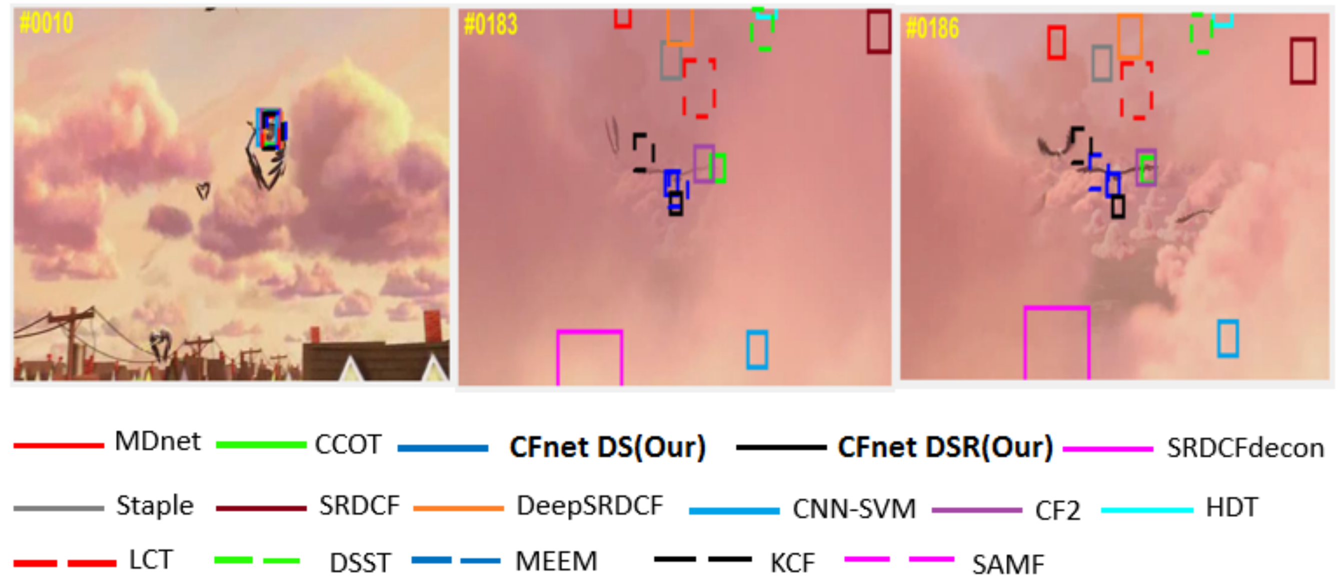
[1]Department of Avionics, Indian Institute of Space Science and Technology, Thiruvananthapuram, Kerala, India [2]Department of Electrical Engineering, Indian Institute of Technology Tirupati, Andhra Pradesh, India

Introduction:

- Visual Object tracking research has undergone significant improvement in the past few years.
- The emergence of tracking by detection approach in tracking paradigm has been quite successful in many ways. Recently, deep convolutional neural networks have been extensively used in most successful trackers.
- Yet, the standard approach has been based on correlation or feature selection with minimal consideration given to motion consistency.
- In this study the major contributions can be summarized as following:
  - A generic approach for incorporating rotation invariance (RI) in object tracking
  - Introduction of motion consistencies
    - Displacement consistency
    - Scale consistency



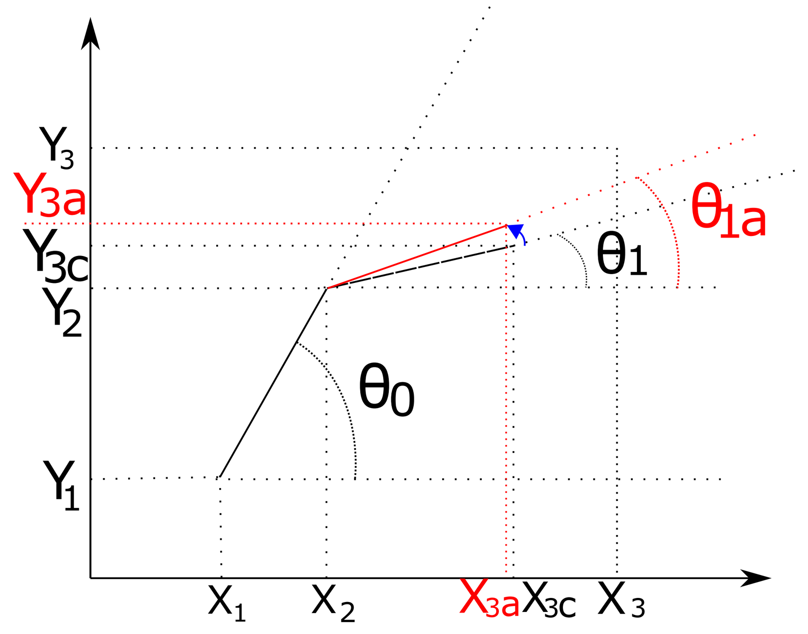
- Sample frames from glove sequence regarded as one of the toughest sequences according to VOT 2016 results. First column indicates the ground truth bounding box in the first frame. Our modified SiameseFC (red) successfully tracks the geometric deformations unlike original SiameseFC(yellow)



- Sample frames from Bird1 sequence, one of the toughest sequences in OTB50. The results are obtained using fully integrated OTB toolkit. Our trackers have not deviated much from the target centroid mainly due to the integration of displacement correction.

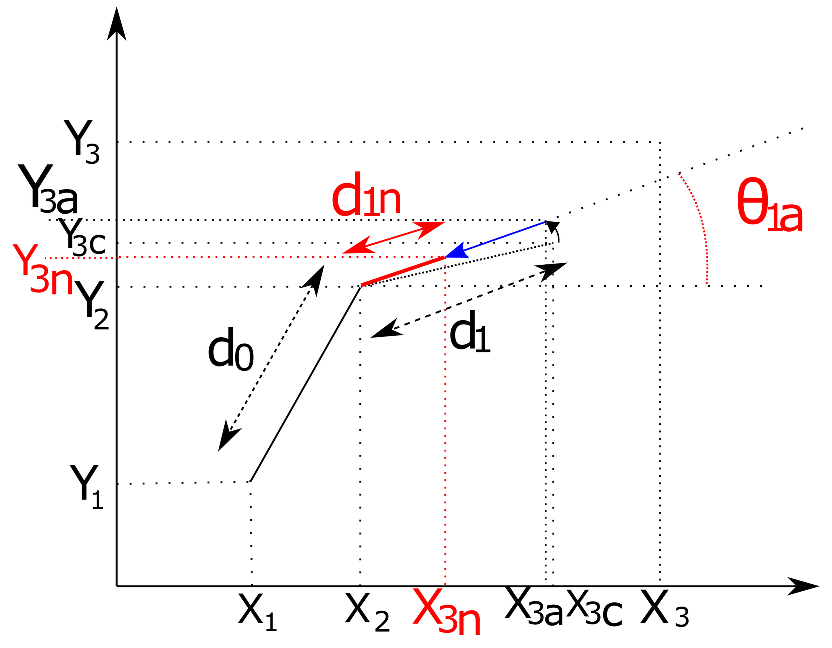
Proposed methodology:

Motion Consistency:



Angle Consistency

$$[X_{3c}, Y_{3c}] = w \times [X_2, Y_2] + (1-w) \times [X_3, Y_3]$$



Displacement Consistency

$$\theta_{1n} = w_\theta \times \theta_0 + (1-w_\theta) \times \theta_1$$

$$d_{1n} = w_d \times d_0 + (1-w_d) \times d_1$$

$$[X_{3n}, Y_{3n}] = [X_2, Y_2] + d_{1n}\angle\theta_{1n}$$

Scale Consistency:

- The conventional approach to estimate size of the target object is to form a scale pyramid and compute response map using each of these images.
- The corresponding scale of the response map having maximum response score among all these response maps determines the size of the target object and target centroid.
- if the position of the target centroid itself is corrupted due to the use of wining response map only, it will persist in subsequent frames. In this standard scenario, the response maps that correspond to different scales aren't used in determining the centroid.
- We propose to use Gaussian weighted average response map centred at the winning map and have variance as an additional hyper parameter. In this way we can incorporate the response maps that correspond to various scales in the scale pyramid.

Algorithm I : Scale Consistency using Gaussian weights

1. **Input parameters :**

   Let *responseMaps* represents the stack of response maps at each scale. μ represents the index of the winning response map. σ_{scale} represents the standard deviation of Gaussian weights. *scaleBins* numerically represents each scale i.e. *scaleBins(1)* represents the first scale, *scaleBins(2)* represents the second scale and so on. Let N represents the total number of scales used in the scale pyramid.

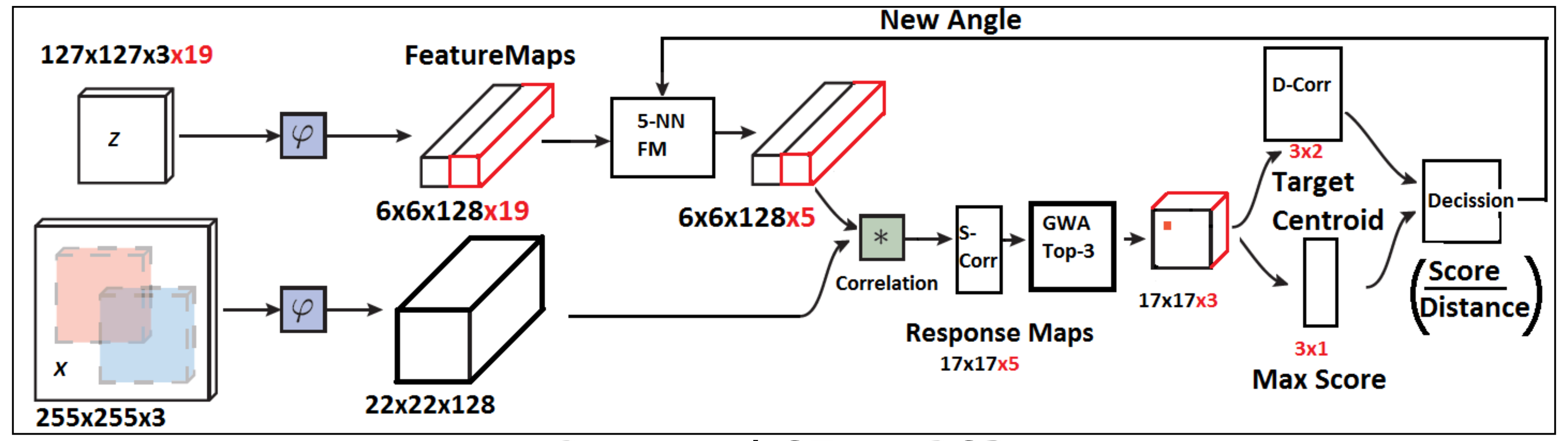
2. **Computation of scale weights and updation of *responseMap*:**

   (a) Define weights for each scale as $scaleWeights = \frac{1}{\sqrt{2\times\pi\times\sigma_{scale}}} \exp^{-(\frac{scaleBins-\mu}{\sigma_{scale}})^2}$

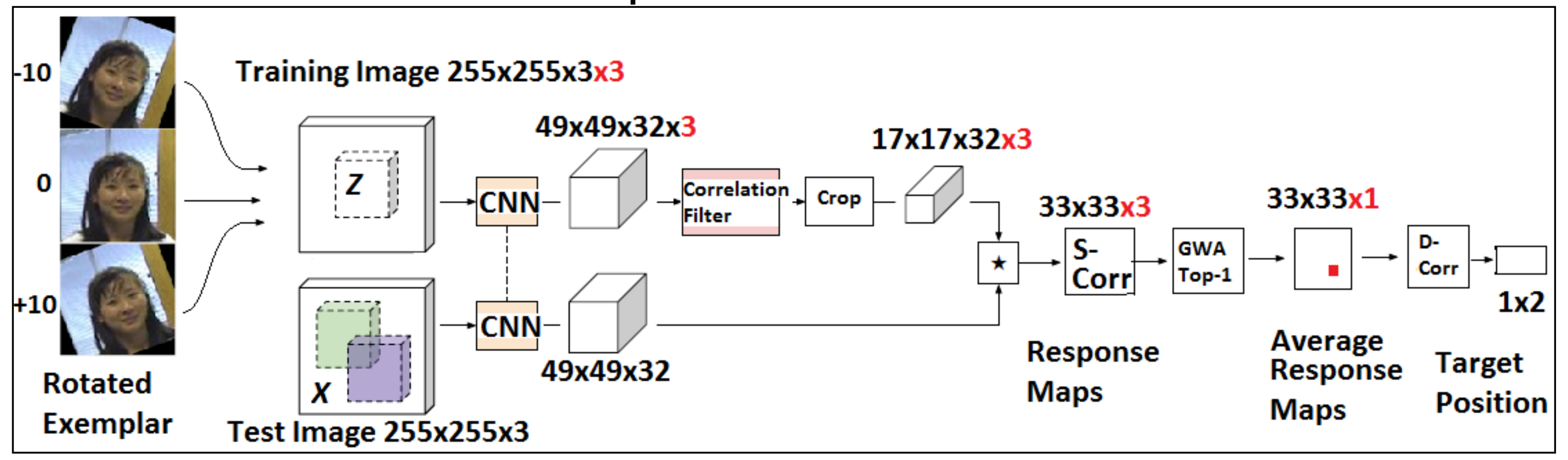
   (b) $responseMap = \sum_{i=1}^{N}[responseMaps(i) \times scaleWeights(i)]$

3. **Output response map :** The output of this algorithm is the Gaussian weighted average *responseMap*.

Proposed SiameseFC DSR

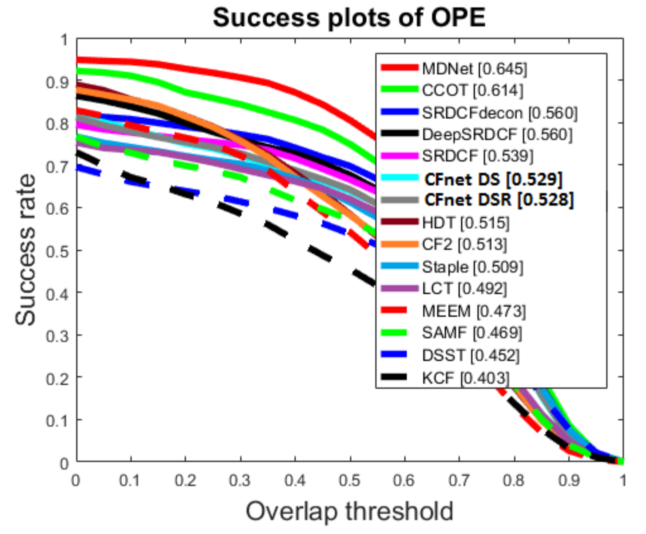


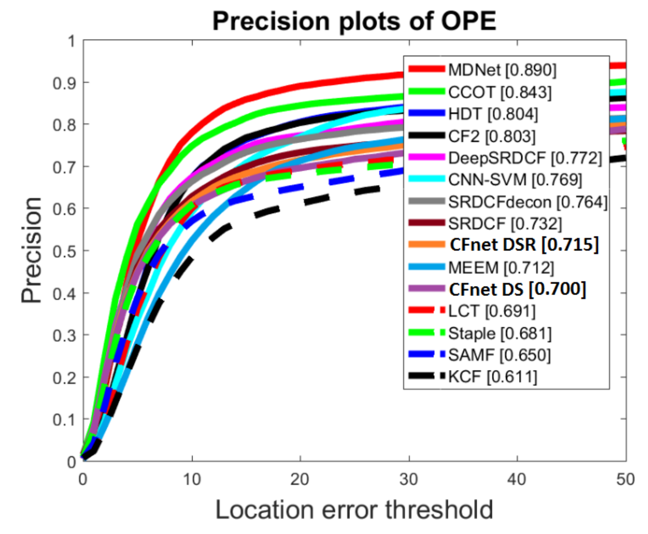
Proposed CFnet DSR



Results and Conclusions:

- The work demonstrated a way to incorporate Rotation Invariance (RI) in generic object tracking.
- The introduction of scale and displacement consistency enhanced the degree of smoothness on physical movement variables such as speed and angles.
- The success rate improved by 4.6% whereas precision, by 6.75% relative to baseline approach on OTB dataset.
- The Proposed Siamese DSR gave a drastic improvement in robustness rank by 15.7% and accuracy rank by 14.3% on VOT 2016 database.
- Our future research may include replacing the simple CNN present in both Siamese and CFnet architectures with a very deep CNN.





Success and Precision plots on OTB50 dataset

Authors contact details: Litu Rout, liturout1997@gmail.com, +91-9439331096
Deepak Mishra, deepak.mishra@iist.ac.in, +91-9446030737