International Journal of Interdisciplinary Innovative Research &Development (IJIIRD)
ISSN: 2456-236X
Vol. 05 Special Issue 01 | 2020

Robust Visual Tracking System via Weighted Spatio – Temporal Context Learning Algorithm

Mr. Shubham Falke[1]

[1] Lecturer, Computer Department, Government Polytechnic Khamgaon, Maharashtra, India

ABSTRACT

Designing a robust visual tracker is a challenging problem due to many disturbed factors such as illumination changes, appearance changes, rotation, partial or full occlusions, etc. Among numerous existed trackers, correlation filter based tracker is a fast and robust method with resistance to the above-mentioned factors. Motivated by that, spatio-temporal context (STC) learning algorithm is proposed, which considers the information of the context around the target and achieved better performance. However, STC treats the whole region of the context equally, which weakens the effectiveness of the context information. In this paper, we propose a novel weighted spatio-temporal context (WSTC) learning algorithm. Our algorithm considers the surrounding context discriminatively and integrates a weighted map by evaluating the importance of different regions. Extensive experimental results on various benchmark databases show that our algorithm outperforms the STC algorithm and the other state-of- the-art algorithms.

Keyword : - STC, WSTC, Visual Tracking, Robust, Context , etc

1. INTRODUCTION

Visual tracking is a hot topic in computer vision and attracts many researchers over the world. In an unconstrained environment, there are a lot of challenge factors to consider when designing a robust tracker, including illumination changes, appearance changes, rotation, partial or full occlusions, etc. To deal with these challenges, Correlation filter is a class of visual tracking algorithms which demonstrates good performance. Average of Synthetic Exact Filters and Unconstrained Minimum Average Correlation Energy are two representative algorithms, after that, Minimum Output Sum of Squared Error was proposed which is robust to variations in lighting, scale, pose, and non-rigid deformations while operating at 669 frames per second.

Visual tracking is a future of research in imaging due to its wide range of applications such as motion analysis, activity recognition, surveillance and human computer interaction. Visual tracking is one of the most active research topics due to its wide range of applications such as motion analysis, activity recognition, surveillance, and human computer interaction. The main challenge for robust visual tracking is to handle large appearance changes of the target object and the background over time due to occlusion, illumination changes, and pose variation. Tracking algorithms can be classified as either discriminative or generative methods, which makes trade-offs between effectiveness and efficiency of an appearance model. The tracking problem is formulated by computing a confidence map, spatial context model and context prior model.

Human visual system that exploits context to help resolving ambiguities in complex scenes efficiently and effectively is the first stage in algorithmic model to compute the spatio-temporal relationships between the object and its local contexts. To design robust visual tracker is a challenging problem, there are many factors such as illumination changes, appearance changes, rotation, partial or full occlusions which causes disturbance in accuracy. Among existing trackers, correlation filter based tracker is a fast and robust method with provides resistance earlier mentioned factors.

2. SPATIO-TEMPORAL CONTEXT LEARNING

Spatio-Temporal Context (STC) Learning algorithm is proposed most recently, which translates the tracking task into a process of locating the target center by calculating a confidence map at every frame. In the current frame, suppose the object center location is , then we can define a context feature set as

$$X^c = \{v(z) = (I(z), z) | z \in \Omega_c(x^*)\}$$

Where I(Z) denotes image intensity at location z and $\Omega_c(x^*)$ (see Figure 1 (a)) is the local context region around of the location x^*. By basic probabilistic formulation, we know that the confidence map c(X) can be represented by

$$\begin{aligned} c(x) &= \sum_{v(z)\in X^c} P(x, v(z)|o) \\ &= \sum_{v(z)\in X^c} P(x|v(z), o)P(v(z)|o) \quad (1) \end{aligned}$$

where x and z are 2D location coordinates, o is the object and the spatial context model $P(x,v(z)|0)$ is the conditional probability that models the spatial relationship between the object location and its context information, and $P(v(z)|0)$is the context prior probability which models the appearance of the local context. So the main task in STC is to learn $p(x,v(z)|0)$ as it is the bridge between the target location and the spatial context.

2.1 Proposed Architecture

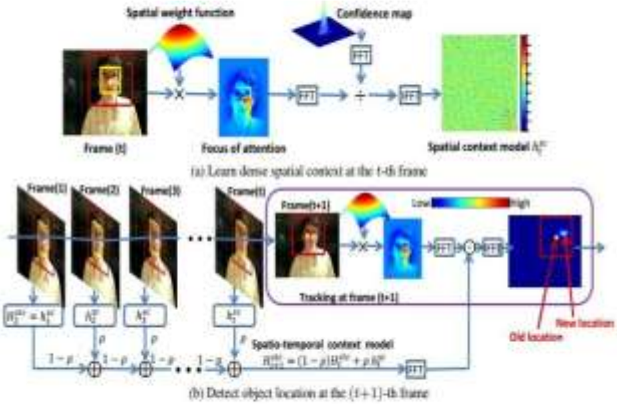


Fig -1: Basic flow of our tracking algorithm

In this paper, we propose a fast and robust tracking algorithm which exploits dense spatio-temporal context information. Figure 1(a) illustrates the basic flow of our algorithm. First, we learn a spatial context model between the target object and its local surrounding background based on their spatial correlations in a scene by solving a deconvolution problem. Next, the learned spatial context model is used to update a spatio-temporal context model for the next frame. Tracking in the next frame is formulated by computing a confidence map as a convolution problem that integrates the dense spatio-temporal context information, and the best object location can be estimated by maximizing the confidence map Finally, based on the estimated confidence map, a novel explicit scale adaptation scheme is presented, which renders an efficient and accurate tracking result.

The key contributions ofthe proposed algorithm are summarized as follows:
– To the best of our knowledge, it is the first work to use dense context information for visual tracking and achieves fast and robust results.
– We propose a novel explicit scale update scheme to deal with the scale variations of the target efficiently and effectively.
– The proposed algorithm is simple and fast that needs only 4 FFTs at 350 FPS in MATLAB.
– The proposed algorithm has the merits of both generative and discriminative methods. On the one hand, the context includes target and its neighbor background, thereby making our method have the merits of discriminative models. On the other hand, the context is a whole of target and background, rendering our method the merits of generative models.

3. OBJECTIVES

- A fast and robust algorithm for tracking of the target object.
- A Spatial Context Prior Model.
- Compute Confidence Map.
- Update of Spatio-Temporal Context by scaling

4. PROPOSED SYSTEM

4.1 Problem formulation:

The problem is implementation of robust visual tracker means such tracker that can be useful under extreme challenging conditions for computer vision; the tracker is applicable in video as well as image sequences. We implement proposed system using weighted spatio-temporal context learning algorithm.

4.2 Proposed system workflow :

Step 1. Start
Step 2. Select whether to load image sequence or load video
Step 3. For image sequence
    a. Select the option load image sequence

b. The tracking process start
   i. First initialize the boundary of object to be tracked
   ii. Set boundaries and build confidence map
   iii. Go for get context so to set boundaries and normalize with weighted map
   iv. Use spatio-temporal context model and build visualization
   v. Position rectangle and build colormap with output

Step 4. For load video part
   a. Select option load video
   b. Choose video we consider only avi/frames files.
   c. We use optical flow method for tracking
   i. Read a file and in its binary form
   ii. Convert file into RGB format so video frames generated are used.
   iii. Build system object Blob analysis for gaps removal in neighborhood
   iv. Next use Morphological erosion to thin out portions
   v. Write out how many object in given case cars are tracked down.

Step 5. Select exit option to quit application.
Step 6. Continue with above steps for different videos to be tested.

5. EXPERIMENTS

In this section we evaluated our WSTC algorithm on 10 video sequences with 5000 frames in total. These videos sequences contain a variety of challenging factors, such as illumination changes, background cluster, heavy occlusion, non-rigid deformation, motion blur. ground truth provided by for all sequences and the parameters of our WSTC algorithm are fixed in all the experiments. We use three popular evaluation criteria to quantitatively evaluate the performance of our trackers, i.e. centre locationerror (CLE), Pascal VOC overlap ratio (VOR) and VOC tracking success rates (SR). VOC overlap ratio is defined as,

$$R = \frac{\mathrm{Area}(B_t \cap B_g)}{\mathrm{Area}(B_t \cup B_g)},$$

where Bt is the tracked bounding box and Bg is the ground truth bounding box. When the VOC overlap ratio is larger than 0.5, the tracking result of one frame is considered to be successful. For space limitation, we only show the result of centre location error (CLE) in this paper (see Table 1), more results of VOC overlap ratio and success rates are shown in supplemental material. From these experimental results, we can easily find that the proposed tracking algorithm outperforms the STC algorithm and achieves the best or second tracking performance in most video sequences.

The implementation of proposed work is done using MATLAB, for conducting experiments we had used the standard dataset available which comprises of videos and sequences that where having different illumination challenges, having variation in contrast, background and blurring, so that for tracking of object subject to different challenges.

6. RESULTS

The video is divided into n number of frames which are stored in .jpeg format in one of the folders which is a prerequisite of our project. Firstly, we initialize the object of our interest by marking the co-ordinates of the rectangle in the first frame. The proposed system tracks the target object in every frame irrespective of occlusion, pose variations, illumination change

Fig 2: Main GUI of proposed system

For tracking we have suggest two options to load image sequence and second to load video, the different types of videos used sample is as shown in table 1.

Table 1 :DESCRIPITON ABOUT VIDEOS USED FOR TRACKING

Sr. No.	Video filename	Bit rate	Dimensions in pixels	No. of frames generated for image sequence
1	car_traffic.avi	238 kbps	160 x 120	120
2.	street.avi	240 kbps	320 x 240	390
3.	highway.avi	240 kbps	320 x 240	540

For the tom_miker.avi video we have about 800 frames generated which are having illumination challenges out of these for tracking of image sequence that his shows sample image frame sequence used in tracking of object.

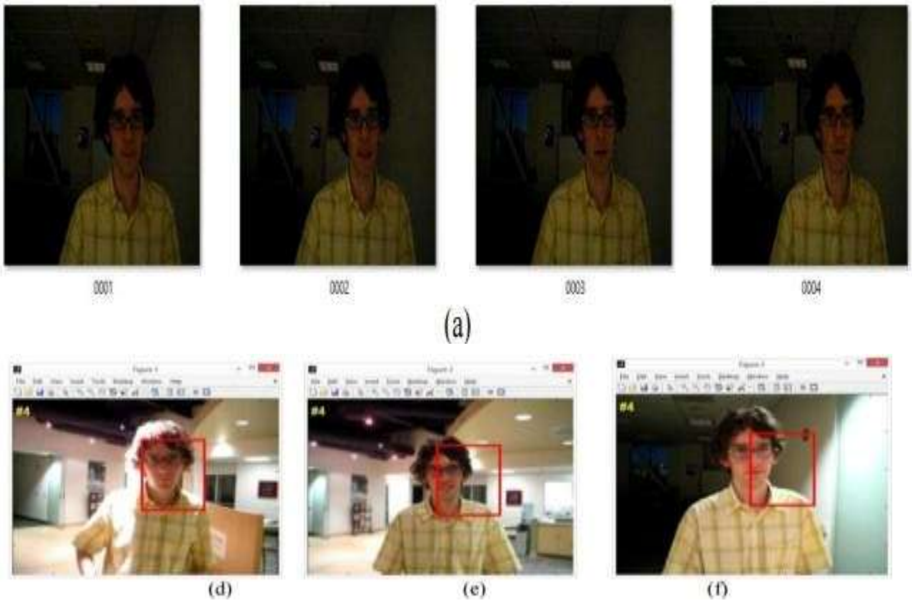


Fig. 3: Results of tracking in image frame sequences for tim_miker.avi image sequence.

Next part is about tracking in for video, after clicking on load video option, after background intermediate steps we have the generated result sequences as shown below in figure 4 for videos tabulated in table 1.

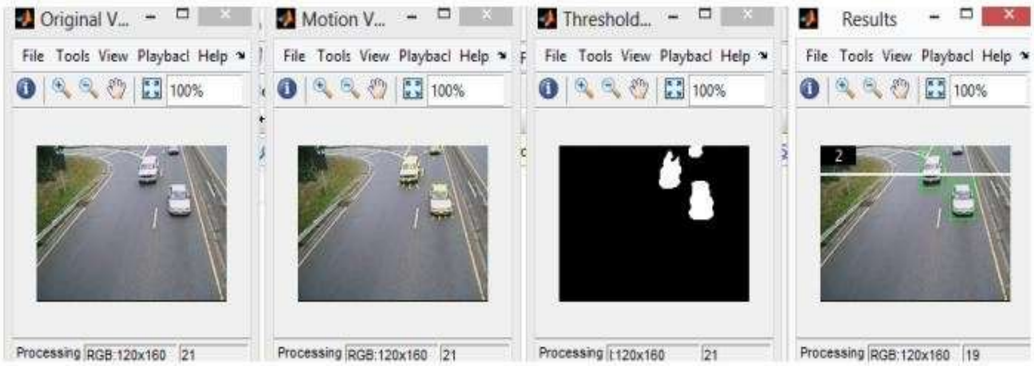
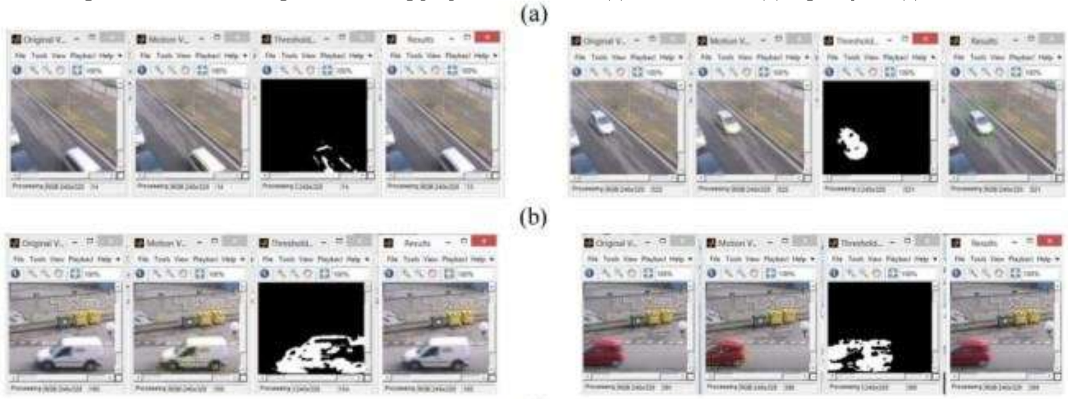


Fig.4 : Results of tracking in video using proposed method for (a) car_traffic.avi (b) highway.avi (c) street.avi

(a)



(c)

7. CONCLUSION

In a paper study a robust algorithm which exploits dense spatio-temporal context information for visual tracking. Two local context models i.e., Spatial context and spatio-temporal context models were proposed which are robust to appearance

variations introduced by occlusion, illumination changes, and pose variations. An explicit scale adaptation scheme is proposed which is able to adapt target scale variations effectively. The Fast Fourier Transform algorithm was used in both online learning and detection, resulting in an efficient tracking method that runs at 35 frames per second with MATLAB 12.0 implemented on i5 machine. Numerous experiments with state of-the- art algorithms on challenging sequences demonstrated that the proposed algorithm achieves favorable results in terms of accuracy, robustness, and speed.

8. REFERENCES

[1] David S Bolme, Bruce A Draper, and J Ross Beveridge,―Average of synthetic exact filters,‖ in Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009, pp. 2105–2112.

[2] AbhijitMahalanobis, BVK Vijaya Kumar, Sewoons Song, SRF Sims, and JF Epperson, ―Unconstrained correlation filters,‖ Applied Optics, vol. 33, no. 17, pp. 3751–3759, 1994.

[3] David S Bolme, J Ross Beveridge, Bruce A Draper, and Yui Man Lui, ―Visual object tracking using adaptive correlation filters,‖ in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, 2010, pp. 2544–2550.

[4] Kaihua Zhang, Lei Zhang, Ming-Hsuan Yang, and David Zhang, ―Fast tracking via spatio-temporal context learning,‖ arXiv preprint arXiv:1311.1939, 2013.

[5] ChenglongBao, Yi Wu, Haibin Ling, and Hui Ji,

[6] ―Real time robust l1 tracker using accelerated proximal gradient approach,‖ in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 1830–1837.

[7] Amit Adam, Ehud Rivlin, and IlanShimshoni,

[8] ―Robust fragments-based tracking using the integral histogram,‖ in Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on. IEEE, 2006, vol. 1, pp. 798–805.

[9] ShaulOron, Aharon Bar-Hillel, Dan Levi, and Shai Avidan, ―Locally orderless tracking,‖ in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 1940–1947.

[10] David A Ross, Jongwoo Lim, Ruei-Sung Lin, and MingHsuan Yang, ―Incremental learning for robust visual tracking,‖ International Journal of Computer Vision, vol. 77, no. 1-3, pp.125–141, 2008.