Real Time Eye Tracking using Kalman Extended Spatio-Temporal Context Learning

Farzeen Munir[a], Fayyaz ul Amir Afsar Minhas[b], Abdul Jalil[c], Moongu Jeon*[a]
[a]School of Engineering and Computer Science, Gwangju Institute of Science and Technology, South Korea. [b] Depatment of Computer and Information Sciences, Pakistan Institute of Engineering and Applied Sciences, Pakistan. [c] Department of Electronic Engineering, International Islamic University, Pakistan.

ABSTRACT

Real time eye tracking has numerous applications in human computer interaction such as a mouse cursor control in a computer system. It is useful for persons with muscular or motion impairments. However, tracking the movement of the eye is complicated by occlusion due to blinking, head movement, screen glare, rapid eye movements, etc. In this work, we present the algorithmic and construction details of a real time eye tracking system. Our proposed system is an extension of Spatio-Temporal context learning through Kalman Filtering. Spatio-Temporal Context Learning offers state of the art accuracy in general object tracking but its performance suffers due to object occlusion. Addition of the Kalman filter allows the proposed method to model the dynamics of the motion of the eye and provide robust eye tracking in cases of occlusion. We demonstrate the effectiveness of this tracking technique by controlling the computer cursor in real time by eye movements.

Keywords: Visual Object Tracking, Eye Tracking, Kalman Filter

1. INTRODUCTION

Eye Tracking is an important application of Human-Computer Interaction [1]. It provides a fast, effortless and non-invasive way of communication with computers. The motivation of our work is to provide a low budget, easy to use and wearable eye tracker for disabled people, so that they can use computer applications by controlling the cursor through eye movements [2].

Over the years, eye tracking has captured the attention of the researchers but biological characteristics of the eye impose a major challenge on real time eye tracking. The complex and saccadic motion of the eye makes real time eye tracking a non-trivial task [3]. Eyes perform sudden and fast movements followed by fixation.

In the literature, feature based and model based tracking techniques have been used for eye tracking [4]. We focus on correlation based techniques that uses template matching to track the eye [5]. However, the classic correlation technique cannot handle to occlusion and template drift. For example, the correlation based Spatio-Temporal Context Learning (STC) approach by Zhang et al. [6] is computationally efficient and robust for visual object tracking. However, it fails to capture rapid eye movements when used for eye tracking as it does not consider the dynamics of the target.

We hypothesized that Kalman filter [7] in combination with a correlation based tracker would overcome these shortcomings. The Kalman filter can model target dynamics and can predict the location of the eye when the eye is occluded due to eye blinking or glare. It can also estimate the correct location of the eye, using raw measurements from the correlation based tracker. Based on this idea, we have developed a Kalman Extended Spatio-Temporal Context Learning algorithm (KSTC) for tracking.

The organization of the paper is as follows. Section 2 discussed the implementation details as well as the performance evaluation method for KSTC. Section 3 presents the result. Conclusion of the paper are presented in section 4. results.

*mgjeon@gist.ac.kr;

This work was supported by the ICT R\&D Program of MSIP/IITP (Grant No. B0101-15-0525, Development of global multi-target tracking and event prediction techniques based on real-time large-scale video analysis) and Center for Inte-grated Smart Sensors as Global Frontier (CISS-2013M3A6A6073718)

2. MATERIAL AND METHODS

In this section, we discuss the algorithmic details of the proposed method.

2.1 Spatio-Temporal Context Learning

Our proposed method is based on the Spatio-Temporal Context Learning method [6]. The details of this method are shown in figure 1(a). In this approach the target is identified in the first frame by prior knowledge. A context window surrounding the target is then extracted. The context window is twice the size of the target, which reduces the computation cost in comparison to using the whole frame and provides a robust template. The context window is multiplied by a spatial weight function to obtain a weighted context model. The Spatial weight function emphasizes the location of the target by assigning a larger weight to target pixels as compared to other pixels. The Fourier representation the weighted context model for frame t is denoted by F_t. A Confidence map indicates the likely location of the target in the weighted context window. The Fourier representation of confidence map is denoted by G_t. The Spatial Context model, H_t is generated using equation 1. The Spatio context model is convoluted with the next frame to obtain the target location by generating Spatio-Temporal context model h_{t+1} using equation 2.

$$H_t = F_t / G_t. \tag{1}$$

$$h_{t+1} = H_t + (1 - \sigma) h_t. \tag{2}$$

Here, σ is the rate of update of the Spatio-Temporal context model. Its value is selected between 0-1. If it is too small, the template will update slowly and template update would not be able to incorporate fast appearance changes in the target. If σ is large, the new template almost replaces the previous template. This tracking filter is unable to handle occlusion and does not model motion dynamics.

2.2 Kalman Extended Spatio-Temporal Context Learning

To overcome the shortcoming of STC, we have extended it by using a Kalman filter. The Kalman filter incorporates the dynamics of the target and helps in predicting the target location in case of occlusion. It can also aid in estimating the correct location during rapid eye movements.

The Kalman filter is a linear quadratic estimator that uses a number of previous measurements to determine the next possible state of the system [7]. The objective in visual object tracking is to track the target location through consecutive frames. If we know the motion dynamics of the target, it is possible to predict the most likely location of the target in the next frame, thus narrowing down the search. Kalman filter predicts the location of the target in case of occlusion, resulting in robust tracking [8]. For modelling the dynamics, we use the equations of the motion proposed by Ahmad. J [9].

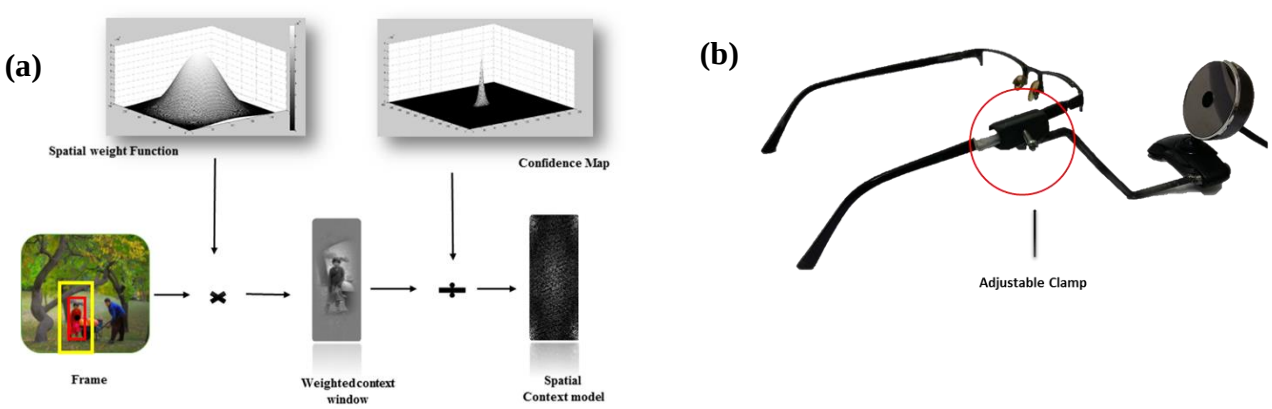

Figure 1: (a) shows Spatio Temporal Context Learning. (b) shows the prototype of eye tracker.

$$X_t = [\, x_t \; x_t' \; x_t'' \; y_t \; y_t' \; y_t'']^T. \tag{3}$$

$$X_{kt} = AX_{kt-1} + W_{kt-1}. \tag{4}$$

$$Z_{kt} = HX_{kt} + V_{kt}. \tag{5}$$

$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 0.5 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0.5 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{6}$$

$$Z_{kt} = [x_{kt} \;\; y_{kt}]^T. \tag{7}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}. \tag{8}$$

X_k, given by equation 3, consists of position, velocity and acceleration along x-axis and y-axis. The equation 4 and 5 gives the target state and observation equations, respectively. A is the transition matrix given by equation 6, which consists of equations of motion for position, velocity and acceleration of the target. Z_k is the observation matrix given by 7. H is the output matrix shown by 8. W_k and V_k models the error in the state and output estimates. We use Gaussian random variable to model W_k and V_k with mean of zero and unity standard deviation.

In our proposed method the location of the target obtained by STC is used as a state vector X_{kt} for Kalman Filter. Z_{kt} is the estimated location of the target depending upon number of previous states. The context window of the correlative filters is set such that it covers the complete range of eye movements. This helps in reducing the computation cost and also provides robust eye tracking. Since the eye does not undergo any major appearance changes, so its template is not updated.

Initialization of the Filter. As the application is developed for people with mobility issues manually selecting the target is not possible. To automate the target selection Minimum Output Sum of Square (MOSSE) correlation filter introduced by Bolme et al. is used [10]. It uses number of training images and their output target images to generate a filter, which when correlated with test image correctly finds the center of the eye.

Hardware Design of Eye Tracker. Figure 1 (b) shows a prototype of the eye tracker. It is a wearable eye tracker that is Cost-effective, easy to use and puts no restriction on the user. A4tech PK-520F camera is mounted on the pair of glasses with help of adjustable clamp, hence giving flexibility to optimize the camera position. The most significant advantage of the prototype is that it provides stability, and can be attached to any pair of glasses. A polarizing filter is placed in front of a camera to block polarized light from the computer screen. This is done to reduce screen glare.
The camera is connected to a computer through a USB port. The cursor control is implemented by mapping the image coordinates to the screen coordinates system. The protocol is implemented in python module. Eye ball coordinates are translated to pixel coordinates of screen.

Evaluation Measures.

The performance of Kalman Extended Spatio-Temporal Context Learning is evaluated on two datasets 1) general object videos [11] and 2) eye movement videos [12]. There are 6 videos in first dataset. These videos are of general objects which contain occlusion, appearance changes, fast movement and illumination variation. The second dataset consists of 4 videos. The videos are recorded in infrared spectrum and exhibit occlusion and rapid eye movement. This dataset helps to evaluate the performance of Kalman Extended Spatio-Temporal Context Learning.

Apart from the two benchmark data sets, we evaluate the performance of KSTC in real time by performing eye tracking on the human test subject. The test subjects used the designed hardware to control the cursor through eye movement.

To evaluate the accuracy of different algorithms on the videos in our data set, we have used center location error and Pascal score (PS). Center location error (CLE) is the mean of the error between actual center of the target and measured center of target. Pascal score measures the degree of overlap of the detected and actual target regions. Its value lies between 0-1. Score above 0.5 are good. Both these measures are widely used in the literature for evaluating the performance of tracking algorithms [6].

3. RESULTS AND DISCUSSION

3.1 Standard Object Tracking and Eye Tracking Results

The performance of the KSTC is compared with STC on the set of six videos. The quantitative result of the comparison is shown in Table 1 (a). On all six videos, KSTC performs better than the STC. STC work well in videos where there is the slow movement like in “David” but where there is a fast movement like in “Sylv” it does not track the target efficiently. In the case of occlusion for the “Face”, it loses the track of target and focuses on the background. The “Girl” and “Face” contains the issue of occlusion. The Pascal score of 0.84 is obtained for the “Girl” and 0.71 in “Face”, which is the significant improvement from STC score. Figure 2(a) shows the visual results of a “Girl”.

Table 1 (b) shows the Pascal score and center location error for the STC and KSTC on offline eye movement videos. Results show that KSTC provides robust eye tracking as compared to the STC. Figure 2 (b) shows visual results, comparing tracking with spatial context learning and Kalman Extended spatial context learning. STC accurately tracks the eye when it is moving slowly. However, when it makes rapid movements and gets occluded by eye lid, the tracking is lost. The results verify our hypothesis that Kalman filter in combination with spatial-temporal context learning performs better. It provides robust tracking in cases of occlusion and fast movement.

Table 1**:** a) and b) showing quantitative results.

(a)	STC		KSTC	
Video	PS	CLE	PS	CLE
David	0.69	7.84	0.74	7.60
girl	0.48	15.17	0.84	4.97
dog	0.18	10.89	0.21	10.00
face	Failed	Failed	0.71	11.17
sylv	0.28	13.65	0.66	7.96
Panda	0.99	1.20	0.99	0.97

(b)	STC		KSTC	
Video	PS	CLE	PS	CLE
Eye 1	0.49	55.4	0.93	18.8
Eye 2	0.56	44.5	0.92	20.9
Eye 3	0.37	51.9	0.92	19.1
Eye 4	0.63	56.1	0.90	21.1
Note: offline eye tracking				

3.2 Real Time Eye Tracking

Kalman Extended Spatio-Temporal Context Learning is tested in real time by using the prototype of the eye tracker. We were able to control the cursor through eye movements. It was used to write in a software called "Dasher" [13]. Dasher is an interface that helps to write with the mouse, providing disable people to express themselves. In implementing the Kalman Extended Spatio-Temporal Context Learning in real time we face challenges like illumination variation and screen glare, since the user is controlling the cursor through the eye movements and looking at the screen. The glare of the screen makes a bright spot on the eye. This bright spot does not move with the eye movements as the screen is fixed. The tracker tracks the bright spot and does not register the eye movements accurately. The glare is removed by using a simple Polarizing Filter.

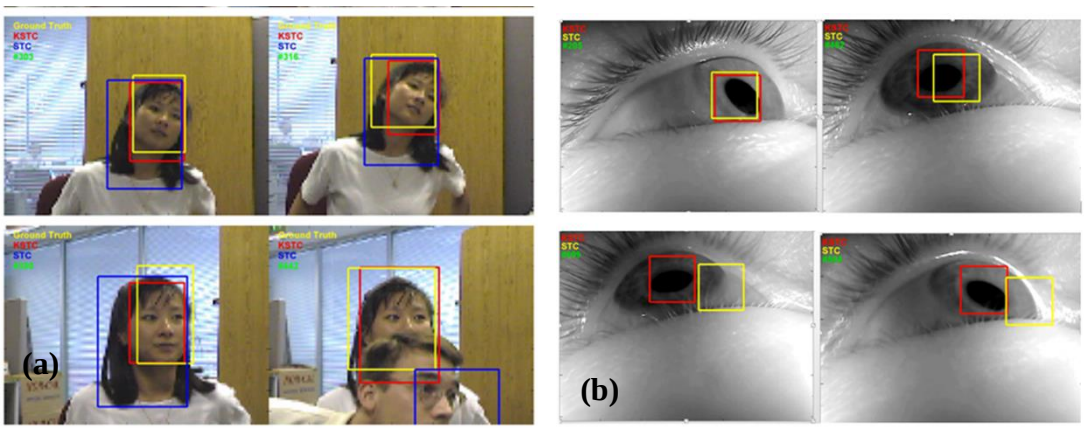

Figure 2: (a) shows tracking result on the "Girl". (b) shows tracking result on the eye video.

4. CONCLUSION

This paper presents a Visual object tracking application in which cursor control was implemented to track the eye movements. A prototype of wearable eye tracker was designed to help disabled people to communicate through their eye movement. For real-time eye tracking, we proposed Kalman Extended Spatio-Temporal Context Learning. From the experimental results it has been shown that KSTC is fast and better adapted to handle occlusion and template drift.

REFERENCES

[1] A. T. Duchowski, "A breadth-first survey of eye-tracking applications," *Behav. Res. Methods Instrum. Comput. J. Psychon. Soc. Inc*, vol. 34, no. 4, pp. 455–470, Nov. 2002.

[2] M. Betke, J. Gips, and P. Fleming, "The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities," *IEEE Trans. Neural Syst. Rehabil. Eng. Publ. IEEE Eng. Med. Biol. Soc.*, vol. 10, no. 1, pp. 1–10, Mar. 2002.

[3] D. J. Singh and H. Singh, "Human Eye Tracking and Related Issues: A Review," *Int. J. Sci. Res. Publ.*, vol. 2, 2012.

[4] A. Kennedy, "Eye tracking: A comprehensive guide to methods and measures," *Q. J. Exp. Psychol.*, Feb. 2016.

[5] M. Chau and M. Betke, "Real time eye tracking and blink detection with USB cameras," *Tech. Rep. 218 Boston Univ. Comput. Sci. Dep.*, 2005.

[6] K. Zhang, L. Zhang, Q. Liu, D. Zhang, and M.-H. Yang, "Fast Visual Tracking via Dense Spatio-temporal Context Learning," presented at the European Conference on Computer Vision, 2014, pp. 127–141.

[7] H. E. RAUCH, C. T. STRIEBEL, and F. TUNG, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, May 2012.

[8] A. Ali, A. Jalil, J. Ahmed, M. A. Iftikhar, and M. Hussain, "Correlation, Kalman filter and adaptive fast mean shift based heuristic approach for robust visual tracking," *Signal Image Video Process.*, vol. 9, no. 7, pp. 1567–1585, Oct. 2015.

[9] J. Ahmad, "Adaptive Edge-Enhanced Correlation Based Robust and Real-Time Visual Tracking 234 Framework and its Deployment in Machine Vision Systems," PhD, Military College of Signals, Pakistan, 2008.

[10] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," presented at the Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 2544–2550.

[11] P. Poschmann, P. Huber, M. Rätsch, J. Kittler, and H.-J. Böhme, "Fusion of Tracking Techniques to Enhance Adaptive Real-time Tracking of Arbitrary Objects," *Procedia Comput. Sci.*, vol. 39, pp. 162–165, Jan. 2014.

[12] L. Świrski, A. Bulling, and N. Dodgson, "Robust real-time pupil tracking in highly off-axis images," presented at the Proceedings of the Symposium on Eye Tracking Research and Applications, 2012, pp. 173–176.

[13] "Inference Group: Dasher Project: Home." [Online]. Available: http://www.inference.phy.cam.ac.uk/dasher/. [Accessed: 02-Feb-2017].

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website
Farzeen Munir	PhD candidate	Computer Vision and Machine Learning	http://mlv.gist.ac.kr/
Fayyaz ul Amir Afsar Minahs	Senior Scientist	Machine Learning and Bioinformatics	http://faculty.pieas.edu.pk/fayyaz/
Abdul Jalil	professor	Image Processing and Machine Learning	http://www.iiu.edu.pk/index.php?page_id=2062
Moongu Jeon	professor	Computer Vision and Machine Learning	http://mlv.gist.ac.kr/