Hierarchical HMM
for Eye Movement Classification

Ye Zhu[1], Yan Yan[1], and Oleg Komogortsev[1]

Texas State University, San Marcos, USA
{ye.zhu, tom_yan, ok}@txstate.edu

Abstract. In this work, we tackle the problem of ternary eye movement classification, which aims to separate fixations, saccades and smooth pursuits from the raw eye positional data. The efficient classification of these different types of eye movements helps to better analyze and utilize the eye tracking data. Different from the existing methods that detect eye movement by several pre-defined threshold values, we propose a hierarchical Hidden Markov Model (HMM) statistical algorithm for detecting fixations, saccades and smooth pursuits. The proposed algorithm leverages different features from the recorded raw eye tracking data with a hierarchical classification strategy, separating one type of eye movement each time. Experimental results demonstrate the effectiveness and robustness of the proposed method by achieving competitive or better performance compared to the state-of-the-art methods.

Keywords: Hidden Markov Model, eye movement, fixation, smooth pursuit, saccade, classification

1 Introduction

Eye tracking technology, which aims to measure the location where a person is looking at, has been widely applied in various research and application fields including the human-computer interaction [16, 4], AR/VR [6, 11], behavioral psychology [10, 3] and usability studies [21, 8] in recent years. With the arising attention and interests from researchers, eye tracking is becoming a potential and promising driver for future immersive technologies.

One fundamental and significant research topic in eye tracking is to identify different types of eye movements. There are several primary types of eye movements: fixations correspond to the situation where the visual gaze is maintained on a single location, saccades are fast movements of the eyes that rapidly change the point of fixation, and smooth pursuits are defined as slower tracking movements of the eyes designed to keep a moving stimulus on the fovea [17]. Ternary eye movement classification [13], which seeks to classify three primary types of eye movement, *i.e.*, fixations, saccades and smooth pursuits, is essential to the above applications.

In this work, we tackle the problem of ternary eye movement classification from a probabilistic perspective by adopting the Hidden Markov Model in a

hierarchical way. The hierarchical structure makes it possible to consider several different data features in different stages of classification. The usage of the Viterbi [9] and Baum-Welch algorithms [2] allows us to avoid the inconvenience of selecting thresholds and to improve the robustness of the classification method. Experiments show that our proposed hierarchical HMM is able to achieve competitive performance compared to the state-of-the-art methods.

2   Related Work

One of the most common and intuitive methods to separate fixations from saccades is threshold based algorithms, such as Velocity Threshold Identification (I-VT) [1] and Dispersion Threshold Identification (I-DT) [18]. The former method assumes the velocity of saccades should be larger than the velocity of fixations, while the latter one relies on the difference of duration and positional dispersion between fixations and saccades. However, the above single threshold-based algorithms are unable to accurately separate smooth pursuits from fixations due to a variety of artifacts usually present in the captured eye movement signal.

The existing threshold-based state-of-the-art methods for ternary eye movement classification mainly combine several different single threshold-based algorithms. Velocity Velocity Threshold Identification (I-VVT) [13] is a basic algorithm that adopts two velocity thresholds, the data points with higher velocity than the larger velocity threshold are classified as saccades, the points with a lower velocity than the smaller threshold are classified as fixations, while the remaining points are considered as smooth pursuits. Velocity Dispersion Threshold Identification (I-VDT) [13] is another algorithm that combines I-VT and I-DT [13]. I-VDT firstly filters out saccades by I-VT, I-DT is then further used to separate fixations from smooth pursuits. Velocity Movement Pattern Identification (I-VMP) [12] uses I-VT to identify saccades, and then employs movement direction information to separate fixations from smooth pursuits. All the methods mentioned above rely on empirically selected thresholds to provide a meaningful classification on a targeted dataset.

In addition to the above threshold-based algorithms, video-based methods to detect gazes have also been exploited in [7, 14]. Dewhurst *et al.* [5] proposes to use geometric vectors to detect eye movements. Nystrom *et al.* [15] aims to classify fixations, saccades and glissades using an adaptive velocity-based algorithm. Santini *et al.* [20] propose Bayesian method (I-BDT) to identify fixations, saccades and smooth pursuits. Another branch of more recent work adopts machine learning techniques to tackle the problem. Identification using Random Forest machine learning technique (IRF) is used to classify fixations, saccades and post-saccadic oscillations in [24]. Zembly *et al.* [23] further propose a gaze-Net to realize end-to-end eye-movement event detection with deep neural networks. Startsev *et al.* [22] tackles the problem of ternary eye movement classification with a 1D-CNN with BLSTM.

Although the machine learning methods, especially deep learning, have been widely applied in multiple research fields including the eye movement classifica-

---

Algorithm 1 Hierarchical HMM for Ternary Eye Movement Classification

---

Require: Eye positional data sequence.

1: ε_n number of epochs for n-th HMM, initial start probability vector π, initial transition probability matrix $\mathbf{A}$ and initial emission probability matrix $\mathbf{B}$. Note that initial parameters for HMM will be optimized and updated by Baum-Welch algorithm.

2: **Step 0: Pre-processing**

3: Compute the position, velocity, acceleration feature sequences for input eye tracking data

4: Select appropriate features for classification.

5: **Step 1: Rough classification**

6: Initialize the parameters of the first HMM for selected feature

7: $e_1 \leftarrow 0$

8: **for** $e_1 < \varepsilon_1$ **do**

9:     Viterbi algorithm

10:     Baum-Welch algorithm

11:     Filter saccades

12: **end for**

13: **Step 2: Refined classification**

14: Initialize the parameters of the second HMM for selected feature, define a threshold value $\mathcal{T}$ of the first feature for fine-tuning

15: $e_2 \leftarrow 0$

16: **for** $e_2 < \varepsilon_2$ **do**

17:     Viterbi algorithm

18:     Baum-Welch algorithm

19:     Classify fixations and smooth pursuits

20:     Fine-tune the classification results by $\mathcal{T}$

21: **end for**

22: **Step 3: Merge function**

23: Merge classified points into complete fixations, saccades and smooth pursuits

24: **return** List of classification results

---

tion, one major drawback of these techniques is that the training process relies on a large amount of data. In addition, the neural networks usually require re-training when applied in a different task or dataset. Our focus in this work is to improve the robustness and performance of the threshold-based state-of-the-art methods, even competing with the recent machine learning based state-of-the-art performance.

3 Methodology

We present the proposed hierarchical HMM method in this section and provide the corresponding pseudo-code.

Hidden Markov Model is a statistical model for time series based on the Markov process with hidden states. The principle of HMM is to determine the hidden state with a maximum probability according to the observable sequence.

A traditional HMM takes three sets of parameters as input, which are the start probability vector, transition probability matrix and emission probability matrix. The ternary eye movement classification can be formulated as a first-order three-state HMM problem, whose hidden states are fixations, saccades and smooth pursuits.

HMM relies on the distinguishable probability distributions of features to correctly define different hidden states, otherwise, the maximum probability of each hidden state may be incorrect if several hidden states have similar probabilities given a certain observation sequence. In the case of eye movement classification, the probability distributions for different movement types are usually represented by continuous Gaussian distributions [19]. The main challenge of ternary eye movement classification using the existing methods lies in the bias of the features for smooth pursuits. While fixations and saccades have very different positional dispersion and velocity features, smooth pursuits are rather ambiguous in terms of dispersion and velocity since the steady state of a smooth pursuit often contains corrective saccades, making the eye tracking data very noisy. To this end, we propose to leverage different features and introduce a hierarchical strategy to tackle the problem. The core idea of our hierarchical HMM is to perform HMM classification for multiple times using different features.

Firstly, we start by analyzing different features of the raw eye positional data as the pre-processing step, *e.g.*, positions, velocities, accelerations, whose objective is to determine the features that can be used to separate three eye movement types in latter steps. After selecting appropriate features from preprocessing, we then perform a first-stage rough classification on the eye tracking data to separate saccades from the fixations and smooth pursuits using the HMM. We refer the first-stage as rough classification due to the reason that the classification results will be fine-tuned in the latter stage. The second-stage classification adopts a different feature to separate fixations and smooth pursuits by another HMM, using the first feature as a fine-tuning criterion at the same time. The final step of our hierarchical HMM is to merge the classified points into complete eye movements with temporal duration criterion.

The pseudo-code of the proposed method is presented in Algo. 1. Compared with the existing threshold-based state-of-the-art methods, our method has several advantages: 1) the usage of HMM avoids the tedious work to select threshold values as for previous algorithms and improves the robustness of the proposed classification method; 2) a hierarchical strategy with fine-tuning technique further improves the classification performance. In the meanwhile, the proposed method does not rely on a large amount of data for training and leverages several different features in a coherent way (*e.g.*, use the feature from the first stage as a fine-tuning criterion in the second stage). Experimental results prove that our proposed hierarchical HMM method is simple, straightforward yet effective.

4 Experiments

In this section, we present the experimental results obtained with the proposed method. The comparisons with other state-of-the-art methods demonstrate the robustness and effectiveness of the proposed hierarchical HMM. An ablation study is also included to show the contributions of the hierarchical structure.

4.1 Dataset and Evaluation Metrics

We use the eye tracking dataset recorded from the previous research work [13] for experiments. The data are recorded by the EyeLink 100 eyetracker at 1000 Hz on a 21-in monitor and contain 11 subjects, containing the human annotations as either clean or noisy data.

Seven different behavior scores are used as quantitative evaluation metrics for our experiments. Saccade quantitative score (SQnS), fixation quantitative score (FQnS) and smooth pursuit quantitative score (PQnS) are used to measure the amount of saccades, fixation and smooth pursuit behavior in response to a stimulus, respectively [12]. Fixation qualitative score (FQlS), smooth pursuit qualitative score for positional accuracy (PQlS_P) and for velocity accuracy (PQlS_V) are to compare the proximity of the detected smooth pursuit signal with the signal presented in the stimuli [13]. Misclassified fixation score (MisFix) of the smooth pursuit is defined as the ration between misclassified smooth pursuit points and the total number of fixation points in the stimuli. The ideal number of MisFix should consider the practical latency situation where smooth pursuit continues when the stimulus changes from smooth pursuit to fixation.

4.2 Implementations

We compare our proposed hierarchical HMM method with six baselines: I-VVT [13], I-VDT [13], I-VMP [12], IRF [24], I-BDT [20] and 1DCNN [22]. Among six baselines, I-VVT, I-VDT, and I-VMP are threshold-based methods, all the threshold values are optimized for different subject recordings in our experiments for comparison. For the IRF, we use the same selected features (*i.e.*, velocity and position) as for our proposed method. For the I-BDT, we follow the parameters chosen in [20]. For the 1DCNN, we take the pre-trained model and fine-tune it on the recording data (except those used for testing) from our targeted dataset.

For our proposed hierarchical HMM method, we use the parameters reported in [19] for the initialization in HMM. We iterate the Viterbi and Baum-Welch algorithms for 3 times to learn and update the parameters for HMM. For the final Merge function, the merge time interval threshold we use is 75ms, the merge distance is 0.5°.

4.3 Experimental Results

Results on Behavior Scores. We compare the behavior scores obtained via the proposed hierarchical HMM and other methods on the subjects that are

Table 1. Comparison of behavior scores obtained by different methods for the subjects that are manually evaluated as "Medium", "Good" and "Bad", respectively. Our proposed method achieves competitive performance close to the human evaluation (*i.e.*, Manual).

Behavior scores	Ideal	Manual [13]	Ours	I-VVT [13]	I-VDT [13]	I-VMP [12]	IRF [24]	I-BDT [20]	1DCNN [22]
SQnS	100%	84%	83%	86%	82%	78%	86%	84%	88%
FQnS	84%	63%	63%	16%	79%	66%	78%	68%	59%
PQnS	52%	47%	48%	38%	55%	61%	43%	51%	40%
MixFix	7.1%	13%	8.9%	42%	6%	22%	12%	15%	20%
FQlS	$0°$	$0.46°$	$0.4°$	$0.5°$	$0.5°$	$0.5°$	$0.5°$	$0.5°$	$0.4°$
PQlS_P	$0°$	$3.07°$	$3.2°$	$3.2°$	$3.2°$	$3.4°$	$3.1°$	$3.2°$	$3.2°$
PQlS_V	$0°/s$	$39°/s$	$30°/s$	$16°/s$	$47°/s$	$40°/s$	$38°/s$	$33°/s$	$40°/s$

Behavior scores	Ideal	Manual [13]	Ours	I-VVT [13]	I-VDT [13]	I-VMP [12]	IRF [24]	I-BDT [20]	1DCNN [22]
SQnS	100%	96%	91%	96%	90%	90%	88%	92%	90%
FQnS	84%	71%	74%	30%	82%	69%	82%	80%	78%
PQnS	52%	39%	44%	39%	30%	56%	58%	42%	38%
MixFix	7.1%	6%	5.5%	50%	4.4%	21%	12%	8%	16%
FQlS	$0°$	$0.44°$	$0.3°$	$0.4°$	$0.4°$	$0.4°$	$0.4°$	$0.4°$	$0.4°$
PQlS_P	$0°$	$3.15°$	$2.9°$	$3.6°$	$3.2°$	$3.7°$	$3.7°$	$3.3°$	$3.4°$
PQlS_V	$0°/s$	$23°/s$	$25°/s$	$16°/s$	$44°/s$	$40°/s$	$38°/s$	$37°/s$	$32°/s$

Behavior scores	Ideal	Manual [13]	Ours	I-VVT [13]	I-VDT [13]	I-VMP [13]	IRF [24]	I-BDT [20]	1DCNN [22]
SQnS	100%	89%	77%	85%	78%	76%	75%	82%	85%
FQnS	84%	42%	49%	17%	61%	51%	43%	45%	50%
PQnS	52%	40%	37%	33%	31%	47%	44%	38%	33%
MixFix	7.1%	33%	11%	33%	22%	36%	18%	12%	20%
FQlS	$0°$	$0.58°$	$0.5°$	$0.6°$	$0.6°$	$0.6°$	$0.6°$	$0.6°$	$0.6°$
PQlS_P	$0°$	$2.58°$	$2.6°$	$3.2°$	$3.5°$	$3.6°$	$3.3°$	$2.4°$	$3.5°$
PQlS_V	$0°/s$	$30°/s$	$32°/s$	$16°/s$	$81°/s$	$52°/s$	$46°/s$	$31°/s$	$44°/s$

manually evaluated as "Medium", "Good" and "Bad" in the eye tracking data quality [13]. The ideal scores and manually classification results are also reported for comparison, the calculation of the ideal behavior scores considers the actual physiological reactions of humans when performing different types of eye movements.

The quantitative experimental results are presented in Table 1. We observe that our proposed hierarchical HMM method achieves promising performance compared to the other baseline methods and is much closer to human performance (*i.e.*, manual). The I-VVT algorithm has the worst performance as expected due to the reason that it only considers the velocity and separates the eye movement by two simple threshold values. The I-VDT and I-VMP are able to separate fixations, saccades and smooth pursuits, and I-VDT has better performance compared to I-VMP, achieving a MixFix score closer to the ideal one. Other machine learning based methods IRF, I-BDT, and 1DCNN achieve good performance in general, however, we observe that 1DCNN has relatively lower behavior scores in detecting the smooth pursuit. One possible reason for its under-performance could be the limited amount of smooth pursuit data for fine-tuning. Since the original model was trained in a different dataset and fine-tuned in our experiments, the transfer learning process may be less effective compared
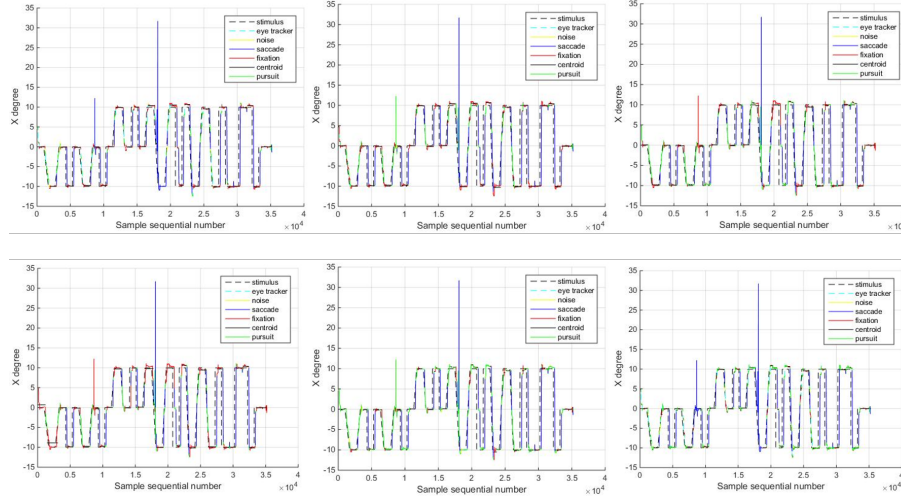
Fig. 1. Qualitative classification results on the "good quality" subject obtained by the proposed hierarchical HMM, I-VDT, I-VMP, IRF, I-BDT and 1D CNN, respectively. Best viewed in color.

to retraining the entire network model, which also reveals one of the limitations about those "data-driven" methods. Our proposed hierarchical HMM outperforms the other considered methods, especially the threshold-based methods, in most of the behavior scores and it is worth noting that the proposed method has behavior scores close to the manually evaluated results.

We also present the qualitative classification results in Fig. 1. Due to the reason that the I-VVT basically fails in this ternary eye movement classification task, we do not include the classification results of I-VVT in the figure. The qualitative results are consistent with the previous quantitative results. Notably, our proposed hierarchical HMM method succeeds in detecting an unexpected saccade while most of the other methods detect it either as smooth pursuit or fixation.

Feature Analysis. In our experiments, we propose a pre-processing step to analyze different features from raw data. We analyze velocity, position and acceleration features, and select velocity and position as two features used for the hierarchical HMM method. Velocity feature is used in the first HMM to filter saccades and also as the fine-tuning criterion in the second stage HMM, and the position features are used as the main feature to separate fixations and smooth pursuits in the second stage classification.

We present the visualization results of the data features on the subject that is manually evaluated as "good" in [13] using the K-Means cluster algorithm. The main objective of the pre-processing is to select appropriate features for further classifications using HMM. In the meanwhile, it also provides some insights for the reasons why the classic threshold-based methods may fail to achieve good
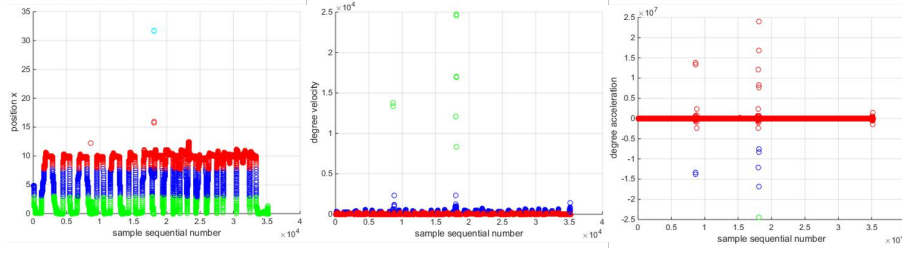
Fig. 2. Visualization of clustering results for position, velocity and acceleration features, respectively, on the data recording from the subject manually evaluated as "good" [13]. Different colors represent different clusters automatically detected by K-means. Best viewed in color.

Table 2. Comparison of behavior scores for the subject manually evaluated as "good" in the ablation study. Our proposed hierarchical strategy especially contributes to the classification of smooth pursuits.

Behavior scores	Ideal	Manual [13]	Ours	3-state HMM
SQnS	100%	96%	91%	91%
FQnS	84%	71%	74%	54%
PQnS	52%	39%	44%	25%
MixFix	7.1%	6%	5.5%	28%
FQlS	$0°$	$0.44°$	$0.3°$	$0.4°$
PQlS_P	$0°$	$3.15°$	$2.9°$	$3.4°$
PQlS_V	$0°/s$	$23°/s$	$25°/s$	$31°/s$

performance in the ternary eye movement classification task. As shown in Fig. 2, the actual eye tracking data usually contain a lot of noises. Take the velocity feature as an example, there exist some noisy data points with extremely large velocities and the difference of velocity among fixations, saccades and smooth pursuits are not very clear. Therefore, the statistical methods with predefined threshold values are more likely to fail in this case.

Ablation Study Analysis. We conduct an additional ablation experiment to demonstrate the necessity and effectiveness of the proposed hierarchical structure. Since the fixations, saccades and smooth pursuits in the ternary eye movement classification can be considered as three hidden states of an HMM, we use the velocity feature as the criterion and adopt a single-stage HMM with the three-hidden state to do the same task, in other words, we remove the second stage classification from the proposed method.

The quantitative and qualitative results are shown in Table 2 and Fig. 3, respectively. Both experimental results prove that the hierarchical structure contributes to better classification results. The qualitative figure shows that the hierarchical structure helps especially with the separation between fixations and smooth pursuits, which further validates the fact that the velocity feature is
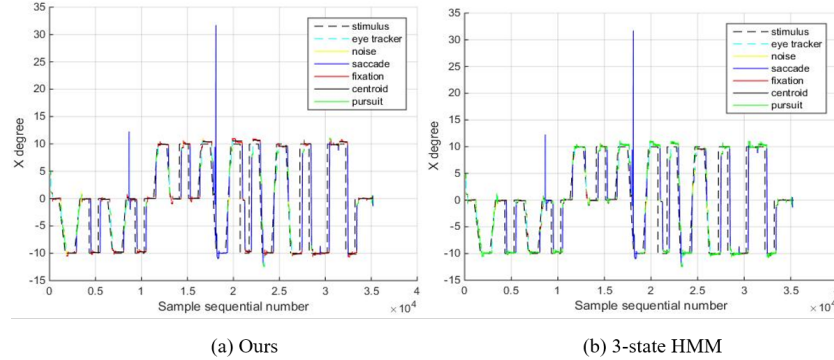
(a) Ours   (b) 3-state HMM

Fig. 3. Qualitative results for the ablation study. Our proposed hierarchical strategy helps with the classification of smooth pursuits, which is consistent with the quantitative results.

biased for fixations and smooth pursuits. Therefore, the analysis of different features and the usage of position features in the second stage of the proposed method are necessary and beneficial.

5   Conclusions

In this paper, we propose a hierarchical HMM algorithm to tackle the problem of ternary eye movement classification from the raw eye positional data. Different features from the data are considered to realize the multi-stage hierarchical classification. Experiments on multiple data records demonstrate the effectiveness and robustness of the proposed method. Possible future directions of this work could be incorporating the pre-processing step into the proposed method in an automated way, and making efforts to classify more categories of eye movements with a deeper hierarchical structure.

References

1. Bahill, A., Brockenbrough, A., Troost, B.: Variability and development of a normative data base for saccadic eye movements. Investigative ophthalmology & visual science **21**(1), 116–125 (1981)
2. Baum, L.E., Petrie, T., Soules, G., Weiss, N.: A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. The annals of mathematical statistics **41**(1), 164–171 (1970)
3. Boraston, Z., Blakemore, S.J.: The application of eye-tracking technology in the study of autism. The Journal of physiology **581**(3), 893–898 (2007)
4. Bruneau, D., Sasse, M.A., McCarthy, J.: The eyes never lie: the use of eyetracking data in hci research. ACM (2002)
5. Dewhurst, R., Nyström, M., Jarodzka, H., Foulsham, T., Johansson, R., Holmqvist, K.: It depends on how you look at it: Scanpath comparison in multiple dimensions with multimatch, a vector-based approach. Behavior research methods **44**(4), 1079–1100 (2012)
6. Duchowski, A.T., Medlin, E., Gramopadhye, A., Melloy, B., Nair, S.: Binocular eye tracking in vr for visual inspection training. In: Proceedings of the ACM symposium on Virtual reality software and technology. pp. 1–8 (2001)
7. Ebisawa, Y.: Improved video-based eye-gaze detection method. IEEE Transactions on instrumentation and measurement **47**(4), 948–955 (1998)
8. Ehmke, C., Wilson, S.: Identifying web usability problems from eye-tracking data. In: Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... but not as we know it-Volume 1. pp. 119–128. British Computer Society (2007)
9. Forney, G.D.: The viterbi algorithm. Proceedings of the IEEE **61**(3), 268–278 (1973)
10. Granka, L.A., Joachims, T., Gay, G.: Eye-tracking analysis of user behavior in www search. In: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 478–479. ACM (2004)
11. Hickson, S., Dufour, N., Sud, A., Kwatra, V., Essa, I.: Eyemotion: Classifying facial expressions in vr using eye-tracking cameras. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1626–1635. IEEE (2019)
12. Komogortsev, O.V., Gobert, D.V., Jayarathna, S., Koh, D.H., Gowda, S.M.: Standardization of automated analyses of oculomotor fixation and saccadic behaviors. IEEE Transactions on Biomedical Engineering **57**(11), 2635–2645 (2010)
13. Komogortsev, O.V., Karpov, A.: Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. Behavior research methods **45**(1), 203–215 (2013)
14. Li, Y., Fathi, A., Rehg, J.M.: Learning to predict gaze in egocentric video. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3216–3223 (2013)
15. Nyström, M., Holmqvist, K.: An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. Behavior research methods **42**(1), 188–204 (2010)
16. Poole, A., Ball, L.J.: Eye tracking in hci and usability research. In: Encyclopedia of human computer interaction, pp. 211–219. IGI Global (2006)
17. Purves, D.: Neuroscience. Scholarpedia **4**(8), 7204 (2009)
18. Salvucci, D.D., Goldberg, J.H.: Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 symposium on Eye tracking research & applications. pp. 71–78. ACM (2000)

19. Salvucci, D.D., R Anderson, J.: Tracing eye movement protocols with cognitive process models (1998)
20. Santini, T., Fuhl, W., Kübler, T., Kasneci, E.: Bayesian identification of fixations, saccades, and smooth pursuits. In: Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications. pp. 163–170 (2016)
21. Schiessl, M., Duda, S., Thölke, A., Fischer, R.: Eye tracking and its application in usability and media research. MMI-interaktiv Journal **6**, 41–50 (2003)
22. Startsev, M., Agtzidis, I., Dorr, M.: 1d cnn with blstm for automated classification of fixations, saccades, and smooth pursuits. Behavior Research Methods **51**(2), 556–572 (2019)
23. Zemblys, R., Niehorster, D.C., Holmqvist, K.: gazenet: End-to-end eye-movement event detection with deep neural networks. Behavior research methods **51**(2), 840–864 (2019)
24. Zemblys, R., Niehorster, D.C., Komogortsev, O., Holmqvist, K.: Using machine learning to detect events in eye-tracking data. Behavior research methods **50**(1), 160–181 (2018)