The Applicability of Probabilistic Methods to the Online Recognition of Fixations and Saccades in Dynamic Scenes

Enkelejda Kasneci*
Computer Engineering
University of Tübingen
Germany

Gjergji Kasneci
Hasso Plattner Institute
Germany

Thomas C. Kübler
Computer Engineering
University of Tübingen
Germany

Wolfgang Rosenstiel
Computer Engineering
University of Tübingen
Germany

Abstract

In many applications involving scanpath analysis, especially when dynamic scenes are viewed, consecutive fixations and saccades, have to be identified and extracted from raw eye-tracking data in an online fashion. Since probabilistic methods can adapt not only to the individual viewing behavior, but also to changes in the scene, they are best suited for such tasks.

In this paper we analyze the applicability of two types of mainstream probabilistic models to the identification of fixations and saccades in dynamic scenes: (1) Hidden Markov Models and (2) Bayesian Online Mixture Models. We analyze and compare the classification performance of the models on eye-tracking data collected during real-world driving experiments.

CR Categories: I.5.1 [Computing Methodologies]: Pattern Recognition—Models; I.6.4 [Computing Methodologies]: Simulation and Modeling—Model Validation and Analysis J.7 [Computer Applications]: Computers in Other Systems—Real time

Keywords: probabilistic, models, eye movements, online, classification, eye tracking, dynamic scene

1 Introduction

Visual perception involves six types of eye movements [Leigh and Zee 2006], among which fixations and saccades are the most studied. During a fixation, the eye is kept relatively stable on an area of interest (AOI), whereas saccades are fast eye movements enabling the fovea to fixate different areas of the scene [Privitera and Stark 2005]. Thus, the process of looking at a scene can be represented by a sequence of fixations and saccades, the so-called visual scanpath. Research on scanpath analysis and visual perception has largely benefited from the recent development of eye trackers. Today's eye-tracking systems allow a precise recording of eye movements at high sampling rates, thus enabling a detailed analysis of the viewing behavior.

While the detection of fixations and saccades in eye-tracking data that is aligned with information from the visual scene is reliably feasible by us humans, reliable automated clustering of eye movements is still challenging; even more so in dynamic scenarios. In many applications, e.g., human-computer gaze-based interaction, driving assistance systems, online adaptation of digital content based on gaze analysis, etc., the identification of fixations and saccades has to occur in an online fashion. There is a wide variety of methods for the online analysis of eye-tracking data and the recognition of fixations and saccades. However, only few of them are suited for online applicability to dynamic scenes. Obviously such methods have to quickly adapt not only to the individual viewing behavior but also to the changes occurring in the viewing scene. This small group of highly promising methods is based on probabilistic formalizations, e.g., as Markov Models [Salvucci and Goldberg 2000; Komogortsev et al. 2010; Sauter et al. 1991], Bayesian Mixture Models [Tafaj et al. 2012; Tafaj et al. 2013a], etc.

In this paper we first discuss some of the most advanced methods for the recognition of fixations and saccades from raw eye-tracking data and highlight the main advantages of probabilistic methods over other state-of-the-art methods. Then we take a closer look at two of the most generic probabilistic methods proposed in literature and empirically evaluate their online performance on hand-labeled, real-world data collected from driving experiments.

*e-mail: enkelejda.kasneci@uni-tuebingen.de

ETRA 2014, March 26 – 28, 2014, Safety Harbor, Florida, USA.

ACM 978-1-4503-2751-0/14/03

2 State of the art methods for the detection of fixations and saccades

Prior techniques for the automated recognition of different types of eye movements from eye-tracking data fall into two main categories: (i) threshold-based methods, where the distinction of fixations from saccades is based on dispersion, velocity, or acceleration thresholds, and (ii) probabilistic methods. These groups of methods will be briefly discussed in the following.

2.1 Threshold-based methods

Algorithms from this group distinguish between fixations and saccades based on the assumption that the distances, velocities, or accelerations occurring between subsequent fixations differ from those occurring between saccades. The goal then is to identify a threshold based on which saccades can be reliably distinguished from fixations.

When distance thresholds are used, fixation clusters are usually identified by searching for data points that are close enough to each other (i.e., below the established threshold) within a predefined time window [Holmqvist et al. 2011]. A representative of this group, is the *Dispersion Threshold Identification (I-DT)* algorithm [Salvucci and Goldberg 2000]. Other similar approaches differ mainly in the way the threshold is calculated [Blignaut 2009; Salvucci and Goldberg 2000; Shic et al. 2008].

Other algorithms in this realm are based on the computation of Minimum Spanning Trees (MST). In [Salvucci and Goldberg 2000] an MST is built on the eye-tracking points within a temporal window of predefined length. An edge (i.e., representing the distance between two points) is classified as a saccade if its length is significantly larger than the lengths of neighboring edges, which have been previously classified as distances between fixations. Yet other methods employ smart clustering algorithms, e.g., [Santella and DeCarlo 2004; Urruty et al. 2007], but have serious limitations with respect to their applicability to dynamic online scenarios, since in such scenarios the cluster properties for fixations and saccades show high variability.

Methods that are based on velocity or acceleration thresholds work similarly. A representative of this group is the *Velocity-Threshold Identification (I-VT)* algorithm, where a point is identified as a saccade point, if the implicit velocity along the distance from the previous data point to that point exceeds a predefined threshold. Otherwise the data point is assigned to a fixation cluster [Salvucci and Goldberg 2000].

Due to their simplicity, threshold-based algorithms have been implemented in several academic and commercial tools, e.g., Tobii[1], faceLab[2], SMI eGaze[3], Gazetracker[4], GazeAlyze [Berger et al. 2011], etc. Several recommendations for task specific settings of dispersion, velocity, or acceleration thresholds have also been made [Holmqvist et al. 2011], mostly considering eye-tracking data from viewing static images. Furthermore, in most cases, the above analysis software tools provide only offline analysis of eye movements.

In summary, the major drawback of threshold-based methods is that they rely on thresholds that have to be empirically adjusted to the individual viewing behavior, the viewing area, and the specific task. Each of these parameters can have significant influence on the classification result [Komogortsev et al. 2010; Komogortsev and Karpov 2013; Salvucci and Goldberg 2000]. For this reason and because of the fact that the viewing behavior is strongly physically- and physiologically-dependent, such methods are not reliable, especially when real-time analysis of eye-tracking data is needed.

2.2 Probabilistic methods

These methods build on soft decision rules, which are formalized as probabilities, e.g., the probability of a data point being a saccade given the previous observations. The probabilities – and thus the decisions – are adjusted to the observations.

HMM One of the most prominent probabilistic methods applied to the identification of fixations and saccades is the Hidden Markov Model (HMM). An HMM is a simple dynamic Bayesian network with variables representing values from a discrete state and observation space. The state of a variable represents the class of the current observation and is only dependent on the state (i.e., class of the previous observation). Because of this sequential nature, such models are a popular choice for the analysis of successively arising data points (i.e., observations). For the detection of fixations and saccades from eye data, HMMs have been used with velocity observations between successive data points, thus allowing the adaptation of the model to the physiological viewing behavior [Salvucci and Goldberg 2000]. In the model of [Salvucci and Goldberg 2000] (coined I-HMM), the two states used represent discretized velocity distributions over fixations and saccades. Transition probabilities between the states represent the probability of the current sample belonging to a fixation cluster or a saccade, given the previous state [Holmqvist et al. 2011]. Due to the above probabilistic representation, no thresholds are needed. The I-HMM is reported to outperform fixed-threshold methods, such as I-VT [Salvucci and Goldberg 2000]. In summary, the sequential, dynamic, and probabilistic nature of HMMs makes them an adequate choice for data arising in an online fashion and containing variability in its features.

Other similar approaches use more complex Markov models but have the same properties, e.g., Kalman Filters [Komogortsev et al. 2010; Sauter et al. 1991].

BMM Probabilistic mixture models, such as the Bayesian Mixture Model (BMM) presented in [Tafaj et al. 2012], build on the assumption that the observed data is generated from mixture of unknown density distributions. The goal is to estimate the parameters of these distributions based on observed data points and to derive the most probable distribution that might have generated a given data point.

The algorithm presented in [Tafaj et al. 2012] could distinguish between fixations and saccades in an online fashion, only by considering the Euclidean distances between subsequent data points. The underlying model is based on the assumption that distances between subsequent fixation points will in general be shorter than distances between subsequent saccade points; that is, distances between subsequent fixation points would be generated from a specific Gaussian distribution and those between subsequent saccade points from another. This intuition was modeled by a Bayesian Online Mixture Model. The benefit of the Bayesian formalization of the mixture model is that the parameters of the two distributions are updated and learned in an online fashion as more and more data is observed. For every new data point, the prior probabilities are replaced by the new estimates. For practical purposes this means that for every new user the algorithm needs a relatively small number of data points to adjust to that user and learn user- or scene-dependent parameters.

In summary, probabilistic methods come with three main advantages over threshold-based ones:

1. No fixed thresholds are needed, instead the parameters of the model (e.g., state transition probabilities, label emission probabilities and other parameters) are learned from labeled data.
2. Both HMMs and BMMs can adapt to the individual (i.e., physiological) viewing behavior of a subject and to the specific task.
3. Given the dynamic nature of the underlying models, the methods are naturally suited for data arising in an online fashion, such as eye-tracking data.

3 Experimental evaluation

In this section, we compare the I-HMM with the BMM with respect to their applicability to online scenarios, based on real-world, hand-labeled eye-tracking data set collected from driving experiments.

3.1 Implementation

We implemented a two-state HMM according to the description of the I-HMM in [Salvucci and Goldberg 2000]. However, the observed sequences for the I-HMM were velocities between the eye-tracking data points, whereas in the HMM version that we have implemented, the sequences consist of distances between successive data points. Based on training data (i.e., manually labeled data points) such distance observations can be mapped to a discrete set of observations, which in our context correspond to the IDs of two Gaussian distributions, i.e., one distribution of distances between saccades and a distribution of distances between fixations. These distributions, the emission probabilities of their IDs, as well as the transition probabilities between the HMM states are learned from labeled data, by computing the corresponding maximum likelihood estimations. As with the I-HMM [Salvucci and Goldberg 2000], the two states of the HMM represent a distribution over fixations and saccades. A Viterbi-based, forward-backward algorithm [Forney Jr. 1973] was implemented to compute the most probable state sequence of the HMM for a given observation sequence.

The BMM was implemented as a Bayesian mixture of two Gaussian distributions, one representing distances between consecutive saccades and another one representing distances between consecutive fixations, as described in [Tafaj et al. 2012; Kasneci 2013]. We used Infer.NET[5] to implement the Bayesian model and variational message passing to perform the inference, i.e., finding the

[1] http://www.tobii.com/en/eye-tracking-research
[2] http://www.seeingmachines.com/product/facelab
[3] http://www.smivision.com/en/gaze-and-eye-tracking-systems
[4] http://www.eyetellect.com/gazetracker
[5] http://research.microsoft.com/en-us/um/cambridge/projects/infernet/

most likely distribution to have generated the observed data points. The source code is also available online[6].

Both models were applied post-experimentally, but in real-time on sequentially arising raw eye-tracking data points from nine different data sets that will be described in the following subsection.

3.2 Data sets

To evaluate the performance of the HMM and BMM, we employed nine real-world, eye-tracking data sets from nine different subjects who took part in our driving experiments [Kasneci et al. 2014]. Each data set consisted of 750 data points recorded while driving, at a sampling rate of 25Hz by a mobile Dikablis eye tracker. The data was analyzed frame-wise by two PhD students. An eye-tracking data point was thereby labeled as belonging to a saccade or fixation only if both of the judges agreed. Blinks and disagreements were excluded from the data. Note that this annotation task is very laborious, as the eye-tracking data has to be labeled manually with respect to the information on the scene. Note that each of the data sets corresponds to a driving sequence of 30 seconds, resulting in a total of 4.5 minutes.

This annotation process led to the data sets presented in Table 1.

	Training data		Test data	
Data set	**Saccade**	**Fixation**	**Saccade**	**Fixation**
Data set 1	33	267	50	398
Data set 2	41	259	51	166
Data set 3	29	271	89	359
Data set 4	41	259	94	258
Data set 5	28	272	53	393
Data set 6	88	212	88	284
Data set 7	67	233	74	366
Data set 8	8	292	9	321
Data set 9	17	283	50	364

Table 1: *Hand-labeled data sets from driving sessions of nine subjects. The first 300 data points in each data set were employed for training, the remaining ones for testing the models.*

Both HMM and BMM were trained on the first 300 eye-tracking points of the above data sets. The HMM derives from the training data the distance distributions (i.e., for distances between consecutive saccades and consecutive fixations) as well as transition and emission probabilities. In contrast, the BMM, updates the learned parameters in an online fashion as new data points are observed. Once the parameters were learned for both models, their prediction quality was tested on each data set.

3.3 Evaluation results

Tables 2 and 3 show a detailed analysis of the quality of both algorithms with respect to the detection of saccades and fixations, respectively. Both tables show quality results in terms of the following measures: *Precision* ($\frac{TP}{TP+FP}$), *Recall* ($\frac{TP}{TP+FN}$), *F1-measure* ($\frac{2 \cdot Precision \cdot Recall}{Precision+Recall}$), which represents the harmonic mean of precision and recall, and *Miss-Classification-Rate (MCR)* ($\frac{FP+FN}{TP+FP+TN+FN}$).

As shown in Table 2, the BMM clearly outperformed the HMM with respect to the detection of saccades, which, in this context, is the more critical class, since a correct detection of saccades implies a correct separation of fixation clusters. Note that in general the proportion of saccade points is much smaller than that of fixation points. For a model such as the HMM, which aims at maximizing the joint probability of a sequence of states and corresponding observations, it is safer to focus on the most probable states and observations; these are fixations and the corresponding distance means. These findings are also in line with the findings presented in [Komogortsev et al. 2010].

In contrast, for the detection of saccades, the BMM achieved an astounding precision between 97% and 100%, Table 2. For most of the data sets it also showed higher recall values than the HMM. This implies that, most of the time, successive fixation clusters are correctly separated by the BMM.

	Model	Precision	Recall	F1	MCR
Data set 1	**HMM**	0.667	0.960	0.787	0.058
	BMM	1.000	0.926	0.962	0.009
Data set 2	**HMM**	0.800	0.941	0.865	0.069
	BMM	1.000	0.927	0.962	0.018
Data set 3	**HMM**	0.742	0.857	0.795	0.076
	BMM	1.000	0.937	0.967	0.013
Data set 4	**HMM**	0.737	0.926	0.821	0.108
	BMM	1.000	0.979	0.989	0.006
Data set 5	**HMM**	0.943	0.758	0.840	0.043
	BMM	1.000	0.964	0.981	0.004
Data set 6	**HMM**	0.625	0.859	0.724	0.113
	BMM	1.000	0.957	0.978	0.011
Data set 7	**HMM**	0.757	0.949	0.842	0.048
	BMM	1.000	0.902	0.949	0.018
Data set 8	**HMM**	1.000	0.375	0.545	0.015
	BMM	1.000	1.000	1.000	0.000
Data set 9	**HMM**	0.698	0.600	0.645	0.080
	BMM	0.971	1.000	0.985	0.003

Table 2: *Quality comparison of the HMM and the BMM for the class of saccade points.*

Indeed, a much better performance of the HMM is shown in Table 3, where the precision with respect to the detection of fixations varies between 96% and 99%. The precision achieved by the BMM is also very high, varying between 97% and 100%. However, the recall of the HMM with respect to the detection of fixation points is remarkably lower than that of the BMM, because the HMM does not manage to adapt well enough to varying distances between fixation points. This is different for the BMM, which can adapt to varying distances in an online fashion. Considering the MCR, which is very low for both models, we found a superior performance of the BMM with values smaller than 2%. In summary, the HMM was outperformed by the BMM with respect to all measures.

These results highlight the superior performance of the online Bayesian mixture model in comparison to a HMM. Especially, on the difficult task of reliably detecting saccades, which is crucial for the correct separation of fixation clusters, the Bayesian model achieves highly satisfiable precision, and recall values.

Beyond the context of dynamic scenarios, the BMM can be integrated into vision research tools (e.g., Vishnoo [Tafaj et al. 2011]) to analyze the viewing behavior during visual search tasks presented on a screen. Furthermore, the BMM can be used in the context of medical testing, e.g., for advanced visual field testing involving online analysis of fixations (e.g., EFOV [Tafaj et al. 2013b]). In

[6] https://www.ti.uni-tuebingen.de/Analysis-of-Eye-Tracking-Data.eyetracking.0.html

summary, whenever fixations and saccades have to be detected in an online fashion, the BMM is a highly reliable choice.

	Model	Precision	Recall	F1	MCR
Data set 1	**HMM**	0.995	0.940	0.966	0.058
	BMM	0.990	1.000	0.995	0.009
Data set 2	**HMM**	0.981	0.928	0.954	0.069
	BMM	0.976	1.000	0.988	0.018
Data set 3	**HMM**	0.969	0.938	0.953	0.076
	BMM	0.983	1.000	0.992	0.013
Data set 4	**HMM**	0.970	0.880	0.923	0.108
	BMM	0.992	1.000	0.996	0.006
Data set 5	**HMM**	0.959	0.992	0.975	0.043
	BMM	0.995	1.000	0.997	0.004
Data set 6	**HMM**	0.968	0.893	0.929	0.113
	BMM	0.986	1.000	0.993	0.011
Data set 7	**HMM**	0.992	0.953	0.972	0.048
	BMM	0.978	1.000	0.989	0.018
Data set 8	**HMM**	0.985	1.000	0.992	0.015
	BMM	1.000	1.000	1.000	0.000
Data set 9	**HMM**	0.946	0.964	0.955	0.080
	BMM	1.000	0.997	0.998	0.003

Table 3: *Quality comparison of the HMM and the BMM for the class of fixation points.*

4 Conclusion

In this paper, we have discussed viable approaches to the challenging task of online detection of saccades and fixations in eye-tracking data from dynamic scenes. While it is known that probabilistic methods are more adequate than threshold-based ones for these kinds of tasks, we showed that, when applied to online and dynamic data, there may be considerable performance differences between probabilistic methods. The experimental evaluation on raw eye-tracking data, collected from on-road and simulated driving sessions has demonstrated a superior classification performance of the Bayesian Mixture Model over the Hidden Markov Model. The main reason for the better performance of the Bayesian Mixture Model is its flexibility with respect to the continuous update of the parameters and thus its ability to adapt to the viewing behavior and changes in the scene.

References

BERGER, C., WINKELS, M., LISCHKE, A., AND HÖPPNER, J. 2011. GazeAlyze: a MATLAB toolbox for the analysis of eye movement data. *Behavioral Research Methods 44*, 2, 1–16.

BLIGNAUT, P. 2009. Fixation identification: The optimum threshold for a dispersion algorithm. *Attention, Perception, & Psychophysics 71*, 4, 881–895.

FORNEY JR., G. D. 1973. The viterbi algorithm. *Proceedings of the IEEE 61*, 3, 268–278.

HOLMQVIST, K., NYSTRÖM, M., ANDERSSON, R., DEWHURST, R., JARODZKA, H., AND VAN DE WEIJER, J. 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

KASNECI, E., SIPPEL, K., AEHLING, K., HEISTER, M., ROSENSTIEL, W., SCHIEFER, U., AND PAPAGEORGIOU, E. 2014. Driving with Binocular Visual Field Loss? A Study on a Supervised On-road Parcours with Simultaneous Eye and Head Tracking. *Plos One*. doi: 10.1371/journal.pone.0087470.

KASNECI, E. 2013. *Towards the Automated Recognition of Assistance Need for Drivers with Impaired Visual Field*. PhD thesis, University of Tübingen, Wilhelmstr. 32, 72074 Tübingen.

KOMOGORTSEV, O. V., AND KARPOV, A. 2013. Automated classification and scoring of smooth pursuit eye movements in the presence of fixations and saccades. *Behavior Research Methods 45*, 1, 203–215.

KOMOGORTSEV, O. V., JAYARATHNA, S., KOH, D. H., AND GOWDA, S. M. 2010. Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '10, 65–68.

LEIGH, R. J., AND ZEE, D. S. 2006. *The neurology of eye movements*. Oxford University Press.

PRIVITERA, C. M., AND STARK, L. W. 2005. Scanpath theory, attention, and image processing algorithms for predicting human eye fixations. In *Neurobiology of Attention*, L. Itti, G. Rees, and J. Tsotsos, Eds., 269–299.

SALVUCCI, D., AND GOLDBERG, J. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '00, 71–78.

SANTELLA, A., AND DECARLO, D. 2004. Robust clustering of eye movement recordings for quantification of visual interest. In *Proceedings of the 2004 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '04, 27–34.

SAUTER, D., MARTIN, B. J., DI RENZO, N., AND VOMSCHEID, C. 1991. Analysis of eye tracking movements using innovations generated by a Kalman filter. *Medical and biological Engineering and Computing 29*, 1, 63–69.

SHIC, F., SCASSELLATI, B., AND CHAWARSKA, K. 2008. The incomplete fixation measure. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*, ACM, New York, NY, USA, ETRA '08, 111–114.

TAFAJ, E., KÜBLER, T., PETER, J., SCHIEFER, U., BOGDAN, M., AND ROSENSTIEL, W. 2011. Vishnoo - an open-source software for vision research. In *Proceedings of the* 24^{th} *IEEE International Symposium on Computer-Based Medical Systems*, IEEE, CBMS' 11, 1–6.

TAFAJ, E., KASNECI, G., ROSENSTIEL, W., AND BOGDAN, M. 2012. Bayesian online clustering of eye movement data. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM, New York, NY, USA, ETRA '12, 285–288.

TAFAJ, E., KÜBLER, T., KASNECI, G., ROSENSTIEL, W., AND BOGDAN, M. 2013. Online classification of eye tracking data for automated analysis of traffic hazard perception. In *Artificial Neural Networks and Machine Learning ICANN 2013*, Springer Berlin Heidelberg, vol. 8131, 442–450.

TAFAJ, E., HEMPEL, S., HEISTER, M., AEHLING, K., SCHAEFFEL, F., DIETZSCH, J., ROSENSTIEL, W., AND SCHIEFER, U. 2013. A New Method for Assessing the Exploratory Field of View (EFOV). In *HEALTHINF 2012*, SciTePress, D. Stacey, J. Sol-Casals, A. L. N. Fred, and H. Gamboa, Eds., 5–11.

URRUTY, T., LEW, S., IHADADDENE, N., AND SIMOVICI, D. A. 2007. Detecting eye fixations by projection clustering. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP) 3*, 4, 1–20.