Real time Object Tracking in video surveillance using Prior context and Image Patching

A Dissertation submitted in partial fulfillment of the requirement for the Award of degree of

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted By

Pullak Gupta

(2K15/ISY/14)

Under the esteemed guidance of

Dr. Anil Singh Parihar
Assistant Professor

Department of Computer Science & Engineering

Delhi Technological University

Bawana Road, Delhi-110042

2015-2017

CERTIFICATE

This is to certify that the thesis entitled “Real time Object Tracking in video surveillance using context and Image Patching(PCIP)” submitted by Pullak Gupta(2K15/ISY/14) to the Delhi Technological University, Delhi for the award of the degree of Master of Technology is a bona-fide record of research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Place: DTU, Delhi

Date: _____

Dr. Anil Singh Parihar
Assistant Professor
Department of Information Technology
Delhi Technological University, Delhi

ACKNOWLEDGEMENT

Firstly I would like to express my gratitude towards my supervisor Dr. Anil Singh Parihar *Assistant Professor, Department of Information Technology* for his able guidance, support and motivation throughout the time. It would not have been possible without the kind support and help of many individuals and Delhi Technological University. I would like to extend my sincere thanks to all of them.

I would like to express my gratitude and thanks to Dr. Kapil Sharma (*Head of Dept.)* for giving me such an opportunity to work on the project.

I would like to express my gratitude towards my parents & staff of Delhi Technological University for their kind co-operation and encouragement which helped me in completion of this project.

My thanks and appreciations also go to my friends and colleagues in developing the project and people who have willingly helped me out with their abilities.

Pullak Gupta
2K15/ISY/14
Dept. of Information & Technology
Delhi Technological University

ABSTRACT

Here we discuss a fast and a robust algorithm for tracking an object in video surveillance, there are various challenges when it comes to dealing with video tracking, such as image illumination, motion object, occlusion, scaling etc. we present our algorithm that handle such measure, we start with generating the weighted map for the image, then we track the object in form of patches, whose combined effect is used as a vote map. Then using image signature and spatio-temporal computation we generate confidence map to predict the next position.

Our algorithm doesn't require training, we just provide the initial position of the object which need to be track. Numerous experiment result show that our algorithm achieve significant improvement our SPC tracker **(Cong Ma, 2017)**

Table of Contents

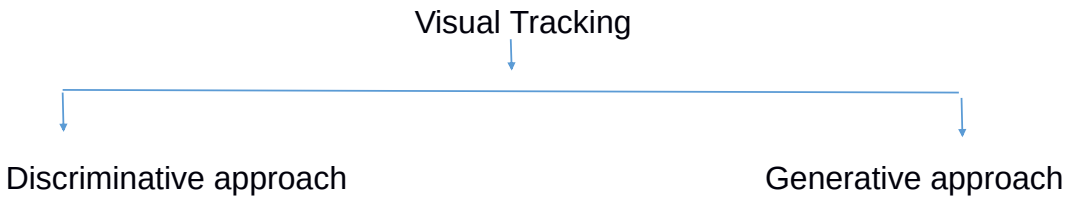
List of Figures

List of Tables :

1. Introduction

Visual tracking has become a key subject in muti-media processing and computer Vision, where the goal is to track the object location when it is in motion, Visual tracking can be seen in exist in various field to exist of Multimedia such as, activity monitoring, human computer interface, surveillance.

Illumination, occlusion and background appearance are few of the factor that affect the visual tracking. Many ideas have been proposed to focus such issues which can be broadly categorized into



1.1 Discriminative approach

Discriminative approach consider tracking target as a binary classification problem & exploits the differences between the target object and its surrounding. Background information plays a key role in exploiting target object, and often work better than the traditional ones. However they are not left behind from various challenges such as classification of boundaries, sample size.

Moreover, little has been known so far about how human eye perceive the target in such discriminative ways. Further, many current methods (S. Hare, 2011) , (K. Zhang, 2012), run at slower speed with high computation cost, hence restricting themselves from practical usage. So, it becomes essential for us to explore a fast and an effective way of tracking the object yet producing the effective result.

1.2 Generative approach

In this, firstly we represent the object to be tracked in form of appearance model, then searching for the image region is applied to find the best score in the

successive frames. The searching result is depend on how well the appearance model has been designed. Building the an effective and an efficient appearance model is a bottle neck for tracking, since it is generally a tradeoff between the computation versus the computation complexity, recently many object representation and learning techniques have been explored ,some of the recent paper **(L. Sevilla-Lara and E. Learned-Miller, 2012) (C. Bao, 2012)** focus on the good representation of the object[target].

Many algorithm exploits features, or region or the contour of object thus providing different result. Target object and its surrounding is called as Context, so within a context there exist a strong possibility of correlation between them, see the **Figure 1**

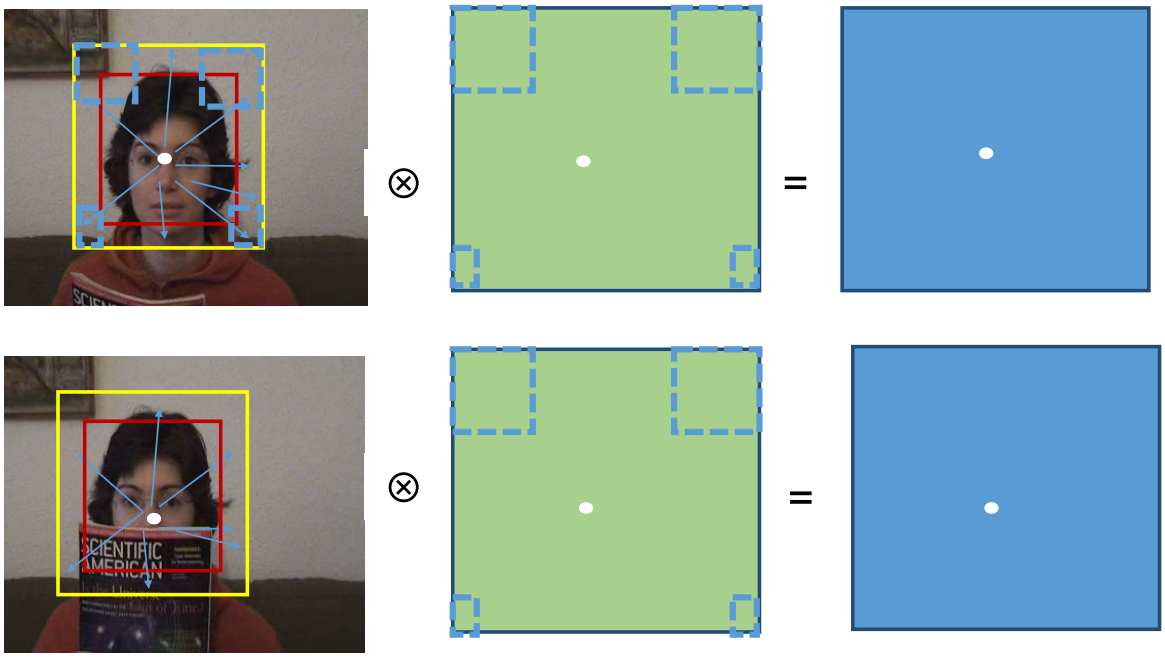


Figure 1: Context and its surrounding

within the yellow box ,the context region has both object and its local surrounding, left : We can see spatial temporal relation between the object(marked with white color) and its surrounding(marked with blue line) remain preserved even though object tends to change it position and even due to occlusion, Middle: the learned spatio-temporal context model the portion inside the blue box have almost same value , which show that consecutive frames have similar spatio – temporal relation with the target center.). Right: generated confidence map, learned using the spatio-temporal approach

1.3 Concept of Patches:

Patches are the small rectangular region, which are sub-region of the target template, which define a portion of the object. patches are very good example of region based tracking, where the idea is to break the entire object in to different sub-region and then process each sub-region as a separate object and then combine the vote map of their result to determine the new location of the target in the current frame There are many ways of applying tracking on the images, such as **(rivlin, 2006)** .

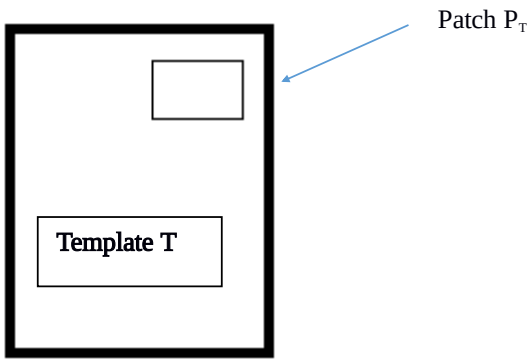


Figure 2: Object Template with Image Patch

Ultimately the goal is how well is matching similar exist decide the location of the tracked object.

Concept of patches over the various measure like mean-shift and kernel tracking is quite unprecedented, mean-shift algorithm (T.collins, 2003) is a non-parametric statistical method for finding the nearest sample distribution , sample points are distributed along the region with some weights *w(a)* ,the sample points are chosen in such a manner that the foreground fixed are given more weight than the background one .the mean shift algorithm specifies how to combine these sample weight over a kernel k(a) to find the centroid of the image , where patches are just the another sub-region on the template which work as a separate image

1.4 Thesis Outline

In this Thesis , we have present our new idea of combing Prior context spatial temporal result using image Patching (PCIP) that track the object location without any increase in the complexity. The organization of the thesis grows in this manner, First we explain the Literature work in chapter two, then in chapter we present the proposed method and algorithm, Chapter Four present the experiment result and extensive comparison with the previous work, Chapter Six Conclude the entire work and the future scope

2. Literature Survey

To perform a video surveillance, an algorithm should analyze the sequential frames and must result in new position of the object in the consecutive frames. There are a many algorithms, each having its limitation and its strength. Depending upon the nature of tracking, we must choose an algorithm to our application. There are two major components of a visual tracking system: target representation and localization, as well as filtering and data association.

Localization of the object and its representation is generally a bottom-up process., Thus depending upon the representation of target different approaches exist to process them, many algorithm exist which tends to explore appearance model , For example, blob tracking (T.collins, 2003).

There are many studies which focus attention on appearance model and context model, and some incorporating the saliency feature of the object,

2.1 Appearance model and context model

Many appearance model have been conceived, such as on-line learning techniques by D.Ross etal using IVT method, which learn and extend the appearance model using incremental PCA.

Further, idea of context model has also grown to popularity , where we incorporate the changes of the surrounding , while we are tracking the object , basically we keep the size of the context as twice the size of the target object

2.2 Feature based Distribution

Feature based distribution ,is a pixel –level context representation and is based on the feature the object possess such as intensity , orientation , color , contrast etc. are most common feature that are widely used by the saliency approach. It is a bottom-up approach that integrate the image gist. For calculation simplicity we omit the various feature except the intensity.

Our approach incorporate this distribution on grayscale image, thus omitting the

computation cost of Color image which require 256*256*256 pixel wise computation over 256 pixel, is good to serve our purpose. Further the performance tradeoff is very low, if we see the computation cost of working with color image. In this thesis our main objective is to track the object speedily. Thus we are using intensity map at position , to model our appearance distribution.

Further, more weight should be assign to the target framework to distinguish it from the background region, thus we multiply the whole image template with Gaussian function

(1)

Thus, applying this function over the Image template we obtain the feature based distribution i.e.

(2)

2.3 Signature Based Distribution

It define the salient foreground of the pixel that comprise to form the object, in this distribution we match the different pattern / signature of the object that depict the object, we have incorporated Image signature method **(X. Hou, 2012)** and derive a sparisty map from it

Computation steps for generating Sparsity map

1) Extract the target object template from the given image
2) Calculate the image signature IS(y) using Sign Discrete Cosine Transform (DCT)

(3)

3) Now generate spatial sparisty map from the result obtain in equation(3), by smoothing the reconstructed image

(4)

(5)

here is the convolution operator and is the hadamard production and g

is 10*10 Gaussian kernel

4) Finally we normalized the compute above to obtain a working model for our computation

(6)

3.

Proposed Work

The idea of context (that is the surrounding or the neighborhood of the target object) has been incorporated to better exploit the local surrounding. Rectangular region with quadruple size shows the region of context see *Figure 3.*

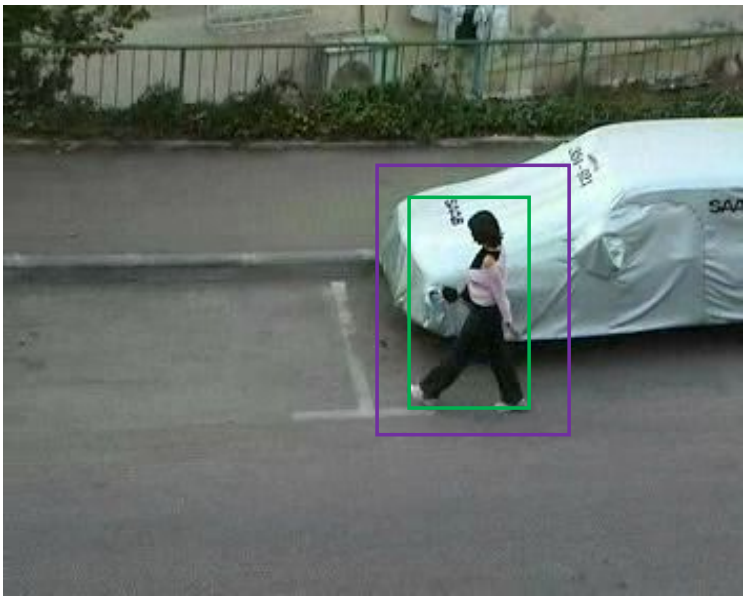


Figure 3: Image depicting the target object context, green rectangle denotes target object and purple show the context

Our Algorithm is based on the integration result of **Sparsity based distribution, feature based distribution, using image patching approach**

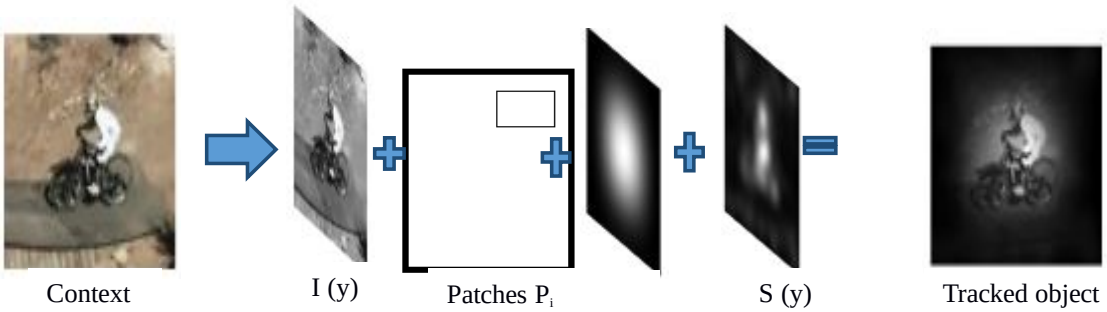


Figure 4 Working model for Proposed Algorithm (PCIP)

Using the literature work, above we sum up our working model as

(7)

Now the right side of the equation we have already discussed above, now to evaluate the left side we have divided the whole approach in to two stages

3.1 Evaluation of the image patch

Given the frame, and Template ‘ , representing the object , aim is to locate the position and scaling factor of in the current frame , approximate to ‘T’ in some regard. Assign the initial position and scale of the target location in the first frame, and now search this object in the neighborhood of the consecutive frame. We represent the (x, y) to denote the initial center of the object to track in the image.

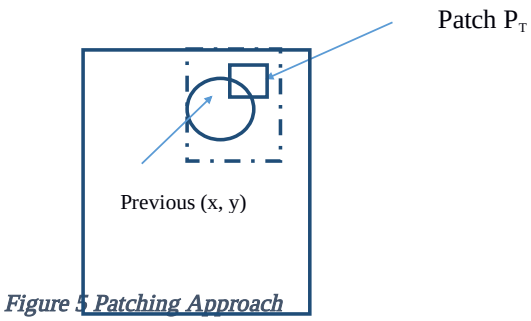


Figure 5 Patching Approach

The circle outing the (x, y) is the hypothesis that the next point will lie within the radius ‘r’ of the previous tracked position and we need to estimate that location, let denote the rectangular patch (d_x, d_y, h, w) in the template T , whose center has displaced by (d_x, d_y,) , with width ‘w’ and height ‘h’

Then position of patch in the image P_{img} (x, y) will be at the Centre

Now given the two patch and we wish to match the similarly between the two patches to validate our hypothesis can be done using the similarity measure), where M and N are Patches

Using the value), we define the term the vote map,

(8)

The lower the value of correspond to the very high chances of sustaining our hypothesis

3.2 Patch Similarity Measures

There are different approaches to measure the patch similarity, we compute them by building the histogram for each patch, and some of them are listed below

3.2.1 Chi-square statistic

This is probably the smallest measure is to compare the bin of their histogram bins using, the two bins are the vectors and computing the difference between the two vectors

3.2.2 Kolmogorov-Smirnov statistic

This is almost similar to chi-square statics, but it compare the two histograms by building their CDF for each histogram, the advantage over the later over the former is smoothing of nearby bin difference as a result of quantization of nearby bin measurement

(9)

3.3.3 Earth Moving Distance

This is what we have used for our computation, in this, actual dissimilarity between two histogram is considered **(Earth mover's distance, n.d.)** , by computing the minimum cost required to move one bin in to another bin.

Let Histogram P has n bin with values where represent the feature and represent the weight of the feature, Similarly Histogram Q has n bin with values ,

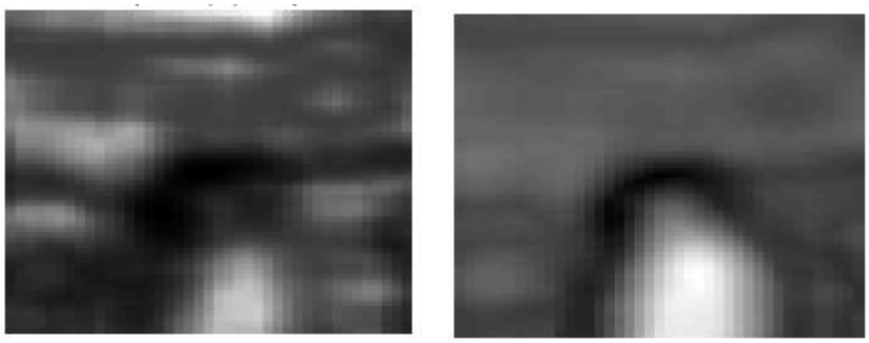
Let be the ground distance between Histogram and . and we want to find the flow that can minimize the distance between the two

(10)

Such that

&

Below is the comparison (Rivlin, 2006) of how EMD is much better than Chi-square statistic see below



Using chi square approach Using EMD

Figure 6 Comparison result of distinguishing image patches uses chi square and EMD

3.3 Incorporating Prior context Model

Now given the vote map , we need to incorporate this vote map with the previously tracked position to give the new position in the current frame ,To do so we need to generate the confidence map

(11)

Where represent the spatial relationship , which handles the ambiguity due to different interpretation of image measurement and represent the context prior probability which model local context appearance.

Now, context prior probability is obtain from the feature based distribution as discussed in the literature review using the equation (2)

Now we are left with spatial context model, which is nothing but the relative distance between the object and the direction of the location ‘y’, thus

(12)

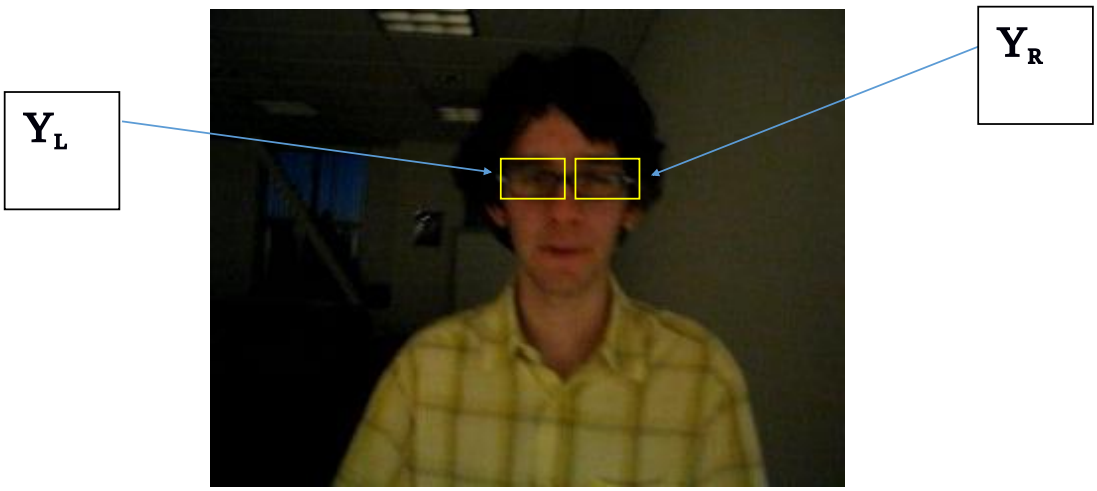


Figure 7 Spatial context Feature extraction

Note that, it is radically symmetric so, , hence reducing the ambiguity when the object are closer in proximity , for example eye tracking , when we are measuring the appearance of left eye(Y_L) , we found a similar object i.e. right eye(Y_R) , but both can be differentiated since the surrounding of the both are different , further their location coordinate are different and thus

Thus

(13)

Proposed Algorithm Step

1. Initialize the image template for object tracking processing.
2. Divide the image template into various patches
3. Calculate the patch score for individual patches and select those patch which have best score using earth moving distance
4. Now use this patch score over the prior context value and spatial context score
5. Find the confidence map to track the next position of the object using equation (16)
6. Update the spatio-temporal value, using the equation (17)
7. Update the scaling parameter using equation (18,19,20,21)

3.3.1 Confidence map

The confidence map of the object location is taken as

(14)

Here ‘a’ is the constant for normalization,

σ is a scaling parameter

is a parameter to determine the shape

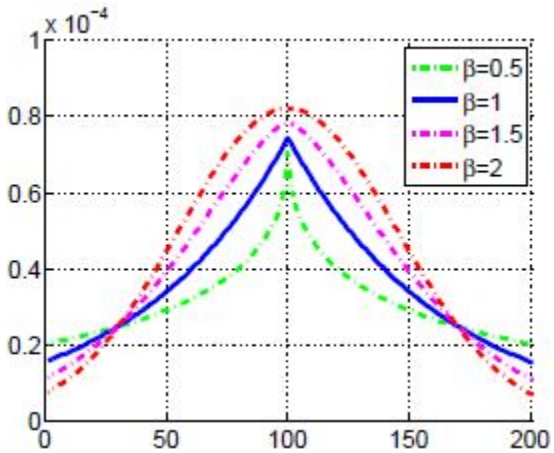


Figure 8: shape parameter describing the object symmetry

Often we observe ambiguity in visual tracking which result due to close proximity of the object location, thus leading to poor performance of tracking. To resolve such issue we have taken in to account various instance learning technique which handle the ambiguity due to object location. The closer the location, the larger probability is that ambiguity will occurs (e.g., predicted new target locations that only differ by few pixels are all possible solutions and hence creating ambiguities). In our approach, we handle such cases by setting the value of shape parameter obtain from the result as shown in Figure 8 , for large value (e.g. = 2) over smoothing occur, On the other hand, a small value (e.g., ,= 0.5) result in a sharp peak near the object center, hence failing to reduce the ambiguity , with value =1 in it achieve near symmetry , thus achieving our goal of spatial context for achieving radial symmetry i.e.

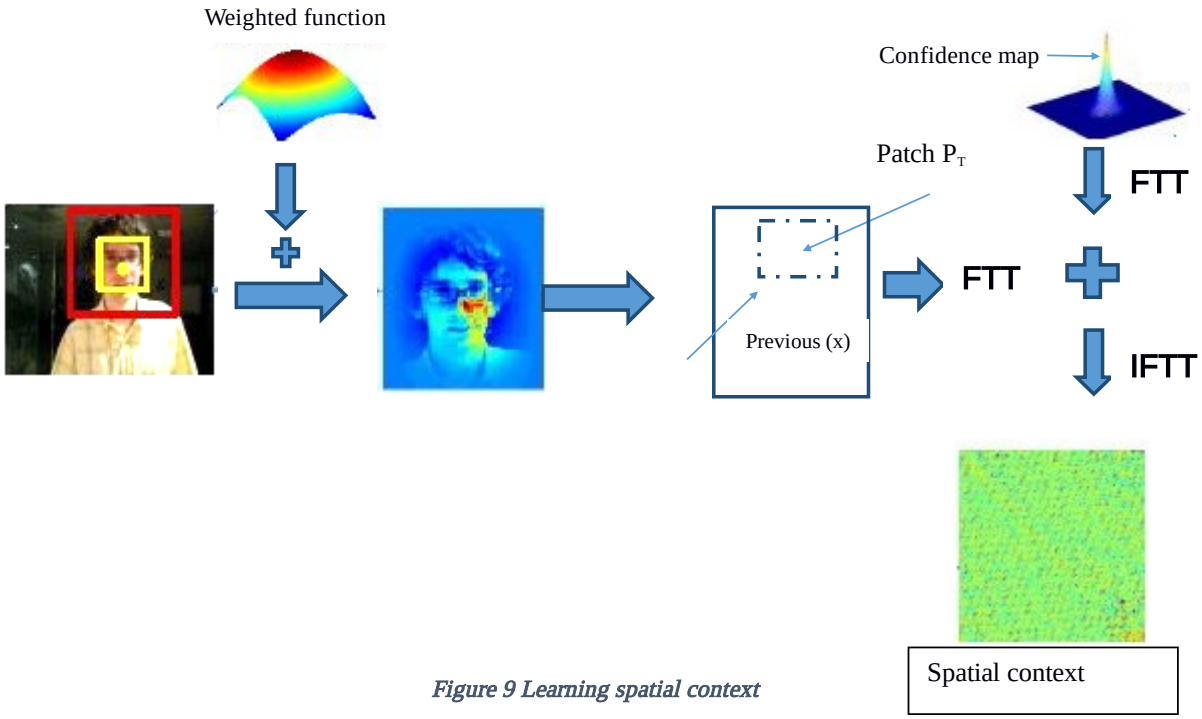


Figure 9 Learning spatial context

3.3.2 Speedy Learning of Spatial context Model

Based on the context prior model and the confidence map, our aim is to learn the spatial model, to put thing together

We know from equation (14)

Also using equation (13)

From above equation (13) and (14) we get

=

Taking the fast Fourier transform, we obtain

)=

Thus

(15)

Now to track the updated location of object, in the t+1 frame, $X_{t+1,}$ we need to find the new confidence map which is calculated as

$$X_{t+1} = \max (C_{t+1}(x))$$

Where $C_{t+1}(x)$ is calculated as ;

$$C_{t+1}(x) = \quad (16)$$

3.3.2 Updation to spatio-temporal context

We update the spatio- temporal using combination of previous spatio –context and spatio – temporal context

(17)

Where is the learning rate

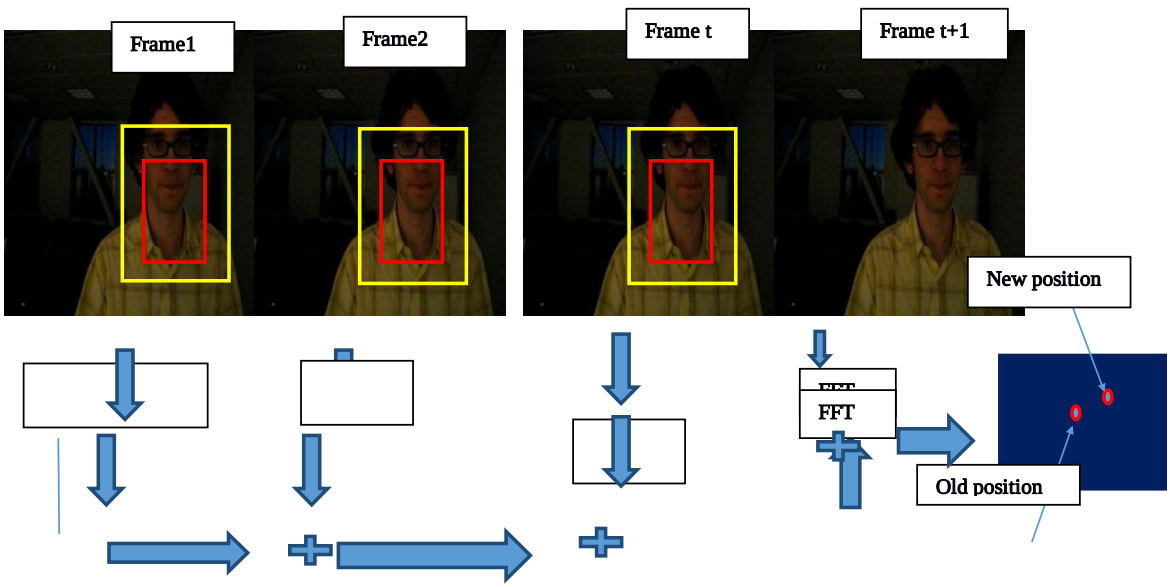


Figure 10 Tracking implementation

3.3.3 Updation of the scale

For consecutive frame, all the scale parameter need to updated, so that accordingly in the next frame these parameter should still be valid, for e.g. scale σ in w_{σ}

Below are the parameter that will be updated using the equation

(18)

(19)

(20)

(21)

4. Experiment Result

We have evaluate the proposes tracking algorithm based on *Prior context spatial temporal result using image Patching* , using 8 videos with various challenges such as occlusion , blurriness , background effect , illumination changes and scale variation using 6 tracker ASLA , CPF , DFT ,IVT,SMS, SPC

Initially we set the initial position of the various object , and set the size of the context to be double the size of the initial position, we set the scale factor() initially as 1 , the learning parameter as 0.25 . the parameter for map function α as 2.25 and frame cycle ‘n; as 5 , to remove the reduce the effect of illumination change which is obtained by subtracting the average value from each intensity value of that region and finally multiplying it with hamming window , thereby removing the boundary effect

We evaluate the performance of our algorithm using Center location Error (CLE) **(Hongbin Zha)** ,

4.1 Center location Error

It a measures to find the error difference between the center location of tracking window and the actual center point, naturally smaller the error better the result and optimum will be the tracker

So, we can write CLE as

(22)

Where (x, y denotes the actual center coordinate and (x’, y’) are the tracked center coordinate

Below table is the comparison result of our approach with generative method,

Sequence	ASLA	CPF	DFT	IVT	SMS	SPC	PCIP
Car	1.79	35.90	61.94	**1.73**	140.65	8.56	8.12
David	5.10	26.13	42.88	**4.82**	24.18	8.51	7.1
woman	153.61	115.46	**8.50**	188.53	99.34	12.31	10.0
David Dull illumination	6.13	45.67	9.12	**2.45**	27.13	8.44	8.04
Car(traffic)	1.85	33.90	51.94	**1.79**	141.65	9.56	9.02
Girl	7.61	18.20	23.98	20.75	24.18	8.05	**7.44**
Instrument	79.54	184.33	26.29	87.15	180.43	8.75	**7.55**
Puppy	18.36	14.68	44.38	36.66	18.36	5.95	5.2
Average CLE	**34.24875**	**59.284**	**33.629**	**42.985**	**81.99**	**8.7663**	**7.8088**

Table 2 : CLE value comparison for various approach, best value is marked in bold

Further, we have illustrated our approach with that of our reference paper, SPC

Figure 11 Comparison Result using Car left is the SPC method and right is PCIP

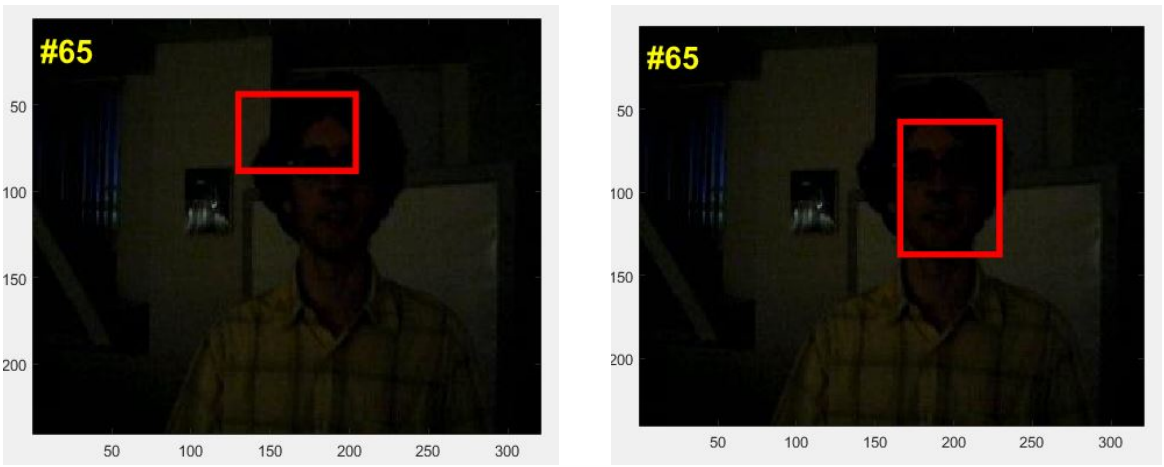


Figure 12 Comparison result using David apperance in low light condition left is the SPC method and right is PCIP

Figure 13 David movement tracking, left is the SPC method and right is PCIP

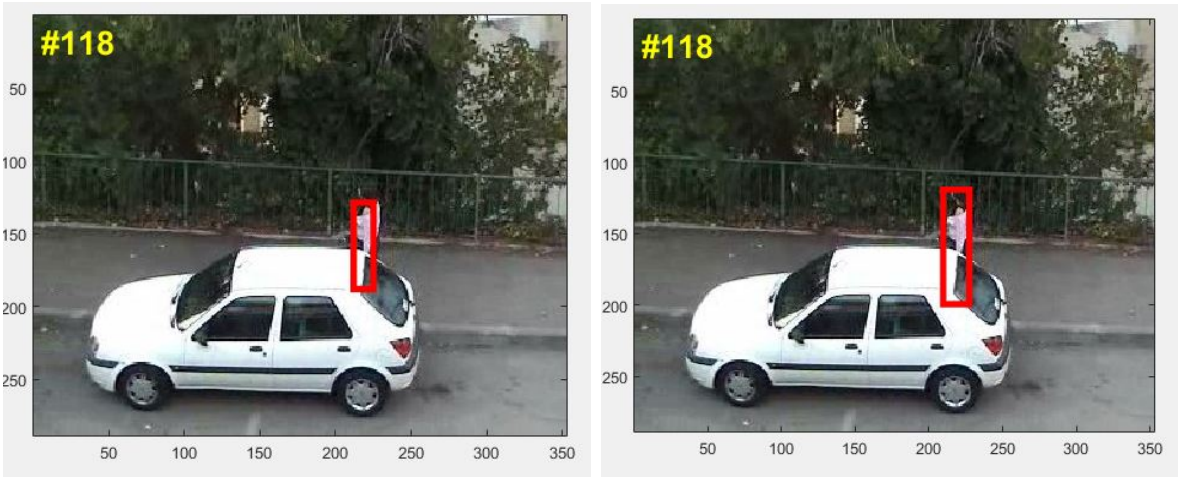


Figure 14 Woman position tracking, left is the SPC method and right is PCIP

Figure 15 Tracking specific car under traffic condition, left is the SPC method and right is PCIP

Figure 16 Girl movement while sitting on chair, left is the SPC method and right is PCIP

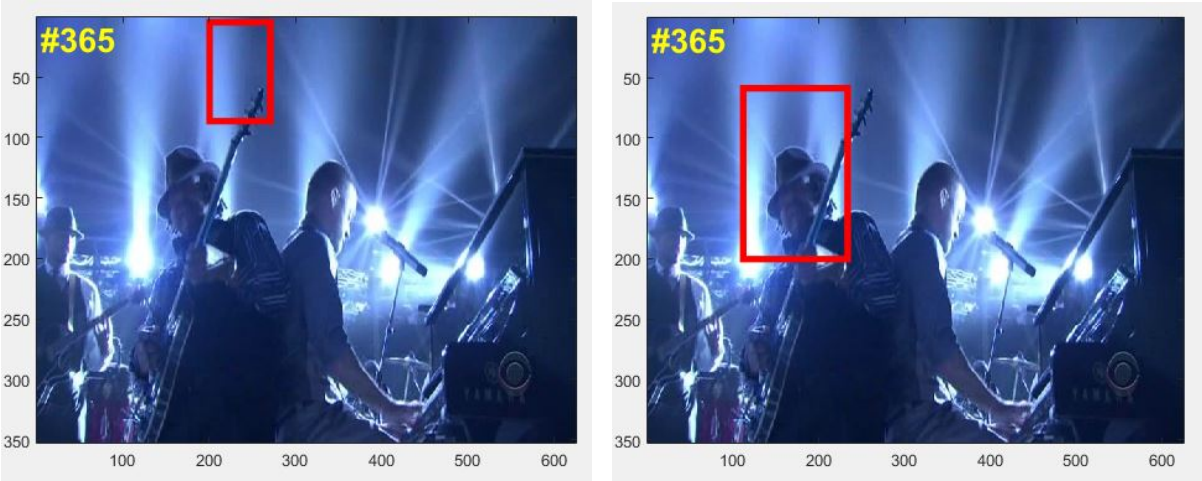


Figure 17 Tracking object, when playing instrument, left is the SPC method and right is PCIP

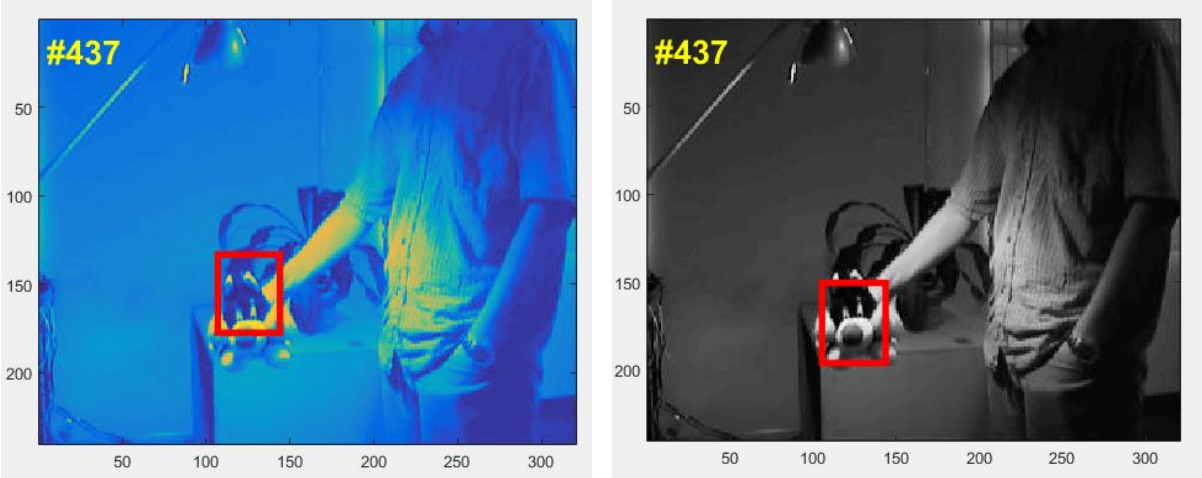


Figure 18 Tracking puppy, left is the SPC method and right is PCIP

5. Conclusion

In this thesis , we have demonstrated fast, but and efficient algorithm for tracking our object using the concept of patching over prior context , the proposed method were tested over various challenges such as occlusion, illumination variance, motion effect, blurriness and pose variation , the result obtained are satisfactory to the previous method . The given algorithm work at the rate of 300 FPS and provide a real time tracking speed

References

1. C. Bao, Y. W. (2012). Real time robust l1 tracker using accelerated proximal gradient approach. *IEEE Conf*, 1830-1837.
2. Cong Ma, Z. M. (2017). A Saliency Prior Context Model for Real-Time Object tracking. *IEEE*.
3. *Earth mover's distance*. (n.d.). Retrieved from www.wikipedia.org: https://en.wikipedia.org/wiki/Earth_mover%27s_distance
4. Hongbin Zha, X. C. (n.d.). Robust mean shift tracking based on apperance model. Springer.
5. K. Zhang, L. Z.-H. (2012). Real-time compressive tracking. *Springer* (pp. 864–877). Proc. Eur. Conf. Comput. Vis. .
6. L. Sevilla-Lara and E. Learned-Miller. (2012). Distribution fields for tracking. *IEEE Conf.*, (pp. 1910-1917).
7. rivlin, a. A. (2006). Robust Fragment-based Tracking using Integral Histogram. *IEEE.*
8. Rivlin, A. A. (2006). Robust Fragments-based Tracking using the Integral Histogram. *IEEE.*
9. S. Hare, A. S. (2011). Struck: Structured output tracking. (pp. 263-270). IEEE 13th Int. Conf. Comput. Vis.,.
10. T.collins, R. (2003). Mean-shift Blob Tracking through Scale Space. IEEE.
11. X. Hou, J. H. (2012). Image signature: Highlighting sparse. *Pattern Anal. Mach. Intell*, IEEE Trans.