Online Recognition of Fixations, Saccades, and Smooth Pursuits for Automated Analysis of Traffic Hazard Perception

Enkelejda Kasneci, Gjergji Kasneci, Thomas C. Kübler, and Wolfgang Rosenstiel

Abstract. Complex and hazardous driving situations often arise with the delayed perception of traffic objects. To automatically detect whether such objects have been perceived by the driver, there is a need for techniques that can reliably recognize whether the driver's eyes have fixated or are pursuing the hazardous object. A prerequisite for such techniques is the reliable recognition of fixations, saccades, and smooth pursuits from raw eye tracking data. This chapter addresses the challenge of analyzing the driver's visual behavior in an adaptive and online fashion to automatically distinguish between fixation clusters, saccades, and smooth pursuits.

1 Introduction

Driving is a complex task requiring proper visual functioning. According to Nagayama [Nag78], more than 50% of collisions in road traffic occur due to missed or delayed hazard perception [VRK$^+$02]. Therefore, numerous studies over the last two decades have investigated eye movements of drivers to identify deficits in visual search patterns or types of hazardous situations that may cause accidents. According to the scanpath theory by Noton and Stark [NS71], a top-down internal cognitive model of what we see drives our eyes efficiently over a scene [PS05] involving six types of eye movements: *fixations, saccades, smooth pursuits, optokinetic reflex, vestibulo-ocular reflex*, and *vergence* [LZ06]. Among these, fixations, saccades, and smooth pursuits are the most studied in driving scenarios. During a

Enkelejda Kasneci · Thomas C. Kübler · Wolfgang Rosenstiel
Department of Computer Engineering, University of Tübingen, Germany
e-mail: {enkelejda.kasneci,thomas.kuebler,
wolfgang.rosenstiel}@uni-tuebingen.de

Gjergji Kasneci
Hasso Plattner Institute, Germany
e-mail: gjergji.kasneci@hpi.uni-potsdam.de

P. Koprinkova-Hristova et al. (eds.), *Artificial Neural Networks*,
Springer Series in Bio-/Neuroinformatics 4, DOI: 10.1007/978-3-319-09903-3_20

fixation the eye is kept relatively stable on an area of interest (AOI), whereas saccades correspond to rapid eye movements enabling the retinal part of sharpest vision (fovea) to fixate different areas of the scene [PS05]. Smooth pursuits occur whenever the eye follows a moving target [Duc07]. Thus, eye movements during driving result in a sequence of fixations, smooth pursuits, and saccades. Such a sequence is also known as *visual scanpath*.

The visual search strategy of a driver seems to be crucial for driving safety. Several studies have reported changes in the viewing patterns of drivers with increasing driving experience [CU98, CUR02, MS99, PHD+05]. Chapman et al. [CUR02] have reported that differences between viewing patterns of novices and experienced drivers are particularly noticeable in demanding or hazardous situations. These changes, however, do not seem to be related to the cognitive resources but rather to the envisioned model of what is likely to happen during the task [UCBC02]. A further issue addressed by several studies concerns changes in viewing patterns with age. While some studies could not find any age-related decline in the subject's visual search behavior in hazardous situations [UPW+05, HDBK+13], others have reported a decline in search efficiency with age [MS99, HSCG01]. At least there is evidence of a significant increase of response times with age [HMM+08].

Based on the above reports, several training techniques have been proposed to improve the visual scanning of drivers – with limited success [KCC12]. The main problem is that viewing patterns are highly individual and vary with the task and scene. Therefore, we hypothesize that the most effective way of hazard avoidance in driving scenarios would be to provide automated and adaptive means for continuously monitoring and analyzing the driver's visual behavior in alignment with entities on the scene; the driver should be warned well in time about upcoming hazards and, in cases where the driver's reaction time would not suffice to avoid an accident, the system (e.g., the automated breaking system) should completely take over to avoid the worst.

Indeed, today's driving assistance systems build on numerous sensors to provide assistance for specific tasks, such as automatic parking, lane keeping, intelligent speed control, emergency braking assistance, and many more. However, many of these tasks are independent of the driver's abilities, let alone her visual deficits. Moreover, the corresponding systems are geared towards reliable, deterministic performance; they do neither adapt to new traffic situations and nor to individual driving capabilities. For example, emergency braking assistance systems are relatively crude in nature, as they are typically applied as the only mean to avoid an accident. In contrast, we are interested in more fine-grained methods, which take the driver's visual search behavior into account to identify overlooked hazardous objects in real-time, possibly without distracting or patronizing the driver.

Going beyond driving assistance systems, efforts toward autonomous driving [Mar10, UAB+08] have come a long way since their beginnings [Pom89]. For a safe navigation, modern self-driving cars use numerous sensors to analyze position, speed, traffic situation, road conditions, etc. The goal of related projects is to take away the burden of driving from human drivers and shift it to the inbuilt autonomous car system. However, despite the reported driving safety of such systems [Mar10], there are still many unanswered questions, especially concerning the legislation of driving, e.g., who is responsible in case of accidents, which insurance company covers the costs, etc. Our approach aims at improving the driving performance and leaves the burden of driving, and thus the whole responsibility, to the human driver. The overarching vision of our work is the synergistic combination of methods that analyze the visual search behavior of the driver as a first step towards the development of driving assistance systems that smoothly guide the driver's visual attention towards potential traffic hazards.

The most important cornerstone for the implementation of such systems is the development of methods for the analysis of the driver's visual behavior in an online fashion. This problem can be reduced to the reliable detection of and distinction between fixation clusters, saccades, and smooth pursuits from eye-tracking data. Most prior research on the detection of fixations and saccades has primarily focused on the offline detection of these types of eye movements. Various approaches such as position, velocity or acceleration based algorithms, methods based on minimum spanning trees, Hidden Markov Models or Kalman filters have been proposed [BWLH12, CNTN08, Git02, KJKG10, MSP08, PS00, SG00, SD04, TGB03, Woo02]. Yet, these approaches show two main drawbacks: (i) they either require several clustering parameters as input, which makes them inadequate for online usage or (ii) they show poor performance in the detection of fixations and saccades in dynamic scenes. A further challenge arises when smooth pursuits have to be distinguished from saccades and fixations. Although new methods have been proposed [KK13, VBG12], their applicability to the detection of smooth pursuits in dynamic scenes is unclear.

In order to identify whether a hazard was perceived by the driver, the driver's visual scanpath has to be analyzed in real time while considering all entities that appear on the visual scene. If a relatively stable target was perceived, we would expect the driver's eye movements to be focused on that target, thus yielding a cluster of fixation points. Any algorithm for clustering such fixation points needs to work in an online fashion and be unparameterized with respect to the number of clusters, as new entities may appear on the scene. Note that the system has to know the driver's AOIs at any point in time. Furthermore, as the viewing behavior differs from person to person, an adaptive algorithm is needed.

In the following, we present a work-flow for the online analysis of hazard perception based on an adaptive online algorithm for the identification of and distinction between fixations, saccades and smooth pursuits.

2 Visual Perception and Eye Movements

The human eye is a foveated system. This means that optimal visual perception (i.e., at highest spatial resolution) is only possible at a small, central area of the retina, which is known as the fovea. The spatial resolution falls by an order of magnitude within a few degrees from the fovea. Thus, when we want to look at an object, we will move our eyes until the image of the object falls on the fovea of each eye [Cor12]. Visual perception would therefore not be possible without eye movements. Although we are mostly unaware of them, when viewing a scene, our eyes are constantly moving to enable the fovea to fixate different parts of the scene. These foveal fixations are also known as *Areas of Interest*, AOIs for short. Although it is possible to deploy attention without an accompanying fixation, under natural viewing conditions this is quite rare. Thus, eye movements are assumed to be preceded by a shift in visual attention to a subsequent target. However, the whole process of visual perception involves not only the sensory input mechanisms, but also memory, attention, cognition, and decision making [HB05, Lan09, LT09].

The question, what drives our eyes over a scene, has been raised in early works of Buswell [Bus35] and Yarbus [Yar67] and, since then, has been in the scope of research in different areas. Especially in recent years, since eye-tracking devices have improved and become broadly available, the number of investigations on eye movements has increased. Several approaches modeling the process of visual attention have been proposed and follow mainly two basic paradigms: the *top-down* and *bottom-up mechanism*. The scanpath theory by Noton and Stark [NS71] suggests that an internal cognitive model of what we see not only controls our vision, but also drives the sequences of rapid eye movements and fixations efficiently over a scene [PS05], e.g., when we are asked to identify a specific object in a scene. This is called the top-down mechanism.

The bottom-up mechanism is based on the idea that an object might attract our attention due to particular features, e.g., a person wearing a red shirt among others wearing white. In this example, the red color is the salient feature. Computer-based models of visual attention have mainly focused on the modeling of bottom-up visual attention by automatically identifying possible fixation targets based on image properties [IK00, IKN98]. The detailed description of these models is beyond the scope of this work. A review of visual attention models can be found in [LT09, SBG11, THLB11]. Although intuitively comprehensible, visual salience as determined by such models can not predict saccade sequences any better than random models [HBCM07].

Despite extensive research in this area, so far, it has not been possible to predict the sequence of fixations of an observer looking at an arbitrary scene [SBG11]. The difficulty lies in the complexity of visual processing, for which eye movements are essential. Six types of eye movements are involved in visual processing of a scene: *fixations*, *saccades*, *smooth pursuits*, *optokinetic reflex*, *vestibulo-ocular reflex*, and *vergence* [LZ06]. In this work we focus on fixations, saccades, and smooth pursuits. Hence, these types of eye movements will be briefly described in the following subsections. A detailed description of the physiology and characteristics of eye movements can be found in [Duc07, LT09].

2.1 Fixation

During a fixation the eye is kept relatively stable on an AOI. It can be assumed that during a fixation, the AOI imaged on the fovea is being visually attended to by the viewer. Fixations usually have a duration of about $200-300ms$, although much longer fixations are possible. The duration of fixations varies depending on the visual search task. Furthermore, fixation durations show inter- and intra-subject variability. Therefore, the duration of fixations is an important research topic in itself, as it relates to the visual information to which the observer is attending as well as to her cognitive state [LT09].

The general assumption while driving is that the visual perception of scene features, such as signs, pedestrians, and obstacles, requires foveated vision (i.e., explicit fixation of the object of interest) [FZ09]. Although peripheral vision is considered as sufficient for some subtasks, such as keeping the vehicle centered in the lane [SNP96], in [MS04] was reported that peripheral vision is insufficient for the detection of traffic hazards. This means, that to have seen a traffic hazard, the driver must have fixated it.

2.2 Saccade

A saccade represents a rapid eye movement to reposition the retinal part of sharpest vision (fovea) to fixate different areas of a scene [PS05]. A saccade occurs at maximum frequency of about $4s^{-1}$ (e.g., during reading) and maximum velocity of approximately $500°/s$ [LT09]. Saccades are performed by both eyes conjugately. As introduced above, a saccade can be triggered in a bottom-up manner, e.g., by a suddenly moving target or a new stimulus, or in a top-down way. Saccade duration varies between $10ms$ and $100ms$ [Duc07]. The reaction time for externally driven saccades is usually $150-200ms$. Furthermore, saccade latency increases with increasing eccentricity [LT09].

2.3 Smooth Pursuit

A smooth pursuit occurs whenever the eye follows a moving target [Duc07]. Similarly to saccades, smooth pursuits are conjugate eye movements that are performed at lower velocities of approximately $15°/s$ [LT09]. They are elicited by moving targets and cannot be executed voluntarily without a target that is visually pursued.

3 Online Recognition of Fixations, Saccades, and Smooth Pursuits from Eye-Tracking Data

The reliable detection of and distinction between the above types of eye movements in an online fashion is a crucial step towards the automated identification of objects that might be overlooked by the driver, i.e., potential traffic hazards. While performing such a detection on eye-tracking data in alignment with information from the visual scene is reliably feasible by us humans, reliable automated clustering of eye movements is still challenging; even more so, when the visual scene changes continuously, such as in driving scenarios.

In the scope of this section will be an online, adaptive algorithm for the reliable detection of the above eye movement types. After discussing related work on the classification of eye movement data, we present a two-step approach to the online detection of the type of eye movements from eye-tracking data.

3.1 State of the Art Methods

Since the first works on classification of eye movements [McC81, Wid84], several approaches have been proposed that fall into three main categories: (i) dispersion-based methods, (ii) velocity- and acceleration-based methods, and (iii) probabilistic methods based on statistical Markov models. All groups of algorithms will be briefly discussed in the following subsections. A detailed review can be found in [HNA$^+$11].

3.1.1 Dispersion-Based Methods

Dispersion-based algorithms distinguish between different types of eye movements (mostly between fixations and saccades) based on the distance between subsequent eye position points. Usually, dispersion-based algorithms aim at detecting fixation clusters by identifying data points that are close to each other within a predefined time window, whereas all other data points are not considered [HNA$^+$11]. A method that is representative of this group, is the *Dispersion Threshold Identification (I-DT)* algorithm [SG00], where the separation of fixation clusters from saccades is based on the dispersion of consecutive data points within a temporal window. In I-DT, the dispersion D is defined as the sum of the maximum and minimum differences between the x and y coordinates of the points within the temporal window, i.e., $D = [max(x) - min(x)] + [max(y) - min(y)]$. If D is below a predefined threshold, the data points within the window belong to a fixation. Otherwise, a saccade is found. Similar approaches that use a temporal threshold as in I-DT, but differ in the way the dispersion is calculated have also been presented in [Bli09, HNA$^+$11, SG00, SSC08].

Another prominent algorithm from this group is the *Identification of eye movements based on Minimum Spanning Trees (I-MST)* [HNA$^+$11, SG00]. As presented in [SG00], this algorithm first builds a minimum spanning tree (MST) on a predefined number of eye position points that fall within a temporal window. Then, from each node v of the MST, all other MST nodes are visited in depth-first search up to a predefined depth threshold. The mean and the variance of the length of the edges that were visited during the depth-first search are assigned as meta information to v. Based on the meta information of the end nodes of an edge, a decision can be taken concerning the type of eye movement it represents, i.e., a saccade or fixation [SG00].

Other dispersion-based approaches are based on clustering algorithms. For example, in [SD04] a mean shift clustering algorithm for the detection of fixation clusters was used. In another approach [ULIS07], fixation clusters have been detected based on projection clustering. Typical clustering algorithms, such as the well-known and widely used k-means algorithm, are only applicable to the problem of identifying fixations clusters, when the number of fixation clusters is known a priori. Hence, such algorithms cannot be applied to the analysis of eye movements when viewing dynamic scenes, e.g., such as those occurring while driving, where the number of resulting clusters is not known.

Other approaches in this realm have divided the viewing area into a regular grid and recorded the time spent inside each square [SG00]. While such approaches are well-suited for static scenes, e.g., reading a page, they are not applicable to dynamic scenarios, where no a priori information about the viewing area is given. The same holds for techniques that provide visualizations of the areas of interest, i.e., fixation clusters, by adapting the original image [Woo02].

Over the last years, dispersion-based approaches have been implemented in both academic and commercial tools [HNA$^+$11], e.g., faceLab [See], SMI BeGaze [Sen], Gazetracker [Eye], etc. Although they offer several useful features, their main drawback is that they come as black-box solutions and can hardly be integrated in self-designed applications. Moreover, in most cases, commercial analysis software provides only offline analysis of eye movements. Academic tools such as the MATLAB-toolbox GazeAlyze [BWLH12] based on ILab [Git02], or Astef [CNTN08] can be easily integrated in self-designed applications but, unfortunately, only for the offline analysis. Another major drawback of dispersion-based methods is that they often rely on thresholds (e.g., length of the temporal window, dispersion threshold, etc.) that have to be empirically adjusted to the individual viewing behavior, the viewing area, and the specific task, thus being inadequate for the task of adaptive, online scanpath analysis [HNA$^+$11].

3.1.2 Velocity- and Acceleration-Based Methods

Velocity- and acceleration-based algorithms distinguish between different types of eye movements based on velocities or accelerations between subsequent eye data points [HNA$^+$11]. An approach that is representative of this group is the *Velocity-Threshold Identification (I-VT)* algorithm. In I-VT, 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 (e.g., $>$ 300 deg/sec). Otherwise the data point is assigned to a fixation cluster [SG00]. Approaches based on acceleration thresholds work similarly. Due to their simplicity, algorithms of this group have been implemented in several commercial software packages, e.g., Tobii [Tob], SMI [Sen], EyeLink [SR]. Several recommendations for task specific settings of velocity thresholds have also been made [HNA$^+$11].

In general, this group of algorithms is best suited for the analysis of eye data tracked at high sampling frequency ($>$ 200 Hz) [HNA$^+$11]. However, as with the dispersion-based methods, a major drawback of the I-VT algorithm and methods related to it is that the applied threshold values need to be empirically adjusted to the eye data at hand. For this reason and because of the fact that velocity profiles are strongly physically- and physiologically- dependent, such methods are not reliable, especially when real-time analysis of eye-tracking data is needed. An additional issue concerns the detection of smooth pursuits. Most velocity-based methods have primarily focused on the separation of fixation clusters from saccades without considering smooth pursuits as an additional class. In [Itt05] a velocity-based approach, which classifies fixation and smooth pursuit clusters into a general "intake" category has been presented. Although saccades and fixations could be separated based on threshold values, smooth pursuits could not be distinguished from fixation clusters. Commercial implementations, such as the Tobii Fixation Algorithms [Tob] and the EyeLink parser [SR], work similarly [HNA+11]. Two-step approaches that use two velocity thresholds to first detect saccades and then separate smooth pursuits from fixations, were presented in [Fer00] and [KK13]. Offline Eigenvector analysis on data points and velocities within a temporal window has been proposed as an approach to distinguish fixations from smooth pursuits in [BBM+09]. Although reliable, the static nature of such offline methods makes them inadequate for the application to dynamically changing scenes. In [KK13], a combination of velocity and dispersion thresholds was used with the I-VT algorithm (coined I-VDT in [KK13]) to classify saccades, smooth pursuits, and fixations. As above, following a two-step approach, first saccades are separated from the other two types of eye movements based on a velocity threshold; then, a dispersion threshold is used to separate fixations from smooth pursuits [KK13]. As with all the presented threshold-based methods, the thresholds of I-VDT have to be chosen carefully, based on thorough data analysis and the specific task.

In summary, most of the dispersion- and velocity-based approaches are based on a considerable number of input parameters that can have significant influence on the classification result. Although several recommendations regarding thresholds for specific tasks have been made, they mostly consider eye-tracking data from viewing behavior on static images.

3.1.3 Probabilistic Methods

The most prominent methods that are representative of this realm are Hidden Markov Models (HMM). Such models are simple dynamic Bayesian networks with variables representing values from a discrete state and observation space. Because of their 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 [SG00]. In the model of [SG00] (coined I-HMM), the two states used represent velocity distributions over fixations and saccades. Transition probabilities between the states represent the probability of the next sample belonging to a fixation cluster or a saccade, given the current state [HNA+11]. Due to the probabilistic representation of velocities (i.e., no thresholds are needed), the I-HMM is reported to perform better than fixed-threshold methods, such as I-VT [SG00]. The dynamic and probabilistic nature of HMMs makes them an adequate choice for sequential data arising in an online fashion and containing variability in its features.

Another similar probabilistic approach was presented in [KJKG10, SMDRV91] and was based on a Kalman-Filter that models and predicts the velocity of eye movements based on previous eye data points. The proposed model could distinguish saccades from fixations.

In summary, with respect to their application to the online analysis of eye data generated in the context of driving, many of the related approaches reviewed in this section suffer from one or more of the following problems: (i) they require several static input parameters, which makes them inadequate for online usage, (ii) they do not adapt to changing scene information or to the subject's viewing behavior, or (iii) they do not successfully generalize to the detection of smooth pursuits in addition to fixations and saccades [KK13].

In contrast, our method, which can be assigned to the family of probabilistic models, performs reliably on the task of online classification of fixations, saccades, and smooth pursuits.

3.2 Method Overview

The method for classifying eye-tracking data consists of two steps: In the first step, a Bayesian online mixture model is employed to distinguish saccades from data points that might represent fixations and smooth pursuits. In a second step, a fine-grained Principal Component Analysis (PCA) is conducted to distinguish fixations from smooth pursuits. Both steps are presented in the following subsections.

3.2.1 Bayesian Mixture Model for the Online Recognition of Fixations and Saccades

In [TKRB12], we presented an effective online clustering algorithm, that could distinguish between fixations and saccades by considering only the Euclidean distance between subsequent data points recorded by the eye tracker. The underlying model was based on the intuition 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. Imagine a temporally ordered sequence of two-dimensional data points, $\mathbf{S} = \{s_i \mid 1 \leq i \leq T\}$, e.g., recorded by an eye tracker and representing the visual scanpath of an observer over time. In such a representation, a dense region of successive points (i.e., successive points that are close to each-other in terms of the Euclidean distance) might reflect an AOI (or, more specifically, an object that attracts the observer's attention). According to [NH10], a fixation location is manifested in eye-tracking recordings as a cloud of points that are normally distributed around the center of the object of interest. Assuming such a normal distribution of fixation points, one could leverage the covariances derived from the coordinates of the data points to approximate the Gaussian distribution that governs them. However, when the number of observation points is rather moderate, such an approximation typically leads to a poor estimation of the underlying distribution. Hence, the proposed algorithm makes use of the intuition that the distances between successive fixation points in an AOI and the distances between successive saccade points come from two distinct Gaussian distributions; thus, following the idea that the structure of two-dimensional data points can be implicitly understood by looking at the distances between them. The resulting (reduced) dimensionality allows the algorithm to effectively deal with a moderate number of observations. The parameters of the two Gaussian distributions (i.e., means and variances) that govern the two different types of distances are learned through a generative mixture model. The Bayesian network in Figure 1 depicts the mixture model for the two Gaussian distributions.

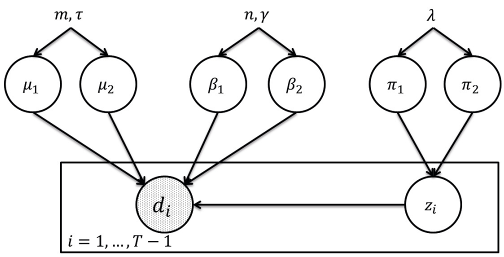


Fig. 1 Bayesian mixture model for the online detection of fixation clusters and saccades

Let $\mathbf{D} = \{d_i \mid 1 \leq i \leq T-1\}$ be the set of distance variables between points $s_i, s_{i+1} \in \mathbf{S}$. $\boldsymbol{\Theta} = \{\mu_1, \beta_1, \mu_2, \beta_2\, \pi_1, \pi_2\}$ denotes the complete parameter set of the mixture model that is depicted in Figure 1. The mixture component is denoted by the variable z_i and the observed distances by the observed variables d_i. The simplifying assumption here is that the distances are generated sequentially in an independent and identically distributed fashion. More specifically, each distance between two successive points is generated independently by the most likely Gaussian distribution.

The joint probability distribution of the model is given by:

$$\begin{aligned} p(\mathbf{D}, \mathbf{z} | \boldsymbol{\Theta}) &= \prod_{i=1}^{T-1} p(z_i | \boldsymbol{\pi}) p(d_i | \mu_{z_i}, \beta_{z_i}) \\ &= \prod_{i=1}^{T-1} \pi_{z_i} N(d_i; \mu_{z_i}, \beta_{z_i}) \end{aligned}$$

where $\mathbf{z} = \{z_1, ..., z_{T-1}\}$, with $z_i \in \{1, 2\}$ being the index of the mixture component chosen for distance d_i, and $\boldsymbol{\pi} = \{\pi_1, \pi_2\}$ denotes the set of mixture parameters.

We have used Infer.NET [MWGK12] to specify the model with the following distributions.

(1) The factorized distribution over the probabilities of each mixture component:

$$p(\mathbf{z}|\boldsymbol{\pi}) = \prod_{i-1}^{T-1} \pi_{z_i}$$

(2) The factorized prior distribution over the model parameters:

$$p(\boldsymbol{\Theta}) = p(\boldsymbol{\pi})p(\boldsymbol{\mu})p(\boldsymbol{\beta})$$

Furthermore, for the online version of the model, the following definitions are needed.

(3) The prior distribution generating the π_1 and π_2 (i.e., the mixture parts) is defined as a symmetric Dirichlet distribution:

$$p(\boldsymbol{\pi}) = Dir(\boldsymbol{\pi};\lambda) \tag{1}$$

Note that the Dirichlet distribution is the so-called conjugate prior of the Multinomial distribution governing π_1 and π_2. This means that the posterior distribution on $\boldsymbol{\pi}$ has a similar mathematical form as the Dirichlet distribution, allowing the Dirichlet prior to be updated by the posterior distribution as more observations are made [Bis06]. Similar reasoning holds for the following definitions.

(4) The prior distribution over the means as a product of Gaussians:

$$p(\boldsymbol{\mu}) = N(\mu_1;m,\tau)N(\mu_2;m,\tau) \tag{2}$$

(5) The prior distribution over the precisions as a product of Gammas:

$$p(\boldsymbol{\beta}) = Gam(\beta_1;n,\gamma)Gam(\beta_2;n,\gamma) \tag{3}$$

Figure 2 represents the Gaussian distributions for fixation clusters and sccades, learned by the above model on a real-world data set.

The above Bayesian model comes with the great benefit that it can be easily turned into an online learning model. In general, for given model parameters $\boldsymbol{\Theta}$ and observations $\mathbf{D}$, after applying Bayes' rule follows that the probability of the parameters $\boldsymbol{\Theta}$ in light of the data $\mathbf{D}$ is:

$$p(\boldsymbol{\Theta}|\mathbf{D}) = \frac{p(\mathbf{D}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})}{p(\mathbf{D})}$$

More generally, we can write:

$$p(\boldsymbol{\Theta}|\mathbf{D}) \propto p(\mathbf{D}|\boldsymbol{\Theta})p(\boldsymbol{\Theta})$$

The above formula suggests that in an online setting the prior of the parameters $p(\boldsymbol{\Theta})$ can be iteratively substituted with the posterior $p(\boldsymbol{\Theta}|\mathbf{D})$, while the likelihood on the parameters $p(\mathbf{D}|\boldsymbol{\Theta})$ helps readjust the model, as more and more observations are made (see also [Bis06]).

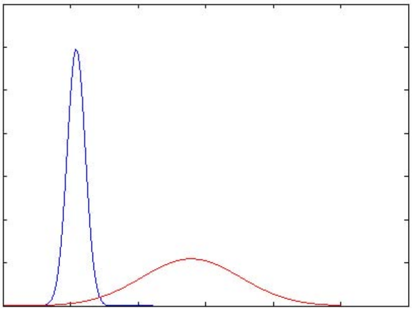


Fig. 2 Two Gaussian distributions learned by the Bayesian mixture model. The left distribution represents distances between successive eye-tracking points that belong to fixation clusters, and the right distribution reflects distances between successive saccadic data points [Kas13].

The above formulation allows the model to readjust the learned parameters $\boldsymbol{\mu}, \boldsymbol{\beta}, \boldsymbol{\pi}$ as more and more eye movement data is available. To this end, we use Gaussian distributions as conjugate priors for the means μ_1, μ_2 (see Equation 2), Gamma distributions as conjugate priors for the precisions β_1, β_2 (see Equation 3), and Dirichlet distributions as conjugate priors for π_1, π_2 (see Equation 1). All these distributions belong to the so-called exponential family of distributions, meaning that they have the same abstract mathematical form. This allows the above prior distributions to be iteratively updated by the posterior distributions on the corresponding variables of the model as new data points are observed.

The whole model was implemented in C# and Infer.NET. The latter not only allows the declarative definition of models such as the above, it also provides various methods for approximate inference; we have used Variational Message Passing as implemented by Infer.NET. In the following we show the fragment of the C# code that is responsible for the update of the priors of the parameters with their posteriors. Initially, we let the inference engine infer all parameters of the model and their posteriors based on a small number (numTrainingPoints) of training points (see lines 1-4). For every other subsequent data point, the priors of the parameters are set to their inferred posterior values (see while loop). The subsequent data point is clustered based on these updated parameters.

```
    ...
 1. engine.InferAll(weights, means, precs);
 2. var meanPost = engine.Infer<Gaussian[]>(means);
 3. var precPost = engine.Infer<Gamma[]>(precs);
 4. var weightPost = engine.Infer<Dirichlet>(weights);
 5. int idxDataPoints = numTrainingPoints+1;
 6. while (iterDataPoints.MoveNext()){
 7.    var newObsData = new double[]{iterDataPoints.Current};
 8.    numData.ObservedValue = newObsData.Length;
 9.    data.ObservedValue = newObsData;
10.    meanPrior.ObservedValue = meanPost;
```

```
    precPrior.ObservedValue = precPost;
    weightPriors.ObservedValue = weightPost;
    ...
    idxDataPoints++;}
```

3.2.2 Principal Component Analysis for the Detection of Smooth Pursuits

Reliably distinguishing fixations from saccades is an important first step towards automated driving support and the prediction of hazardous situations. However, in driving scenarios, objects are typically in motion relative to the driver. Therefore it is crucial to automatically recognize objects that are being pursued by the driver's gaze and others that are not. To address this issue, in [TKK$^+$13] we have extended the above method to also recognize smooth pursuits. We first describe the general idea and then the details of the algorithm.

Let us assume that the last k gaze points were labeled by the above mixture model as fixation points. The key question is whether these points are centered around a relatively stable target or correspond to a moving object that is being pursued by the driver's gaze. In the former case, the vector that represents the highest variability in the k data points and the one representing the second highest variability will have approximately similar lengths. In the latter case though, there will be a notable difference in the lengths of the two vectors. The direction of highest variability corresponds the direction of movement of the followed object relative to the observer. Note that these vectors correspond to the first and the second eigenvectors of the covariance matrix of the data points. We rely on Principal Component Analysis [Jol86] to efficiently retrieve these vectors.

More specifically, let $\mathbf{M}$ be the matrix holding the last k successive data points that were all labeled as fixation points, such that each row of $\mathbf{M}$ contains the coordinates of a point after subtracting the (empirical) mean of the distribution of all points. Through singular value decomposition, $\mathbf{M}$ can be decomposed into $\mathbf{U\Sigma V}^T$, where $\mathbf{U}$ contains the orthonormal eigenvectors of the covariance matrix $\mathbf{MM}^T$, $\mathbf{\Sigma}$ is a diagonal matrix containing the positive roots of the eigenvalues of $\mathbf{MM}^T$, and $\mathbf{V}$ contains the orthonormal eigenvectors of $\mathbf{M}^T\mathbf{M}$. This decomposition is unique up to different orderings of the values in $\mathbf{\Sigma}$. This means that if the values are ordered decreasingly (with the largest value in the upper-left corner of the matrix) in $\mathbf{\Sigma}$, then the first and the second column of $\mathbf{U}$ correspond to the first and second eigenvector of $\mathbf{MM}^T$, respectively.

In order to decide whether the last k data points describe a smooth pursuit, we compute:

$$\frac{\sigma_2^2 \cdot \|u_2\|}{\sigma_1^2 \cdot \|u_1\|} = \frac{\sigma_2^2}{\sigma_1^2} \leq t$$

where t is an empirically established threshold, σ_1 and σ_2 are the largest and the second largest values in $\mathbf{\Sigma}$, respectively, and u_1 and u_2 are the corresponding eigenvectors.

An example scenario of an online smooth pursuit detection is depicted in Figure 3, where a hazardous situation arises from the white car cutting into the lane from the right. The figure shows video frames from a driving scene that was generated in a driving simulator. The driver's eye movements were tracked by means of a mobile Dikablis eye tracker at a sampling rate of 25Hz.

Figure 3(a) shows the moment when the driver's attention is caught for the first time by the white car. The black arrow shows the shift of visual attention in the most recent time frame of 1s. Figure 3(b) depicts the situation 400 ms later, when the driver has come closer to the white car. During this time, the driver's gaze has pursued the relative movement of the white car. According to the spread of the gaze points, which were labeled as a fixation cluster by the online Bayesian mixture model, the PCA-based analysis has classified them as belonging to a smooth pursuit.

(a) (b)

Fig. 3 An example scenario of a smooth pursuit [TKK+13]

4 Experimental Evaluation

In this section, we first demonstrate the superior quality of the online Bayesian mixture model over a state-of-the-art model for detecting fixations and saccades and then showcase the quality of the PCA-based detection of fixations, saccades, and smooth pursuits based on a real-world, hand-labeled data set.

4.1 Evaluation of the Bayesian Mixture Model in Comparison with a Hidden Markov Model

For the quality evaluation of the Bayesian Mixture Model (BMM), in [KKKR14] we compared its prediction performance to that of a Hidden Markov Model (HMM), such as the one presented in [KJKG10, SG00]. These types of models come with several advantages over threshold-based methods: (1) no fixed thresholds are needed, instead the parameters of the model (i.e., state transition probabilities and label emission probabilities) are learned from labeled data, (2) in consequence, HMMs can adapt to the individual (i.e., physiological) viewing behavior of a subject and to the specific task, and (3) given their dynamic sequential structure, they are naturally suited for sequentially arising data points, such as eye-tracking data. Note also that both models, HMM and BMM, belong to the family of probabilistic models.

We implemented a two-state HMM according to the description of the I-HMM in [SG00]. 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 [KKKR14]. 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 estimated Gaussian distributions, i.e., one representing the distribution of distances between saccades and another one standing for the 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. A Viterbi-based, forward-backward algorithm [FJ73] was implemented to compute the most probable state sequence of the HMM corresponding to an observation sequence.

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 [KSA$^+$14]. 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. Each of the data sets corresponds to a driving sequence of 30 seconds, resulting in a total of 4.5 minutes.

Both models were applied post-experimentally, but in real-time on sequentially arising raw eye-tracking data points from nine different data sets. Both the HMM and the BMM were trained on the first 300 eye-tracking points of the each data sets. From such training data, the HMM derives 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.

Table 1 shows a summary of the evaluation of both algorithms with respect to the detection of saccades and fixations from [KKKR14]. The quality results present average values from the evaluation of the models on the above nine data sets 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)* ($\frac{FP+FN}{TP+FP+TN+FN}$).

Table 1 Quality comparison of the Hidden Markov Model (HMM) and the Bayesian Mixture Model (BMM) with respect to the detection of fixations and saccades from raw eye-tracking data

	Data points	Model	Precision	Recall	F1	MCR
Saccade	**910**	**HMM**	0.774	0.803	0.763	0.068
		BMM	0.997	0.955	0.975	0.009
Fixation	**5840**	**HMM**	0.974	0.943	0.958	0.068
		BMM	0.989	1.000	0.994	0.009

As shown in Table 1, the BMM clearly outperformed the HMM, especially in the detection of saccades where an astounding average precision of 99.7% is achieved. Note that the reliable recognition of saccade points is in the context of driving very critical, since a correct detection of saccades implies a correct separation of fixation clusters. 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 distributions. These findings are also in line with the findings presented in [KJKG10].

A much better performance of the HMM is shown in the detection of fixation points, with an average precision of 97.5%. The precision achieved by the BMM is even higher, i.e., 98.9%. 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 1%. In summary, the HMM was outperformed by the BMM with respect to all measures.

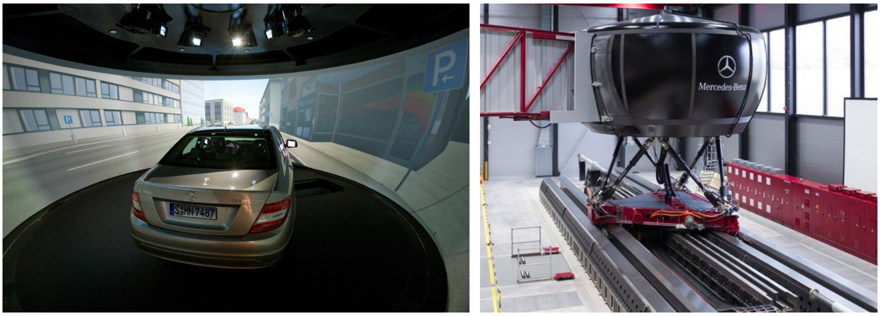
These results highlight the superior performance of the online Bayesian mixture model in comparison to an HMM. Especially, on the difficult task of reliably detecting saccades, which is crucial for the correct separation of fixation clusters, the Bayesian model achieves highly satisfiable precision,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

Beyond the context of driving, the BMM can be integrated into vision research tools (e.g., Vishnoo [TKP$^+$11]) to analyze the viewing behavior during visual search tasks presented on a screen. Furthermore, the BMM can be used in the context of medical testing, e.g., for advanced visual field testing involving online analysis of fixations (e.g., EFOV [THH$^+$13]). In summary, whenever fixations and saccades have to be detected in an online fashion, the BMM is a highly reliable choice.

4.2 *Overall Evaluation of the Classification Technique*

The PCA-based extension of the Bayesian mixture model was evaluated on data from a driving simulator experiment with 27 subjects. The results were presented in [TKK$^+$13].

The driving experiment was conducted in the moving-base driving simulator [Zee10] shown in Figure 4 at the Mercedes-Benz Technology Center in Sindelfingen, Germany. The facility allows simulating acceleration forces in all directions, with up to $1g$ into the direction of a twelve-meter long rail. The cabin, Figure 4(b), contained a real car body (Mercedes S class with automatic transmission) amidst a 360° projected virtual reality, Figure 4(a). The car body was oriented perpendicular to the rail. Thus curve and lane changing maneuvers appear most realistic, while acceleration and braking result in a movement often described as "diving". All in all, acceleration, sound effects, and car environment contributed to a near-to-reality driving experience.



(a) (b)

Fig. 4 Moving base driving simulator. (a) View from inside the cabin onto the environment projection. (b) The entire cabin is mounted on a hexapod, moving along the 12m rail resulting in up to 1g acceleration force. Figures were provided by Daimler AG.

Nine hazardous situations were placed along the 37.5 km long driving route. These traffic hazards, e.g., pedestrians suddenly appearing behind parking cars and trying to cross the road (Figure 5) and risky overtaking maneuvers, were the same for each driver. In case of an overseen hazard, the participants did not experience a crash. For example, it was not possible to run over pedestrians. Instead pedestrians would leap backwards and overtakers would return to their original lane to avoid crashes and subsequent psychological stress. The course contained rural as well as urban areas with different speed limits up to 100 km/h. For the 27 study subjects this would result in a total of 243 hazardous situations. However, some of the study participants aborted the session due to motion sickness or technical problems, resulting in a total of 184 hazardous situations.

(a) (b)

Fig. 5 Scenes from the virtual reality [TKK+13]: (a) a pedestrian intending to cross the road, (b) and a white car coming from the right side.

Driver's eye movements were recorded at 25Hz using a Dikablis mobile eye tracker, whereas head movements were recorded by a LaserBird head tracker. Prior to the driving session each subject underwent a brief training session of 5 km length, in order to adjust to the car and the driving environment. This session was also used to learn an initial visual behavior model with parameters adjusted to the current driver. The training session began with a straight road and became more complex as oncoming traffic became successively denser and traffic signs more frequent.

We evaluated our method on 184 hazardous situations. The spatial extent of the traffic hazards was manually annotated using bounding boxes, separately for each video frame. The analysis of the driver's viewing behavior and the detection of fixations, smooth pursuits and saccades was performed online using the algorithms presented in the previous section. A traffic hazard was considered as perceived, if a fixation or a smooth pursuit cluster intersected with the bounding box around it at any frame. An example of an intersection between the driver's gaze and the bounding box is shown in Figure 6. The results are presented in Table 2.

In 169 of the 171 situations where the hazard was considered as perceived (i.e., an intersection between the fixation cluster and the bounding box occurred) the driver reacted by performing a braking or obstacle avoidance maneuver. In 7 of the 13 hazardous situations where the bounding box was not intersected by a fixation or smooth pursuit cluster no reaction to the situation happened (i.e., the target was missed). Note that in a real-world scenario, these situations would have led to accidents. In 6 situations the bounding box was not intersected but the driver reacted nevertheless. These hazards were perceived, even though the driver did not explicitly look at them. In 2 situations, where the bounding box was intersected by a fixation or smooth pursuit cluster, the driver did not react. Although the driver looked at the targets, they were not perceived. Again, in reality, such situations would result in accidents. In terms of predicting the recognition of traffic hazards by the driver (using only raw eye-tracking data), the algorithm showed an overall accuracy of 95,7%, a specificity of 96,5%, and a sensitivity of 77,8%. These results highlight the overwhelming reliability of the proposed method.

Fig. 6 A car appears on the left of the driving scene and is about to cut the driver's way. The red bounding box around the approaching hazard and the driver's fixation cluster are shown. Once the bounding box is intersected by the fixation cluster, the traffic hazard is marked as "perceived", highlighted by the green bounding box [TKK^{+}13].

In order to look into the detailed per-class performance of our algorithm (i.e., the detection of fixations, saccades, and smooth pursuits), we randomly picked the eye tracking data of one of the subjects. For a six-minute-long driving sequence, two PhD students manually annotated the data points as fixations, saccades, or smooth pursuits. Note that this annotation task is very laborious, as the data has to be labeled frame-wise. Overall, there were 46 eye movement events that were labeled; 27 as fixations, 8 as smooth pursuits, and 11 as saccades. Table 2 shows the per-class true-positive and false-positive counts.

Table 2 True and false positive counts for the detection of fixations, smooth pursuits, and saccades

Eye movement type	Annotation	True Positives	False Positives
Fixation clusters	27	26	1
Smooth pursuit clusters	8	7	2
Saccades (single points)	11	11	0

Seven out of eight smooth pursuits and 26 out of 27 fixations were identified correctly. While our algorithm detected all saccades correctly, two fixations were falsely classified as smooth pursuits and one smooth pursuit as fixation. Although preliminary in nature, these results are very promising and we plan to further evaluate the algorithm on larger labeled datasets.

5 Conclusion

We presented an online adaptive, classification algorithm for detecting fixations, saccades, and smooth pursuits in driving scenarios. This algorithm was primarily evaluated with respect to its ability to detect hazardous traffic situations that might have been overlooked by the driver. In a user study with a state-of-the-art driving simulator, the method showed an impressive detection accuracy, which we think can be mainly explained by the method's ability to adjust the underlying model to the driver's visual behavior. An evaluation on the per-class detection of fixations, saccades, and smooth pursuits hints at the method's ability to recognize and distinguish between different types of eye movements. Apart from experiments on larger, labeled datasets, we also plan to investigate the integration of physiological models, which take heart rate and skin conductance into account to predict the driver's stress levels [KKR+14]. Such models could supplement models that are based on gaze recordings to predict traffic hazard perception and the driver's ability to react.

Acknowledgements. The authors would like to thank Daimler AG and the driving simulator experts for the technical support during the whole experimental study.

References

[BBM+09] Berg, D.J., Boehnke, S.E., Marino, R.A., Munoz, D.P., Itti, L.: Free viewing of dynamic stimuli by humans and monkeys. Journal of Vision 9(5), 1–15 (2009)

[Bis06] Bishop, C.M.: Machine Learning and Pattern Recognition. Springer-Verlag, New York, Inc., Secaucus (2006)

[Bli09] Blignaut, P.: Fixation identification: The optimum threshold for a dispersion algorithm. Attention, Perception, & Psychophysics 71(4), 881–895 (2009)

[Bus35] Buswell, G.T.: How people look at pictures. University of Chicago Press, Chicago (1935)

[BWLH12] Berger, C., Winkels, M., Lischke, A., Höppner, J.: GazeAlyze: A MATLAB toolbox for the analysis of eye movement data. Behavior Research Methods 44(2), 404–419 (2012)

[CNTN08] Camilli, M., Nacchia, R., Terenzi, M., Di Nocera, F.: Astef: A simple tool for examining fixations. Behavior Research Methods 40, 373–382 (2008)

[Cor12] Cornsweet, T.: Visual perception. Academic Press (2012)

[CU98] Chapman, P.R., Underwood, G.: Visual search of driving situations: Danger and experience. Perception London 27, 951–964 (1998)

[CUR02] Chapman, P., Underwood, G., Roberts, K.: Visual search patterns in trained and untrained novice drivers. Transportation Research Part F: Traffic Psychology and Behaviour 5(2), 157–167 (2002)

[Duc07] Duchowski, A.: Eye tracking methodology: Theory and practice. Springer, London (2007)

[Eye] Eyetellect. GazeTracker, `http://www.eyetellect.com/gazetracker/`

[Fer00] Ferrera, V.P.: Task-dependent modulation of the sensorimotor transformation for smooth pursuit eye movements. Journal of Neurophysiology 84(6), 2725–2738 (2000)

[FJ73] Forney Jr., G.D.: The viterbi algorithm. Proceedings of the IEEE 61(3), 268–278 (1973)

[FZ09] Fletcher, L., Zelinsky, A.: Driver inattention detection based on eye gaze-road event correlation. The International Journal of Robotics Research 28(6), 774–801 (2009)

[Git02] Gitelman, D.R.: ILAB: A program for postexperimental eye movement analysis. Behavioral Research Methods, Instruments and Computers 34(4), 605–612 (2002)

[HB05] Hayhoe, M., Ballard, D.: Eye movements in natural behavior. Trends in Cognitive Science 9(4), 188–194 (2005)

[HBCM07] Henderson, J.M., Brockmole, J.R., Castelhano, M.S., Mack, M.: Visual saliency does not account for eye movements during visual search in real-world scenes. In: Eye movements: A Window on Mind and Brain, pp. 537–562 (2007)

[HDBK+13] Hamel, J., De Beukelear, S., Kraft, A., Ohl, S., Audebert, H.J., Brandt, S.A.: Age-related changes in visual exploratory behavior in a natural scene setting. Frontiers in Psychology 4(339) (2013)

[HMM+08] Horswill, M.S., Marrington, S.A., McCullough, C.M., Wood, J., Pachana, N.A., McWilliam, J., Raikos, M.K.: The hazard perception ability of older drivers. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 63(4), P212–P218 (2008)

[HNA+11] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J.: Eye tracking: A comprehensive guide to methods and measures. Oxford University Press (2011)

[HSCG01] Ho, G., Scialfa, C.T., Caird, J.K., Graw, T.: Visual search for traffic signs: The effects of clutter, luminance, and aging. Human Factors: The Journal of the Human Factors and Ergonomics Society 43(2), 194–207 (2001)

[IK00] Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40(10-12), 1489–1506 (2000)

[IKN98] Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11), 1254–1259 (1998)

[Itt05] Itti, L.: Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. Visual Cognition 12(6), 1093–1123 (2005)

[Jol86] Jolliffe, I.T.: Principal Component Analysis. Springer, New York (1986)

[Kas13] Kasneci, E.: Towards the Automated Recognition of Assistance Need for Drivers with Impaired Visual Field. PhD thesis, University of Tübingen, Wilhelmstr. 32, 72074 Tübingen (2013)

[KCC12] Konstantopoulos, P., Chapman, P., Crundall, D.: Exploring the ability to identify visual search differences when observing drivers' eye movements. Transportation Research Part F: Traffic Psychology and Behaviour 15(3), 378–386 (2012)

[KJKG10] Komogortsev, O.V., Jayarathna, S., Koh, D.H., Gowda, S.M.: Qualitative and quantitative scoring and evaluation of the eye movement classification algorithms. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ETRA 2010, pp. 65–68. ACM, New York (2010)

[KK13] 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, 203–215 (2013)

[KKKR14] Kasneci, E., Kasneci, G., Kübler, T.C., Rosenstiel, W.: The applicability of probabilistic methods to the online recognition of fixations and saccades in dynamic scenes. In: Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA 2014, pp. 323–326. ACM, New York (2014)

[KKR+14] Kübler, T.C., Kasneci, E., Rosenstiel, W., Schiefer, U., Nagel, K., Papageorgiou, E.: Stress-indicators and exploratory gaze for the analysis of hazard perception in patients with visual field loss. Transportation Research Part F: Traffic Psychology and Behaviour 24, 231–243 (2014)

[KSA+14] Kasneci, E., Sippel, K., Aehling, K., Heister, M., Rosenstiel, W., Schiefer, U., Papageorgiou, E.: Driving with Binocular Visual Field Loss? A Study on a Supervised On-Road Parcours with Simultaneous Eye and Head Tracking. PLoS ONE 9(2), e87470 (2014)

[Lan09] Land, M.F.: Vision, eye movements, and natural behavior. Visual Neuroscience 26(1), 51–62 (2009)

[LT09] Land, M.F., Tatler, B.W.: Looking and acting: vision and eye movements in natural behaviour. Oxford University Press (2009)

[LZ06] Leigh, R.J., Zee, D.S.: The neurology of eye movements. Oxford University Press (2006)

[Mar10] Markoff, J.: Google cars drive themselves, in traffic. The New York Times 10, A1 (2010)

[McC81] McConkie, G.W.: Evaluating and reporting data quality in eye movement research. Behavior Research Methods & Instrumentation 13(2), 97–106 (1981)

[MS99] Maltz, M., Shinar, D.: Eye movements of younger and older drivers. Human Factors: The Journal of the Human Factors and Ergonomics Society 41(1), 15–25 (1999)

[MS04] Maltz, M., Shinar, D.: Imperfect in-vehicle collision avoidance warning systems can aid drivers. Human Factors: The Journal of the Human Factors and Ergonomics Society 46(2), 357–366 (2004)

[MSP08] Munn, S.M., Stefano, L., Pelz, J.B.: Fixation-identification in dynamic scenes: comparing an automated algorithm to manual coding. In: Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization, APGV 2008, pp. 33–42. ACM, New York (2008)

[MWGK12] Minka, T., Winn, J.M., Guiver, J.P., Knowles, D.A.: Infer.NET 2.5. Microsoft Research Cambridge (2012), http://research.microsoft.com/infernet

[Nag78] Nagayama, Y.: Role of visual perception in driving. IATSS Research 2, 64–73 (1978)

[NH10] Nuthmann, A., Henderson, J.M.: Object-based attentional selection in scene viewing. Journal of Vision 10(8), 20 (2010)

[NS71] Noton, D., Stark, L.W.: Eye movements and visual perception. Scientific American 224(6), 34–43 (1971)

[PHD+05] Pradhan, A.K., Hammel, K.R., DeRamus, R., Pollatsek, A., Noyce, D.A., Fisher, D.L.: Using Eye Movements To Evaluate Effects of Driver Age on Risk Perception in a Driving Simulator. Human Factors: The Journal of the Human Factors and Ergonomics Society 47(4), 840–852 (2005)

[Pom89] Pomerleau, D.A.: ALVINN: An autonomous land vehicle in a neural network. In: Touretzky, D.S. (ed.) Advances in Neural Information Processing Systems 1, pp. 305–313. Morgan Kaufmann, San Francisco (1989)

[PS00] Privitera, C.M., Stark, L.W.: Algorithms for defining visual regions-of-interest: Comparison with eye fixations. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(9), 970–982 (2000)

[PS05] Privitera, C.M., Stark, L.W.: Scanpath theory, attention, and image processing algorithms for predicting human eye fixations. In: Itti, L., Rees, G., Tsotsos, J. (eds.) Neurobiology of Attention, pp. 269–299 (2005)

[SBG11] Schütz, A.C., Braun, D.I., Gegenfurtner, K.R.: Eye movements and perception: a selective review. Journal of Vision 11(9), 1–30 (2011)

[SD04] Santella, A., De Carlo, D.: Robust clustering of eye movement recordings for quantification of visual interest. In: Proceedings of the 2004 Symposium on Eye Tracking Research & Applications, ETRA 2004, pp. 27–34. ACM, New York (2004)

[See] Seeing Machines Inc. faceLab 5, `http://www.seeingmachines.com/product/facelab/`

[Sen] SensoMotoric Instruments GmbH. SMI BeGaze Eye Tracking Analysis Software, `http://www.smivision.com/en/gaze-and-eye-tracking-systems/products/begaze-analysis-software.html`

[SG00] Salvucci, D., Goldberg, J.: Identifying fixations and saccades in eye-tracking protocols. In: Proceedings of the 2000 Symposium on Eye Tracking Tesearch & Applications, ETRA 2000, pp. 71–78. ACM, New York (2000)

[SMDRV91] Sauter, D., Martin, B.J., Di Renzo, N., Vomscheid, C.: Analysis of eye tracking movements using innovations generated by a Kalman filter. Medical and biological Engineering and Computing 29(1), 63–69 (1991)

[SNP96] Summala, H., Nieminen, T., Punto, M.: Maintaining lane position with peripheral vision during in-vehicle tasks. Human Factors: The Journal of the Human Factors and Ergonomics Society 38(3), 442–451 (1996)

[SR] SR Research Ltd. EyeLink 1000 and EyeLink II, `http://www.sr-research.com/index.html`.

[SSC08] Shic, F., Scassellati, B., Chawarska, K.: The incomplete fixation measure. In: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications, ETRA 2008, pp. 111–114. ACM, New York (2008)

[TGB03] Turano, K.A., Geruschat, D.R., Baker, F.H.: Oculomotor strategies for the direction of gaze tested with a real-world activity. Vision Research 43, 333–346 (2003)

[THH+13] Tafaj, E., Hempel, S., Heister, M., Aehling, K., Schaeffel, F., Dietzsch, J., Rosenstiel, W., Schiefer, U.: A New Method for Assessing the Exploratory Field of View (EFOV). In: Stacey, D., SoléCasals, J., Fred, A.L.N., Gamboa, H. (eds.) HEALTHINF 2013, pp. 5–11. SciTePress (2013)

[THLB11] Tatler, B.W., Hayhoe, M.M., Land, M.F., Ballard, D.H.: Eye guidance in natural vision: reinterpreting salience. Journal of Vision 11(5), 5 (2011)

[TKK+13] Tafaj, E., Kübler, T.C., Kasneci, G., Rosenstiel, W., Bogdan, M.: Online classification of eye tracking data for automated analysis of traffic hazard perception. In: Mladenov, V., Koprinkova-Hristova, P., Palm, G., Villa, A.E.P., Appollini, B., Kasabov, N. (eds.) ICANN 2013. LNCS, vol. 8131, pp. 442–450. Springer, Heidelberg (2013)

[TKP+11] Tafaj, E., Kübler, T., Peter, J., Schiefer, U., Bogdan, M., Rosenstiel, W.: Vishnoo - an open-source software for vision research. In: Proceedings of the 24th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2011, pp. 1–6. IEEE (2011)

[TKRB12] Tafaj, E., Kasneci, G., Rosenstiel, W., Bogdan, M.: Bayesian online clustering of eye movement data. In: Proceedings of the Symposium on Eye Tracking Research & Applications, ETRA 2012, pp. 285–288. ACM, New York (2012)

[Tob] Tobii Technology AB. Eye Tracking for Research and Analysis, http://www.tobii.com/en/eye-tracking-research/global/.

[UAB+08] Urmson, C., Anhalt, J., Bagnell, D., Baker, C., Bittner, R., et al.: Autonomous driving in urban environments: Boss and the urban challenge. Journal of Field Robotics 25(8), 425–466 (2008)

[UCBC02] Underwood, G., Chapman, P., Bowden, K., Crundall, D.: Visual search while driving: skill and awareness during inspection of the scene. Transportation Research Part F: Traffic Psychology and Behaviour 5(2), 87–97 (2002)

[ULIS07] Urruty, T., Lew, S., Ihadaddene, N., Simovici, D.A.: Detecting eye fixations by projection clustering. In: ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), vol. 3(4), pp. 1–20 (2007)

[UPW+05] Underwood, G., Phelps, N., Wright, C., Van Loon, E., Galpin, A.: Eye fixation scanpaths of younger and older drivers in a hazard perception task. Ophthalmic and Physiological Optics 25(4), 346–356 (2005)

[VBG12] Vidal, M., Bulling, A., Gellersen, H.: Detection of smooth pursuits using eye movement shape features. In: Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA 2012, pp. 177–180. ACM, New York (2012)

[VRK+02] Velichkovsky, B.M., Rothert, A., Kopf, M., Dornhöfer, S.M., Joos, M.: Towards an express-diagnostics for level of processing and hazard perception. Transportation Research Part F: Traffic Psychology and Behaviour 5(2), 145–156 (2002)

[Wid84] Widdel, H.: Operational problems in analysing eye movements. In: Gale, A.G., Johnson, F. (eds.) Theoretical and Applied Aspects of Eye Movement Research Selected/Edited Proceedings of The Second European Conference on Eye Movements. Advances in Psychology, vol. 22, pp. 21–29. North-Holland (1984)

[Woo02] Wooding, D.S.: Fixation maps: quantifying eye-movement traces. In: Proceedings of the Eye Tracking Research and Applications, pp. 31–36 (2002)

[Yar67] Yarbus, A.L.: Eye movements and vision. Plenum Press, New York (1967)

[Zee10] Zeeb, E.: Daimler's New Full-Scale, High-dynamic Driving Simulator–A Technical Overview. In: Proceedings of the Driving Simulator Conference Europe, pp. 157–165. Institut national de recherche sur les transports et leur sécurité (2010)