Bayesian Identification of Fixations, Saccades, and Smooth Pursuits

Thiago Santini *
Perception Engineering Group
University of Tübingen
Germany

Wolfgang Fuhl †
Perception Engineering Group
University of Tübingen
Germany

Thomas Kübler ‡
Perception Engineering Group
University of Tübingen
Germany

Enkelejda Kasneci §
Perception Engineering Group
University of Tübingen
Germany

Abstract

Smooth pursuit eye movements provide meaningful insights and information on subject's behavior and health and may, in particular situations, disturb the performance of typical fixation/saccade classification algorithms. Thus, an automatic and efficient algorithm to identify these eye movements is paramount for eye-tracking research involving dynamic stimuli. In this paper, we propose the Bayesian Decision Theory Identification (I-BDT) algorithm, a novel algorithm for ternary classification of eye movements that is able to reliably separate fixations, saccades, and smooth pursuits in an online fashion, even for low-resolution eye trackers. The proposed algorithm is evaluated on four datasets with distinct mixtures of eye movements, including fixations, saccades, as well as straight and circular smooth pursuits; data was collected with a sample rate of 30 Hz from six subjects, totaling 24 evaluation datasets. The algorithm exhibits high and consistent performance across all datasets and movements relative to a manual annotation by a domain expert (recall: $\mu = 91.42\%$, $\sigma = 9.52\%$; precision: $\mu = 95.60\%$, $\sigma = 5.29\%$; specificity $\mu = 95.41\%$, $\sigma = 7.02\%$) and displays a significant improvement when compared to I-VDT, an state-of-the-art algorithm (recall: $\mu = 87.67\%$, $\sigma = 14.73\%$; precision: $\mu = 89.57\%$, $\sigma = 8.05\%$; specificity $\mu = 92.10\%$, $\sigma = 11.21\%$). Algorithm implementation and annotated datasets are openly available at www.perception.uni-tuebingen.de

Keywords: smooth pursuit, eye-tracking, probabilistic, model, online, classification, dynamic stimuli, open-source

Concepts: •**Computing methodologies** → **Machine learning approaches;** *Model development and analysis;* •**Applied computing** → *Computers in other domains;* •**Computer systems organization** → *Real-time systems;*

*e-mail: thiago.santini@uni-tuebingen.de
†e-mail: wolfgang.fuhl@uni-tuebingen.de
‡e-mail: thomas.kuebler@uni-tuebingen.de
§e-mail: enkelejda.kasneci@uni-tuebingen.de

ETRA '16,, March 14-17, 2016, Charlston, SC, USA
ISBN: 978-1-4503-4125-7/16/03
DOI: http://dx.doi.org/10.1145/2857491.2857512

1 Introduction

The human visual perception involves mainly six types of eye movements: fixations, saccades, smooth pursuits, optokinetic reflex, vestibulo-ocular reflex, and vergence [Leigh and Zee 2015]. The automatic and correct identification of these eye movements based on the raw eye-position signal is critical for several applications, such as driver's activity recognition [Braunagel et al. 2015] or detection of hazard perception during driving [Kasneci et al. 2015]), marketing applications, and Human Computer Interfaces (HCI) [Vidal et al. 2013].

Initially, eye-tracking research restrained head movements and employed *static stimuli*, such as images and text. In this scenario, the only relevant movements considered were *fixations* (in which the eyes are relatively still) and *saccades* (rapid transitions from one fixation point to another); thus, early algorithms for the automatic classification of eye movements focused on segregating only between these two movements. Nowadays, there is an increasing interest in using *dynamic stimuli* (e.g., video clips) [Larsson et al. 2015], where an object of interest moves through the subject's field of view and is kept on the fovea, producing a fluent eye motion – which we denominate a *smooth pursuit*. It is worth noticing that during these pursuits, minor eye movements such as tremors and microssacades exist, albeit these minor movements do not show on low-resolution eye-tracking data.

The presence of smooth pursuits disturbs the performance of established event classification algorithms since these pursuits end up spread over the two classification classes. Moreover, they also provide valuable information on subject's health and behavior; for instance, smooth pursuit impairment and dysfunction have been linked to mental illnesses, such as schizophrenia [ODriscoll and Callahan 2008] and Alzheimer's disease [Fletcher and Sharpe 1988]. Thus, an automatic and efficient algorithm to distinguish between fixations, saccades, and smooth pursuits is paramount for eye-tracking research involving dynamic stimuli. Furthermore, some of the possible applications must be in the form of embedded systems (e.g., driving assistance) and impose real-time, processing, and energy consumption constrains on the eye-tracking system. To meet these constraints, typically eye trackers with a lower sample rate are used. Consequently, such an algorithm must not only work in real-time, but also be able to deal with the low resolution arising from such eye trackers.

In this paper, we propose a novel algorithm for ternary classification of oculomotor events. Our main contributions are:

- We propose the Bayesian Decision Theory Identification (I-BDT) algorithm to identify fixations, saccades, and smooth pursuits in real-time for low-resolution eye trackers. Additionally, the algorithm operates directly on the eye-position signal and, thus, requires no calibration.

- The proposed algorithm is evaluated relative to manual annotation by a domain expert, and performance is measured in terms of recall, precision, specificity, and accuracy; on average, the proposed algorithm scores above 90% on all metrics.
- I-BDT's performance is compared to that of a state-of-the-art algorithm (Velocity and Dispersion Threshold Identification), showing a significant improvement in terms of average score and variability.
- Additionally, we openly provide a *MATLAB* implementation for the I-BDT algorithm as well as the annotated datasets used for evaluation at www.perception.uni-tuebingen.de.

2 Related Work

In 1991, [Sauter et al. 1991] proposed using a Kalman filter coupled with a χ^2-test to separate saccades from other eye movements. This approach was later extended as the Attention Focus Kalman Filter (AFKF) by [Komogortsev and Khan 2007], using velocity and temporal thresholds to separate fixations from smooth pursuits. Similarly, several methods use a simple *velocity threshold* to isolate saccades, followed by a second step to distinguish between fixations and smooth pursuits. These are typically identified by a name following the pattern *I-V**. [Komogortsev and Karpov 2013] proposed to distinguish between the remaining movements through a *second velocity* threshold (Velocity and Velocity Threshold Identification (I-VVT)) and through a *dispersion* threshold combined with a temporal window (Velocity and Dispersion Threshold Identification (I-VDT)). [Berg et al. 2009] proposed analyzing the ratio between first and second principal components to identify smooth pursuits (Principal Component Analysis Identification (I-PCA)) on the intuition that fixations would have a ratio close to one. [Lopez 2009] started a subgroup that uses the *movement pattern* to identify smooth pursuits, hence the common prefix *I-VMP*; [Lopez 2009] used the standard deviation of the movement directions in a time window to isolate fixations (Velocity Movement Pattern Standard Deviation Identification (I-VMPStd)). [Larsson 2010] used a Rayleigh test to identify smooth pursuits by rejecting the hypothesis of uniformity of inter-sample vectors around the unit circle (Velocity Movement Pattern Rayleigh Identification (I-VMPRay)); more recently, this algorithm was extended with four different spatial features (dispersion, consistent direction, positional displacement, and spatial range) in [Larsson et al. 2015].

[Tafaj et al. 2012] used a Bayesian Mixture Model based on the Euclidean distance between sequential points to discern fixations from saccades, which was later extended in [Kasneci 2013] with a principal component analysis similar to I-PCA to identify smooth pursuits. This method is called the Bayesian Mixture Model Identification (I-BMM). [Vidal et al. 2012] defined a set of shape features, whose expected range is derived from training data. A k-nearest neighbors classifier ($k = 3$) is then used to isolate smooth pursuits from other movements.

As illustrated in Figure 1, the above methods fall mainly into two classes: *threshold-based* and *probabilistic* methods. While threshold-based algorithms tend to be simpler to implement, their major drawback is that they usually depend on the eye movements being clearly discernible from each other. On the other hand, probabilistic methods work based on softer decision rules in the form of probabilities, making them more flexible. Hybrid methods combine insights from physiological limits to define clear thresholds (e.g., only during saccades the eyes reach velocities above $100\,^\circ$ s [Meyer et al. 1985]) with a probabilistic approach in other cases. I-BDT, the method proposed in this paper, falls into the probabilistic group.

Furthermore, most previous work has focused on eye trackers with high sampling rates (i.e., above 250 Hz). However, in dynamic scenarios where a non-intrusive head-mounted eye tracker is required (e.g., driving assistance), such high sampling rates are not available. Currently, mostly head-mounted eye trackers present an upper limit of 60 Hz for binocular tracking (e.g., Dikablis Pro, SMI Glasses 2, ASL H7 Optics, Tobii Pro Glasses 2). The exception is SR Research's EyeLink II, which has a binocular sampling rate of 500 Hz. Despite its clear advantage in temporal resolution, this eye tracker is rather intrusive, occupying a large part of the subject's field of view; for comparison, EyeLink II's eye cameras measure each approximately 5 cm × 5 cm × 1 cm whereas Dikablis Pro's eye cameras measure approximately only 2.5 cm × 2 cm × 1 cm, resulting in a volume difference of five times.

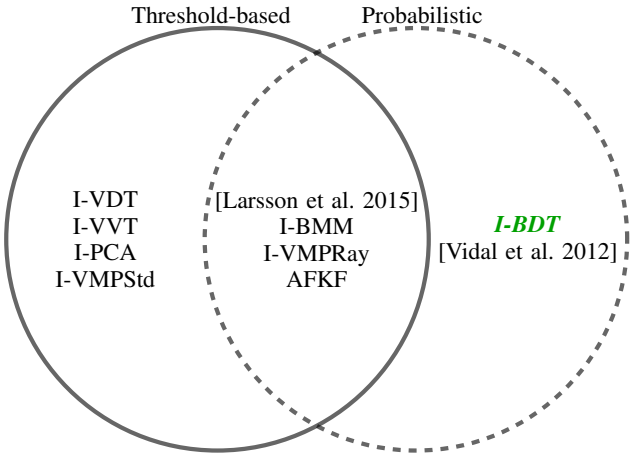

Figure 1: *Algorithms for the automatic identification of smooth pursuits according to a broad classification based on their underlying mechanisms. The algorithm proposed on this work (I-BDT) falls within the probabilistic group.*

3 Bayesian Decision Theory Identification

3.1 Problem Statement

Let $S = \{s_i | 1 \leq i \leq N\}$ be a set of N temporally ordered tuples, each containing two-dimensional pupil position estimates (x_i, y_i) and a timestamp (t_i) generated by an eye tracker (i.e., an eye-tracker protocol). The problem, thus, is to classify all periods between two subsequent tuples according to the set of possible events $E = \{fix, sac, pur\}$, where fix, sac, and pur stand respectively for fixation, saccade, and smooth pursuit.

3.2 Model

In this paper, we propose a Bayesian decision theory approach to solve the stated problem based on a pair of features derived from S. In other words, given some data D, we are interested in defining the *likelihoods* $p(D|e)$ and *priors* $p(e)$ for each event $e \in E$ in order to calculate the *posteriors* $p(e|D)$ of these events. Following the naming convention from [Komogortsev and Karpov 2013] and [Salvucci and Goldberg 2000], we will hereby refer to this method as the Bayesian Decision Theory Identification (I-BDT) algorithm.

The first feature derived from S is the estimated eye speed (v_i) between two subsequent tuples, defined as

$$v_i = \frac{\sqrt{\Delta {x_i}^2 + \Delta {y_i}^2}}{\Delta t_i} \tag{1}$$

where $\Delta x_i = x_i - x_{i-1}$, $\Delta y_i = y_i - y_{i-1}$, and $\Delta t_i = t_i - t_{i-1}$.

The second derived feature is the movement ratio r_i over the window $W_i = \{v_j | i - N_w < j \leq i\}$ of the latest N_w tuples. For simplicity, we define it as the amount of non-zero eye speed estimates relative to the window size, conveying the idea that the more movement in the window, the more likely a smooth pursuit is; thus,

$$r_i = \frac{1}{N_w} \sum_{v_j \in W_i} [v_j > 0] = \frac{1}{N_w} \sum([W_i > 0]) \tag{2}$$

where [...] is the Iverson bracket notation [Knuth 1992] given by

$$[X] = \begin{cases} 1 & \text{if X is true;} \\ 0 & \text{otherwise.} \end{cases}$$

It is paramount to note that this feature's definition is heavily dependent on the eye tracker used to record the data and its temporal and spatial resolution; zero speed may not be an appropriate representation for fixations. Nevertheless, the intuition behind this feature is that fixations exhibit little continuous movement, saccades are brief and usually separated by fixations, and smooth pursuits tend to exhibit continuous movement during larger periods of time (see Figure 2). Therefore, r_i should be a good smooth pursuit indicator if

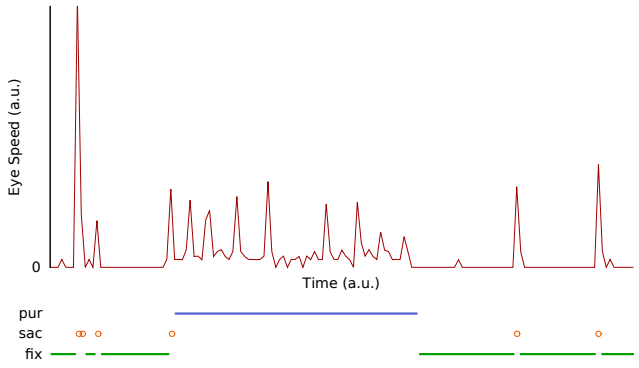

Figure 2: *Eye speed compared to manual classification by a domain expert. Fixations (fix) tend to be mostly still, with only few deviations due to micro eye movements and measurement noise, whereas saccades (sac) result in brief spikes in the eye speed signal. On the contrary, smooth pursuits (pur) show a distinct speed pattern during a longer period of time.*

an adequate window size is chosen; this time window should be large enough to encompass the maximum saccade duration, otherwise misclassification of saccades as pursuits may be exacerbated. In our model, we use this feature directly as the smooth pursuit likelihood, i.e.,

$$p(r_i | pur) = r_i. \tag{3}$$

Once a smooth pursuit has started, it tends to continue for an arbitrary period; thus, this should be reflected in one's belief before any evidence is taken into account. For this reason, we model the smooth pursuit prior as the mean of previous smooth pursuit likelihoods (i.e., the set $L_i = \{p(r_j|pur) | i - N_w < j < i\}$) such that

$$p(pur) = \frac{1}{N_w - 1} \sum_{p(r_j|pur) \in L_i} p(r_j | pur). \tag{4}$$

Naturally, the joint probability of priors must sum to one. With no further evidence, we do not have reason to believe either fixations or saccades are more probable, and, thus, we divide the remaining joint prior probability equally between these movements such that

$$p(fix) = p(sac) = \frac{1 - p(pur)}{2}. \tag{5}$$

It is worth noticing that if information on the task being performed by the subject is available, one could improve these priors based on the duration of the current event. For instance, imagine a task characterized by fixations with a relatively constant duration: after a first fixation is found, the following events are likely to be fixations until the average fixation duration is reached. At this point the next event becomes less and less probable to be a fixation. Such behaviour could be taken into account by adjusting the priors.

The fixational and saccadic likelihoods are deemed to be dependent only on the current eye speed (v_i) feature. This feature can be used to reliably separate high-speed saccades from other events as it has been shown that no other event can reach a velocity higher than V_{sac}, estimated to be around 100 °/s [Meyer et al. 1985]. However, the speed spectra of different eye movements overlap for lower velocities. Nonetheless, it is intuitive that velocities closer to zero are more likely to stem from fixations whereas velocities closer to V_{sac} are more likely to stem from saccades. In fact, [Tafaj et al. 2012] have shown that saccades and fixations can be represented by a mixture model of two Gaussian distributions based on the distance between sequential points – one Gaussian generating fixations, and another one generating saccades. Therefore, we assume the eye speed feature to also be generated by two such Gaussian distributions. Intuitively, saccade likelihood should be at its maximum for speeds above V_{sac}. Ideally, fixations would exhibit zero speed; however, as they typically include small movements, such as microsaccades and tremors, there is a threshold speed V_{fix} that encompasses these combination of movements. Thus, fixation likelihood should be at its maximum for speeds below V_{fix}. In the interval between these thresholds, we assume the likelihood to be generated by two Gaussian[1] distributions, one centered around V_{fix} and the other around V_{sac} (see Figure 3). Thus,

$$p(v_i | fix) = \begin{cases} N(V_{fix} | V_{fix}, \sigma_{fix}) & \text{if } v_i < V_{fix} \\ N(v_i \quad | V_{fix}, \sigma_{fix}) & \text{if } v_i \geq V_{fix} \end{cases}, \tag{6}$$

and

$$p(v_i | sac) = \begin{cases} N(v_i \quad | V_{sac}, \sigma_{sac}) & \text{if } v_i < V_{sac} \\ N(V_{sac} | V_{sac}, \sigma_{sac}) & \text{if } v_i \geq V_{sac} \end{cases}. \tag{7}$$

Having defined the priors and likelihoods for all events, we can calculate the posterior for each event $e \in E$ given the data $D = \{v_i, r_i\}$ using Bayes' Theorem; thus,

$$p(e|D) = \frac{p(e)p(D|e)}{p(D)}, \tag{8}$$

and the period is classified as the event with highest posterior probability. Here, $p(D)$ is merely a scaling factor that guarantees that the sum of the posterior probabilities sum to one. It is worth noticing that this model can be extended to include other eye movements in the future by determining their priors and likelihoods, and taking these into account when computing $p(D)$.

[1] Denoted as

$$N(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

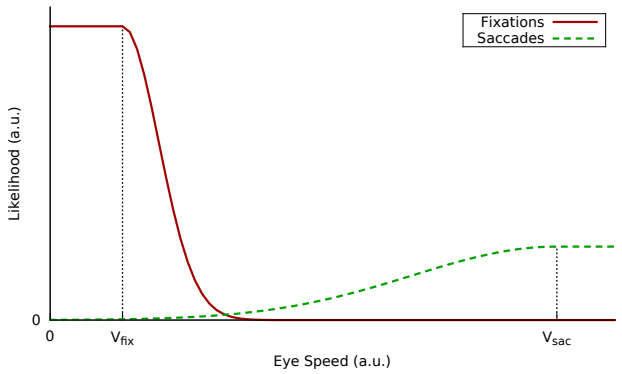

Figure 3: *Resulting fixational and saccadic likelihoods based on the eye speed feature (v_i).*

4 Experimental Setup

4.1 Dataset

To evaluate the proposed algorithm, we designed an experiment to cover a wide range of induced as well as natural eye movements. The induced movements are characterized in Table 1.

Movement	Amplitude[a]/Radius[b] (°)	Velocity (°/s)
Saccade[a]	6, 11, 14	—
Straight Pursuit[a]	6, 12, 22, 28	10, 20, 30
Circular Pursuit[b]	6, 8, 14	18, 25, 44

Table 1: *Induced movements used within the experiment. Degrees are expressed in terms of visual angle. Straight pursuit amplitudes and velocities were combined such that their durations were within 0.4 and 2 seconds to account for subject latency while keeping pursuit duration realistic. Circular pursuits were conducted at a constant angular velocity of 180°/s. Pursuits were separated from other movements by one second fixations. Saccades were separated from each other by fixations of 0.75 seconds. The directions of the movements were chosen randomly and differ per subject.*

Prior to the recording, each user was shown a tutorial with detailed on-screen instructions and examples of movements for each class in Table 1. Four datasets were recorded per subject, and all datasets had a common beginning: first, four dots were shown at 15° of visual angle diagonally from the screen center for five seconds (Figure 4a); subjects were instructed to look at these stimuli at will. During this period natural saccades and fixations are collected; saccades of ≈ 20 and 30° of visual angle were expected, separated by fixations of arbitrary duration. Afterwards, a single dot appeared at the screen center for two seconds (Figure 4b); subjects were instructed to focus on and follow this target. The subsequent movements differ per dataset and are listed in Table 2.

Dataset	Movements
I	Fixations, saccades, and all possible straight pursuits.
II	Fixations and saccades. No pursuits.
III	Fixations, saccades, and all circular pursuits.
IV	Fixations, saccades, straight and circular pursuits.

Table 2: *Movements distribution per dataset.*

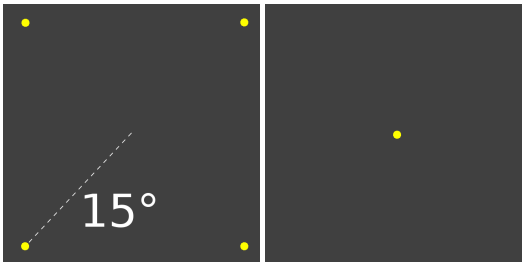

Figure 4: *Common stimuli at the beginning of each dataset. In this figure, the color of the targets was changed from red to yellow to facilitate visualization.*

Targets were red dots (with a width of 1° of visual angle) on a dark gray background displayed using *MATLAB* (r2013a) and the *Psychtoolbox* (3.0.12) [Kleiner et al. 2007] on a Windows 64-bit machine. Subjects' heads were supported by a chin rest at a distance of 300 mm from a *Samsung SyncMaster 2443BW*[2] color display unit. Ocular dominance was determined using the Miles test, and data was collected only from the dominant eye using a *Dikablis Pro* eye tracker (eye images of 384x288 pixels with a 30 Hz sampling rate) and *EyeRec* [Santini et al. 2016] (1.2.2) running the *ExCuSe* [Fuhl et al. 2015] pupil detection algorithm on a distinct Windows 64-bit machine. To avoid gaze estimation noise and calibration requirements, we use the pupil position signal as input; as such, no calibration step was performed. An unjittering function was applied to this input prior to processing to remove obvious jitter artifacts (e.g., one sample spikes [Stampe 1993]). Six adult subjects (age: $\mu = 31.50$, $\sigma = 2.59$ years; 4 males, 2 females) took part in the experiment. Eye location relative to the eye tracker varied greatly between subjects to exacerbate differences in the input signal and stress the algorithm (see Figure 5). Two of the subjects wore corrective glasses for myopia (-13 dpt and 1.5 dpt).

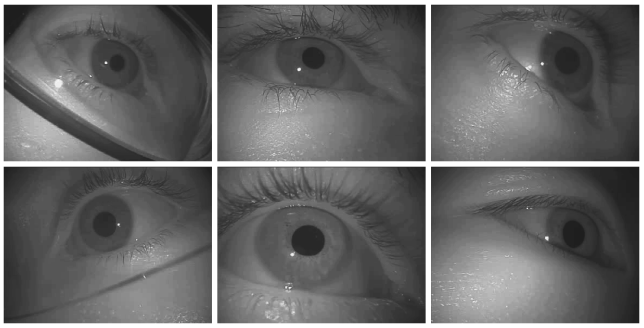

Figure 5: *Example of eye location relative to the eye tracker during experiments. Note the distinct proximities, positions, and rotations.*

4.2 Baseline and Metrics

The collected data was manually classified by one domain expert in order to identify data that is not coherent with the stimulus information, e.g., because the subject did not follow the stimulus as instructed. This manual classification was used as the ground truth.

Nonetheless, it is worth noticing that the manual classification is a subjective task, especially for data with a temporal resolution where

[2]Width: 520 mm. Height: 320 mm. Resolution: 1920x1200 pixels. Screen refresh rate: 60 Hz. Luminance: 0.08 cd/m^2.

a measurement period may contain a mixture of the end of a saccade and the beginning of a fixation. For this reason, we provide our annotated dataset openly to allow for review and potential improvements. Initial corrective saccades during pursuit onset were classified as saccades, whereas catch-up saccades during pursuit were classified as smooth pursuits. Fixation classifications encompass small eye tracker noise, drift and microsaccades. Blinks, partial pupil occlusions, and pupil detection failures were marked as noise and are ignored for performance evaluation; these represent $\approx 1.76\%$ of samples.

Overall, 18,682 fixations, 1,296 saccades, and 4,143 smooth pursuits were classified. Performance is measured through four metrics per movement class, namely: recall $\left(\frac{TP}{TP+FN}\right)$, precision $\left(\frac{TP}{TP+FP}\right)$, specificity $\left(\frac{TN}{TN+FP}\right)$, and accuracy $\left(\frac{TP+TN}{TP+FP+TN+FN}\right)$, where TP, FP, TN, and FN stand for True Positive, False Positive, True Negative, and False Negative, respectively. Moreover, we compare the performance of the proposed algorithm to that of the I-VDT algorithm as implemented by [Komogortsev and Karpov 2013; Komogortsev et al. 2010]; I-VDT was chosen as it can be easily adapted to perform online classification on low-resolution eye trackers, and because it has been shown to exhibit a competitive performance with smaller variability relative to other algorithms [Gyllensten 2014; Komogortsev and Karpov 2013]. Additionally, we also provide *Cohen's Kappa* [Galar et al. 2011] values for the overall classification agreement between the algorithms and the domain expert to account for agreement merely due to chance.

4.3 Algorithm's Parameters

I-BDT: We have chosen a window size to fit 1.5 times the maximum saccade duration (80 ms [Holmqvist et al. 2011]). This value was chosen to fill the minimum size requirement while keeping the window size to a minimum, thus minimizing the duration of the pursuit detection onset. For each subject-dataset pair, the Gaussian distributions parameters are derived from an approximately 15 s of data to demonstrate an online training procedure. Initially, the Expectation-Maximization algorithm was used to derive a mixture of two Gaussian distributions based on speed samples from this period (with the smallest positive scalar supported by the platform added to the estimated covariance matrices to ensure they were positive definite). The parameters of the Gaussian distribution with the highest mean are used as parameters for saccades in Equation (7). However, due to the low resolution of the eye tracker, the Gaussian distribution with the smaller mean is heavily biased towards zero and does not describe fixations adequately; we chose instead to derive the parameters for Equation (6) based on the inherent eye tracker resolution: the minimum dispersion between two samples larger than zero divided by the inter-sample period was taken as V_{fix}, and σ_{fix} was set to $\frac{2}{3}V_{fix}$ such that $\approx 99.7\%$ of the distribution values lie within the interval $[0, 2V_{fix}]$. Furthermore, this low resolution also leads to speed samples with null value during slow smooth pursuits; thus, we have redefined Equation (2) as

$$r_i = \frac{1}{N_w}\sum(smooth([\,0 < W_i < V_{sac}\,])) \tag{9}$$

where the $smooth$ function applies the following logical substitutions over the entire temporal window

$$\begin{cases} 1x1 \;\rightarrow 111 & \text{always} \\ 1xx1 \rightarrow 1111 & \text{if sample } i-1 \text{ was classified as a smooth pursuit} \end{cases},$$

with x representing a *don't care* term. In other words, r_i tolerates a single isolated null speed sample if not currently in a smooth pursuit; otherwise, it is more lenient and tolerates up to two isolated null speed samples. This redefinition implies the temporal window must include at least four samples.

I-VDT: In order to get I-VDT's optimal performance, we give it an advantage by defining pareto-optimal thresholds that maximize Z1 scores based on the ground truth. First, the Z1 score for saccade classification is evaluated for all the inter-sample velocities that can be derived from the eye-tracker protocol; the velocity that maximizes this score is chosen as the *velocity threshold*. Second, the minimum fixation duration is derived from the ground truth and is used as the *temporal window size threshold* (generally around 100 ms). Lastly, fixing the previously defined thresholds, the Z1 score for pursuit classification is evaluated for all the inter-sample dispersions that can be derived from the eye-tracker protocol; the dispersion that maximizes this score is chosen as the *dispersion threshold*. If the ground truth contains no pursuits, the Z1 score for fixation classification is used instead.

5 Experimental Results

First we look at an overview that encompasses all datasets and eye movements to show the overall performance of the proposed algorithm. Afterwards, we analyze our results for separate movements and datasets to provide a comprehensive understanding on the I-BDT behavior. Results are reported using *boxplots* (a box is drawn between the first and third quartiles, a horizontal line represents the median value, and whiskers extend to the minimum and maximum values). Ideally, the value for all metrics should be as close to one as possible. The method introduces no delays, and the average time required to classify a new sample was 0.44 ms, thus attesting for the real-time capabilities of the proposed approach.

5.1 Overall Results

Table 3 and Figure 6 indicate the high performance of the I-BDT algorithm. It is clear that not only I-BDT presents better scores throughout all metrics relative to I-VDT, it also exhibits less variability. Moreover, the high Cohen's kappa score indicates that the inter-rater agreement between expert and algorithm was not due to chance. Note that, in its current form, the algorithm seems to favor precision instead of recall; this is true for smooth pursuits (which can sometimes be misclassified as fixations, specially during onset) and saccades (which are rarely misclassified as smooth pursuits); however, fixations are very seldom misclassified but tend to encompass other movements in its class more often.

	I-BDT	I-VDT
Recall	$\mu = 91.42\%, \sigma = 9.52\%$	$\mu = 87.67\%, \sigma = 14.73\%$
Precision	$\mu = 95.60\%, \sigma = 5.29\%$	$\mu = 89.57\%, \sigma = 8.05\%$
Specificity	$\mu = 95.41\%, \sigma = 7.02\%$	$\mu = 92.10\%, \sigma = 11.21\%$
Accuracy	$\mu = 96.95\%, \sigma = 2.54\%$	$\mu = 94.65\%, \sigma = 4.50\%$

Table 3: *Average algorithm performance per dataset per subject per movement class ($n = 3 \times 6 \times 3 + 1 \times 6 \times 2 = 66$).*

5.2 In-depth Analysis

We start our in-depth analysis by looking at the algorithms performance per dataset for fixations. Figure 7 shows that the algorithm scores highly for the recall and precision metrics for this class, consistently above 90%, and generally above 95%. However, since fixations are the prevalent class in all datasets, false positives are drowned in the larger number of true positives; as a result, it is

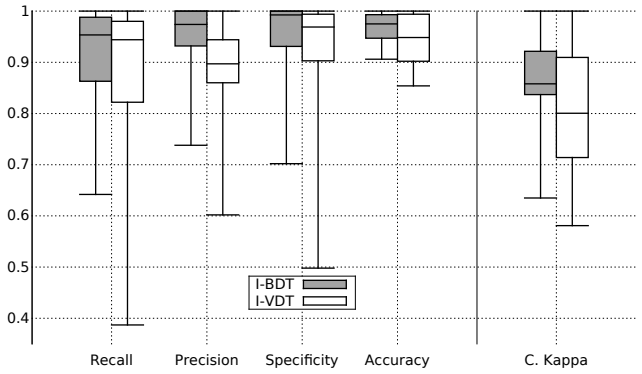

Figure 6: *Overall algorithm performance. Recall, precision, specificity, and accuracy per dataset per subject per movement class ($n = 3 \times 6 \times 3 + 1 \times 6 \times 2 = 66$). Cohen's kappa per dataset per subject ($n = 4 \times 6 = 24$).*

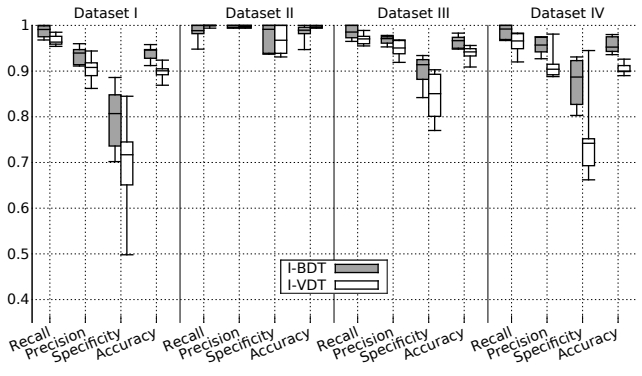

Figure 7: *Performance metrics per dataset for fixations.*

of great importance to look at the specificity when evaluating fixation classification performance. In this case, I-BDT scored above 80% reliably. It is plain that the specificity for dataset II is well above the others, which suggests that the false positives are mostly misclassified smooth pursuits. This is supported by evidence that slow smooth pursuits are the ones being misclassified; specificity for dataset I is almost consistently lower than for dataset III and IV, presumably due to dataset I always including the slowest smooth pursuits. Likewise, specificity for dataset IV is only sometimes lower than that of dataset III because dataset IV only randomly includes the slowest smooth pursuits.

As can be seen in Figure 8 and Figure 9, specificity for both saccades and smooth pursuits classification is persistently high ($> 95\%$). However, similarly to how precision can be misleading for the performance evaluation of fixation classification, specificity can be deceptive for saccades and smooth pursuits classification as false positives get masked by the larger amount of true negatives. Thus, we analyze saccade and smooth pursuit classification through the recall and precision metrics.

Figure 8 shows that saccade classification is very precise ($> 90\%$) in the majority of cases. While the proposed algorithm also displayed a good recall (mostly above 80%), it is clear that some saccades are being misclassified; these are usually saccades surrounded by noise, which the algorithm ends up interpreting as a high movement ratio and, thus, classifying as smooth pursuits. This effect also leads to the I-VDT algorithm outperforming I-BDT for

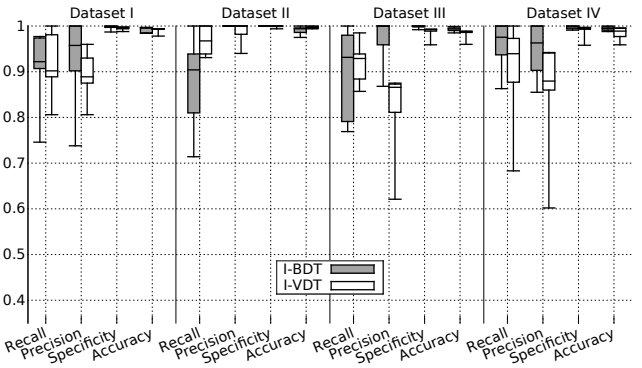

Figure 8: *Performance metrics per dataset for saccades.*

saccade recall for dataset II. Since this dataset contains no smooth pursuits, there is a clear velocity threshold separating the remaining movements, and, thus, I-VDT can clearly distinguish between them. I-BDT, however, is still affected by saccades surrounded by noise, on average classifying 2.18% of the samples as smooth pursuits. In contrast, dataset III exposes one of the I-VDT weaknesses as it contains smooth pursuits with higher speeds (i.e., $44\,^{\circ}$/s); as a result, smooth pursuit and saccade speeds overlap, yielding the misclassification of some high-speed pursuits and decreasing saccade classification precision.

Regarding smooth pursuit classification performance, Figure 7 highlights the consistent good precision ($> 80\%$) through all datasets, scoring above 90% in the great majority of cases. I-BDT exhibits good recall ($> 85\%$) for datasets III and IV. As mentioned previously, for dataset I there is a struggle to classify slow smooth pursuits, resulting in the smaller recall for this dataset. Furthermore, it is worth noticing that I-BDT cannot reach maximum recall by design; since the algorithm relies on a temporal window to consider smooth pursuits, there is an onset period after the smooth pursuit has started until I-BDT starts classifying samples as such.

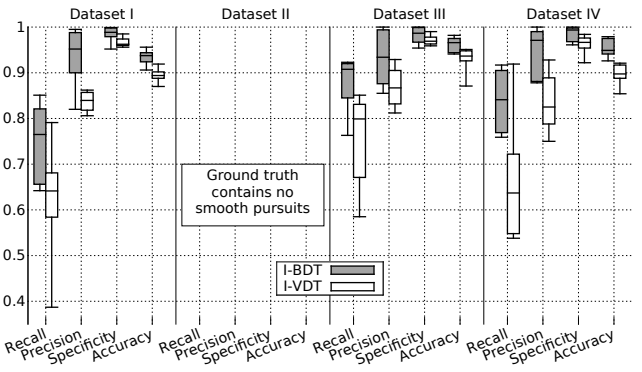

Figure 9: *Performance metrics per dataset for smooth pursuits. Dataset II contains no smooth pursuits in the ground truth; thus, the resulting performance metrics are irrelevant and not reported.*

Figure 10 illustrates I-BDT's smooth pursuit classification relative to that of a domain expert. Notice how the algorithm detects a false short smooth pursuit sequence at the beginning due to a saccade surrounded by noise. In an offline version, such misclassifications could be eliminated, for example, by using a minimum duration threshold for smooth pursuits; the one in question, has a duration

of approximately only 100 ms. Moreover, it is possible to perceive the onset period for the smooth pursuit detection at the beginning of each smooth pursuit; this onset period could also be dealt with in an offline version by employing a similar detection technique but reversing the order of the samples. Furthermore, notice that during the second smooth pursuit the eye speed quickly switches between zero and close to zero values, misleading the algorithm, which does not detect the whole slow pursuit successfully. Thus, we do not advise the usage of I-BDT as is for very slow smooth pursuits when using low-resolution eye trackers; higher resolutions should alleviated this problem, but further investigation is required. It is worth noticing that, despite this weakness, low-resolution eye trackers are more appealing for embedded use in dynamic scenarios because these systems are cheaper, less computationally intensive, and consume less power than their high-resolution counterparts.

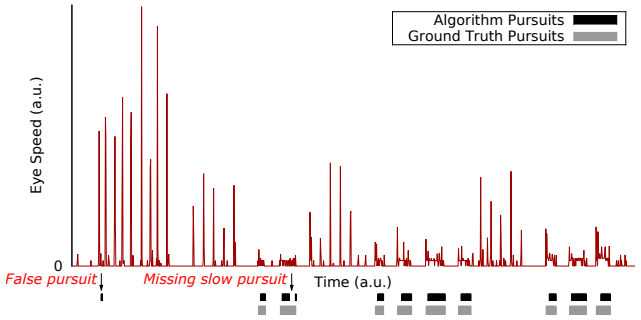

Figure 10: *I-BDT smooth pursuit classification compared to that of a domain expert, accompanied by the eye-speed signal. A wrongly detected smooth pursuit and a partially detected slow smooth pursuit are highlighted. Moreover, notice the onset period required by the algorithm to classify the smooth pursuits.*

Comparing our results to those of related work is relatively complicated, mainly due to the lack of openness regarding algorithms and datasets, and due to differences in eye-tracking systems and metrics used for evaluation. Regarding dataset, eye-tracking system, and online constraints, our work is most similar to [Vidal et al. 2012]. Our dataset design was heavily influenced by the dataset used in [Vidal et al. 2011] and [Larsson et al. 2013]; the main differences are 1) smooth pursuits in this work are not restricted to horizontal and vertical directions, and 2) we chose not to include short smooth pursuits (e.g., amplitude of 2° and velocity of $30\,^\circ$/s) as their durations are smaller than an acceptable latency for the subject to start tracking the target.

In their work, [Vidal et al. 2012] report an accuracy for smooth pursuit detection up to 92% whereas, in this work, I-BDT reached an average accuracy of 94.98%, ranging from 90.57% to 98.19%. It is worth noticing that accuracy alone does not allow us to completely evaluate algorithm performance [Ben-David 2007]. Unfortunately, the machine learning-based classifier presented in [Vidal et al. 2012] is not available for evaluation on our dataset, nor is their dataset available for evaluation with other algorithms. Thus, a direct comparison of both methods could not be performed.

[Larsson et al. 2015] use a subset of the dataset from [Larsson et al. 2013]; however, their algorithm is designed for offline analysis of high-resolution eye-tracking data. Thus, it cannot be applied to low-resolution eye trackers such as the one used in this work – mainly due to the preliminary segmentation stage relying on hypothesis testing, which would require a long time interval from the low-resolution eye tracker to be statistically significant (363 ms compared to the 22 ms used in their work). Nonetheless, their static *image* dataset can to some extent be compared to dataset II (in the sense that both do not contain smooth pursuits inducing elements). Similarly, their *video* and *moving dot* datasets can be compared to datasets I, III, and IV. Since their algorithm uses the same mechanism as I-VDT to separate saccades from other eye movements, their algorithm performance in this regard is clear. Thus, we briefly draw a parallel between their results for smooth pursuit and fixation classification and our results. Table 4 reports recall and specificity values from I-BDT mean results from this work, as well as best case results from [Larsson et al. 2015] – to pick a best case scenario, we utilize the maximum value independent from which expert (1 or 2) was used as ground truth. Although I-BDT seems to provide better performance despite working under harder constraints, a fair and valid conclusion could only be drawn from similar experiments. Nonetheless, it is worth noticing that such an experiment is possible as I-BDT could be applied to the datasets from [Larsson et al. 2015] (e.g., by coalizing the data into a lower resolution or applying I-BDT with adapted parameters). Unfortunately, neither dataset nor algorithm implementation from [Larsson et al. 2015] are available.

		Recall		Specificity	
		I-BDT	**Larsson**	**I-BDT**	**Larsson**
Static	Fixation	0.985	≈0.93	0.977	≈0.98
	Pursuit	N/A	≈0.75	N/A	≈0.97
Dynamic	Fixation	0.986	≈0.90	0.859	≈0.85
	Pursuit	0.822	≈0.80	0.984	≈0.95

Table 4: *Performance comparison between I-BDT and [Larsson et al. 2015]. Static represents the average performance for dataset II compared to the best performance for the images dataset. Dynamic represents the average performance for datasets I, III, and IV compared to the best performance for the videos/moving dot datasets.*

6 Final Remarks

In this paper, we have proposed and evaluated a novel algorithm for the real-time identification of fixations, saccades, and smooth pursuits. Since the algorithm operates directly on the eye-position signal, it requires no calibration step. The proposed algorithm displayed higher and more consistent performance than a state-of-the-art algorithm, demonstrating the capability of I-BDT to provide meaningful ternary classification. Moreover, an open-source *MATLAB* implementation of the algorithm is provided.

One of the main difficulties during evaluation, was the lack of open annotated datasets. The manual coding of eye movements is a subjective, laborious, and time-consuming task; thus, having to create one from scratch is far from ideal. In an effort to allow for review and to kick-start an open-access benchmark for the evaluation of eye movement identification algorithms, we provide our annotated datasets openly at www.perception.uni-tuebingen.de.

For future work, we are interested in analyzing additional features for I-BDT to further improve its performance as well as evaluating the algorithm with higher-resolution eye trackers. Moreover, an important step to enable the fully automation of eye movements classification is a reliable detection of blinks, which the proposed algorithm does not take into account at the moment. Furthermore, we are intent on developing solutions to account for head movements in order to reliably distinguish smooth pursuits from vestibulo-ocular reflexes.

References

BEN-DAVID, A. 2007. A lot of randomness is hiding in accuracy. *Engineering Applications of Artificial Intelligence 20*, 7, 875–885.

BERG, D. J., BOEHNKE, S. E., MARINO, R. A., MUNOZ, D. P., AND ITTI, L. 2009. Free viewing of dynamic stimuli by humans and monkeys. *Journal of Vision 9*, 5, 19.

BRAUNAGEL, C., KASNECI, E., STOLZMANN, W., AND ROSENSTIEL, W. 2015. Driver-activity recognition in the context of conditionally autonomous driving. In *IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*.

FLETCHER, W. A., AND SHARPE, J. A. 1988. Smooth pursuit dysfunction in alzheimer's disease. *Neurology 38*, 2, 272–272.

FUHL, W., KÜBLER, T., SIPPEL, K., ROSENSTIEL, W., AND KASNECI, E. 2015. Excuse: Robust pupil detection in real-world scenarios. In *Computer Analysis of Images and Patterns 2015. CAIP 2015. 16th International Conference*, IEEE.

GALAR, M., FERNÁNDEZ, A., BARRENECHEA, E., BUSTINCE, H., AND HERRERA, F. 2011. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition 44*, 8, 1761–1776.

GYLLENSTEN, O. C. 2014. *Evaluating current algorithms for smooth pursuit detection on Tobii Eye Trackers*. Master's thesis, Royal Institute of Technology, Sweden.

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., KASNECI, G., KÜBLER, T., AND ROSENSTIEL, W. 2015. Online recognition of fixations, saccades, and smooth pursuits for automated analysis of traffic hazard perception. In *Artificial Neural Networks*, P. Koprinkova-Hristova, V. Mladenov, and N. K. Kasabov, Eds., vol. 4 of *Springer Series in Bio-/Neuroinformatics*. Springer International Publishing, 411–434.

KASNECI, E. 2013. *Towards the automated recognition of assistance need for drivers with impaired visual field*. PhD thesis, Universität Tübingen, Germany.

KLEINER, M., BRAINARD, D., PELLI, D., INGLING, A., MURRAY, R., AND BROUSSARD, C. 2007. Whats new in psychtoolbox-3. *Perception 36*, 14, 1.

KNUTH, D. E. 1992. Two notes on notation. *American Mathematical Monthly*, 403–422.

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., AND KHAN, J. I. 2007. Kalman filtering in the design of eye-gaze-guided computer interfaces. In *Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments*. Springer, 679–689.

KOMOGORTSEV, O. V., GOBERT, D. V., JAYARATHNA, S., KOH, D. H., AND GOWDA, S. M. 2010. Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *Biomedical Engineering, IEEE Transactions on 57*, 11, 2635–2645.

LARSSON, L., NYSTROM, M., AND STRIDH, M. 2013. Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit. *Biomedical Engineering, IEEE Transactions on 60*, 9, 2484–2493.

LARSSON, L., NYSTRÖM, M., ANDERSSON, R., AND STRIDH, M. 2015. Detection of fixations and smooth pursuit movements in high-speed eye-tracking data. *Biomedical Signal Processing and Control 18*, 145–152.

LARSSON, L. 2010. *Event detection in eye-tracking data*. Master's thesis, Lund University, Sweden.

LEIGH, R. J., AND ZEE, D. S. 2015. *The neurology of eye movements*. Oxford University Press.

LOPEZ, J. S. A. 2009. *Off-the-shelf Gaze Interaction*. PhD thesis, IT University of Copenhagen, Denmark.

MEYER, C. H., LASKER, A. G., AND ROBINSON, D. A. 1985. The upper limit of human smooth pursuit velocity. *Vision Research 25*, 4, 561–563.

ODRISCOLL, G. A., AND CALLAHAN, B. L. 2008. Smooth pursuit in schizophrenia: a meta-analytic review of research since 1993. *Brain and cognition 68*, 3, 359–370.

SALVUCCI, D. D., AND GOLDBERG, J. H. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, 71–78.

SANTINI, T., FUHL, W., KÜBLER, T., AND KASNECI, E. 2016. Eyerec: An open-source data acquisition software for head-mounted eye-tracking. In *International Conference on Vision Theory and Applications (VISAPP)*. Forthcoming.

SAUTER, D., MARTIN, B., 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.

STAMPE, D. M. 1993. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, & Computers 25*, 2, 137–142.

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.

VIDAL, M., BULLING, A., AND GELLERSEN, H. 2011. Analysing eog signal features for the discrimination of eye movements with wearable devices. In *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*, ACM, 15–20.

VIDAL, M., BULLING, A., AND GELLERSEN, H. 2012. Detection of smooth pursuits using eye movement shape features. In *Proceedings of the symposium on eye tracking research and applications*, ACM, 177–180.

VIDAL, M., BULLING, A., AND GELLERSEN, H. 2013. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM, 439–448.