bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

New Measures of Visual Scanning Efficiency and Cognitive Effort

Zezhong Lv, Qing Xu[§]
College of Intelligence and Computing
Tianjin University
Tianjin, China
zezhonglv0306@gmail.com
qingxu@tju.edu.cn

Klaus Schoeffmann
Institute of Information Technology
Alpen-Adria Universitat Klagenfurt
Klagenfurt, Austria
ks@itec.aau.at

Simon Parkinson
School of Computing and Engineering
University of Huddersfield
Huddersfield, UK
S.Parkinson@hud.ac.uk

***Abstract*—Visual scanning plays an important role in sampling visual information from the surrounding environments for a lot of everyday sensorimotor tasks, such as walking and car driving. In this paper, we consider the problem of visual scanning mechanism underpinning sensorimotor tasks in *3D* dynamic environments. We exploit the use of eye tracking data as a behaviometric, for indicating the visuo-motor behavioral measures in the context of virtual driving. A new metric of visual scanning efficiency (*VSE*), which is defined as a mathematical divergence between a fixation distribution and a distribution of optical flows induced by fixations, is proposed by making use of a widely-known information theoretic tool, namely the square root of *Jensen-Shannon divergence*. Based on the proposed efficiency metric, a cognitive effort measure (*CEM*) is developed by using the concept of quantity of information. Psychophysical eye tracking studies, in virtual reality based driving, are conducted to reveal that the new metric of visual scanning efficiency can be employed very well as a proxy evaluation for driving performance. In addition, the effectiveness of the proposed cognitive effort measure is demonstrated by a strong correlation between this measure and pupil size change. These results suggest that the exploitation of eye tracking data provides an effective behaviometric for sensorimotor activity.**

***Index Terms*—visual scanning efficiency, eye tracking, behaviometric, *Jensen-Shannon divergence* (*JSD*), cognitive effort**

I. Introduction

A. Background

Visual scanning is critical for the living of any human in natural surroundings [1]. Visual scanning is indeed the cornerstone for a human to perform common and everyday sensorimotor tasks, such as walking and car driving. Actually, the understanding of the mechanism behind visual scanning has been very critical since late 1970 and, is especially helpful and beneficial to making stark and essential progress in both theoretical and practical perspectives [2]. As has been widely accepted, visual scanning in sensorimotor activities is a typical task directed behavior contributed by top-down modulation [3]. And this is the focused consideration in this paper.

Basically, it is significant to make clear how well a human can achieve for sampling visual information through visual scanning. In this case, to define an efficiency measure objectively is one of the most fundamental points involved in the understanding of visual scanning mechanism [2]. Also it is notably argued that the measure of efficiency for visual scanning should be indicative for the performance of the corresponding sensorimotor task [4].

But unfortunately, all the relevant research findings available do have some issues, which will be described in Section II. Indeed, it is challenging to measure the visual scanning behavior, due to that the behavior itself, which results from both the bottom-up influence by visual stimuli and the top-down modulation impact by human activity, is very sophisticated, according to the theory of predictive processing and active inference [5]. Besides, there is no established work on defining a measure of visual scanning efficiency as a true math metric, and obviously this remarkable mathematical property makes the measure more usable in theory and practice [6]. In order to combat all the disadvantages mentioned above, a metric of visual scanning efficiency is proposed based on making use of the information theoretic tools, behaving very well as a proxy for the driving performance in the context of virtual driving.

B. Contributions

The main contributions of this paper are as follows.

- The optical flow induced by a fixation is introduced, based on the detection of the visual motion resulted from a fixated dynamic stimulus.
- We propose a methodology to define a mathematical metric of visual scanning efficiency in an objective and quantitative way, by the square root of *JSD* divergence between fixation distribution and optical flow distribution.
- Based on an extended exploitation of our proposed visual scanning efficiency, a cognitive effort measure, in a perspective of the quantity of information, is developed.

To the best of our knowledge, this is the first time to effectively define the metric of visual scanning efficiency and cognitive effort measure, based on the exploitation of information theoretic tools. This paper paves a novel path for behaviometric discovery by the utilization of eye tracking data.

[§] Qing Xu is corresponding author. This work has been funded by Natural Science Foundation of China under Grants No. 61471261 and No. 61771335.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

C. Organization

Firstly, the related works are discussed in Section II. The philosophy and methodology of our proposed technique are detailed in Section III. Then the experimental methods are described in Section IV, followed by the results and discussion presented in Section V. Finally, conclusions and future works are given in Section VI.

II. Related Works

The initial consideration for discussing the efficiency measure of visual scanning is to select suitable eye tracking indices naturally associated with cognitive processing. Fixation and saccade are classic indices for this purpose. But, direct and indirect usages of these indices (for examples, rate/duration of them and their simple combinations) are more applicable in specific application scenarios [7], than in the general efficiency definition. In addition, pupil dilation and blink rate are two widely used eye tracking indices for the study of cognition and psychology [8].

The efficiency evaluation of visual scanning has started from the study of the complexity of visuo-motor behavior and, this is the usual manner in relevant field [2]. Information entropy, which is a good measure of complexity [9], in effect has been used a lot for the issue of visual scanning complexity [2]. The so-called entropy rate, originally used for the measurement of task load [10], is defined as a multiplication between the normalized entropy of fixation sequence and the sum of inverse transition durations. Entropy of fixation sequence (called *EoFS* for brevity) defines based on the probability distribution of fixation sequences, giving the degree of complexity for spatial patterns of fixation sequence [11]. *EoFS* has shown its interaction to visual content complexity [12] and to human factors such as expertise skill [11].

Differently from *EoFS*, gaze transition entropy (*GTE*), in which transition between fixations is depicted by the probability transition between *Markov* states (in this paper, meaning the areas of interest (*AOI*s)) and, conditional entropy based on transition probabilities is used for its definition to express the degree of complexity for temporal patterns of fixation transition, with respect to overall spatial fixation dispersion. *GTE* has demonstrated its relationship with visual content complexity [13] and also with human factors, such as anxiety [14]. There exists a closely relevant gaze entropy, called stationary gaze entropy (*SGE*) [13], in which the *Shannon* entropy, based on an equilibrium probability distribution for *Markov* transition matrix, is used to describe the degree of complexity for spatial patterns of fixation dispersion. *SGE* has also exhibited, similar to *GTE*, its connection to visual content complexity [13], and also to human factors, such as emotional valence [15].

Notice specifically that, it is necessary to know the relationship between entropy based measures discussed above and the performance of sensorimotor tasks. *EoFS* has not been indicated to have correlation with the task performance. It has been reported that *GTE* [1] could not predict the performance of sensorimotor tasks. As for *SGE*, it relates to the complexity of visual content, but can not indicate the performance of task directed sensorimotor activities [3]. As a mater of fact, eye tracking has been expected as a strong estimator for task performance in many professional fields [4]. Obviously, it is indeed indispensable for the efficiency measure of visual scanning to take the role as measure of task performance.

III. Methodology

A. The Philosophy of the Proposed VSE and CEM

It is well understood that, in a sensorimotor task, visual scanning is managed by a performer through a series of fixations, so that meeting the requirement of visual sampling the surrounding environments for fulfilling this task [16]. The probability distribution of fixations, namely the fixation distribution, which is constructed based on the normalized histogram of fixation locations in a *3D* environment, is used as a basis for understanding the mechanism of visual scanning in this paper. In addition, the optical flow, which effectively reflects a visual motion resulted from a dynamic stimulus [17], always happens in driving situations [18]. In this paper, the optical flows induced by fixations are detected and used to establish the optical flow probability distribution. And then a divergence between fixation distribution and optical flow distribution is proposed to define the visual scanning efficiency. Notice that, optical flow represents a vector, according to its original definition [17] [18]; but, for the purpose of this paper, unless otherwise specified, optical flow is always used for representing its magnitude.

The rationale behind our proposed divergence comes from two points. Firstly, optical flow distribution in fact contains the visual motions that are embedded in fixation distribution, the divergence between these two distributions represents visual motions perceived by a performer in visual scanning. Secondly, bear in mind that visuo-motor behavior is usually influenced by both top-down and bottom-up factors and that, the combination of these two factors reflects the visual scanning efficiency [2]. In fact, all the visual motions, which are essential information due to the dynamic nature of driving environment, are considered to result from bottom-up factors [2]. To go further, visual motions perceived by a performer in the procedure of driving actually indicate an effective component of all the visual motions necessarily used for the purpose of fulfilling the driving task. As a result, this so-called effective component can be easily considered to be caused by both top-down modulations and bottom-up influences. And in this case, we hypothesize that the proposed divergence takes as an indicator for visual scanning efficiency.

The square root of *JSD* (*SRJSD*) is chosen for computing the divergence and for extracting the visual motions, because it is largely accepted as a true mathematical metric [19]. Actually, *SRJSD* offers a kind of distance between two probability distributions in a normalized manner [6] [19]. Thus, *SRJSD* between fixation distribution and optical flow distribution in effect indicates a normalized degree to which the visual motions are perceived. To sum up, *SRJSD* between fixation and optical flow distributions should behave as an evaluation

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

function for visual scanning efficiency (*VSE*). The larger the *SRJSD*, the more visual motions are perceived, the higher the *VSE* becomes, and *vice versa*.

Additionally, the proposed *VSE* can be studied from the perspective of probability and information theories. That is, the *VSE*, in fact, can be considered as a probability p an event occurs, because it ranges from 0 to 1. As a result, the *VSE* points out the probability of perception of the visual motions. According to information theory, the logarithmic probability of occurrence ($-\log_2 p$) represents the quantity of information conveyed by the occurrence [9]. We hypothesize that the information quantity may indicate the attentional resources needed for the perception of all the visual motions. Because the amount of attentional resource reflects cognitive effort [8], and thus the logarithm of *VSE* may be the central part of cognitive effort.

A scheme of the calculation of the two proposed measures is shown in Fig. 1. Firstly, optical flows induced by fixations are extracted, to construct the optical flow distribution that actually embeds visual motions into the fixation distribution. Then the

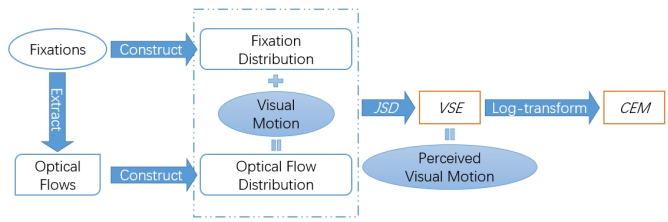

Fig. 1. The scheme of the calculation of the proposed measures.

VSE is obtained by *SRJSD* between these two distributions, for measuring the perceived visual motions in driving. Finally, the *CEM* is obtained by the logarithm of *VSE*.

B. Optical Flow Induced by Fixation

A fixation f_n (its index is n) with duration τ_n by an observer O in a *3D* environment is shown in Fig. 2; in this figure, ρ_n is the distance between O and f_n in the direction of the line of sight, and ρ_n is the depth of f_n from the perspective of O. Considering there is a relative displacement $\boldsymbol{u_n}$, during τ_n, between O and f_n, $\boldsymbol{v_n}$ is defined as the optical flow vector [17] (bold-type letters are used to denote vectors in this paper). Actually, $\boldsymbol{v_n}$ is the projection of $\boldsymbol{u_n}$ in the direction of optical flow vector, and thus perpendicular to the line of sight. The so-called visual motion perceived by the observer, m_n, is defined as the trajectory segment of the circular motion of f_n centered on O in this paper. Note, m_n acts as an approximated magnitude of $\boldsymbol{v_n}$. The central angle I_n subtended by m_n,

$$I_n = \frac{m_n}{\rho_n}, \tag{1}$$

is used to approximately define a perceived magnitude of optical flow, according to the usual practice that angle is widely used to represent the magnitude in eye tracking [20]. Notice that, the definition of the perceived magnitude of optical flow explicitly uses the depth, enabling it to convey an important information in *3D* environments. For the sake of this paper, I_n is called as optical flow.

C. Visual Scanning Efficiency

In this paper, fixations are assigned to *AOI*s, and *AOI*s are defined based on *3D* objects in virtual environment [21]. There exists a one-one correspondence between the *AOI*s and the objects.

The probability distribution of fixations, which is the normalized fixation histogram, is based on the frequencies of fixations located at *AOI*s. There exist a series of fixations $F = \langle f_1, \cdots, f_n, \cdots, f_{N_F} \rangle$, $(n \in C_F = \{1, \cdots, N_F\})$ that happens in a driving procedure, the *AOI*s hit by these fixations are $X_F = \{\chi_i | i = 1, \cdots, N_A\}$. For an *AOI* $\chi_i \in X_F$, all the indices of the fixations $\{f_h | h \in C_F\}$ hitting χ_i constitute a set,

$$\Gamma_F(\chi_i) = \{h\}.$$

Then the fixation distribution $P_f = \{p_f(\chi_i) | \chi_i \in X_F\}$ and the optical flow distribution $P_o = \{p_o(\chi_i) | \chi_i \in X_F\}$ are obtained respectively as

$$\begin{cases} p_f(\chi_i) = \dfrac{|\Gamma_F(\chi_i)|}{|C_F|} \\ p_o(\chi_i) = \dfrac{\sum_{n \in \Gamma_F(\chi_i)} I_n}{\sum_{m \in C_F} I_m} \end{cases}. \tag{2}$$

It is noted that, visual motions, indicated by optical flows, are embedded into the fixation distribution P_f to construct the optical flow distribution P_o.

A special instance of *JSD* between two distributions $P = \{p(x) | x \in X\}$ and $Q = \{q(x) | x \in X\}$ with equal weights [6] is defined as

$$JSD(P||Q) = \frac{\sum_{x \in X} (p(x) \log_2 \frac{2p(x)}{p(x)+q(x)} + q(x) \log_2 \frac{2q(x)}{p(x)+q(x)})}{2}. \tag{3}$$

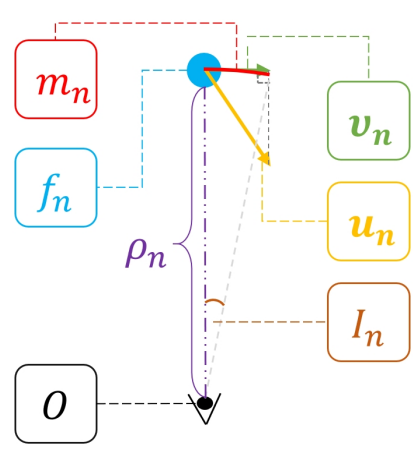

Fig. 2. An illustration of optical flow I_n induced by fixation f_n.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The square root of $JSD(P||Q)$

$$SRJSD(P||Q) = \sqrt{JSD(P||Q)} \tag{4}$$

is used to obtain the metric of visual scanning efficiency,

$$VSE = SRJSD(P_f||P_o), \tag{5}$$

because *SRJSD* is a true mathematical metric [19].

D. Cognitive Effort Measure

Because the distributions of fixation and fixation sequence reflect different aspects of visual scanning, a combination of both these two distributions should provide a more complete understanding of visual scanning behavior, as has been pointed out in relevant work [2]. Therefore, we propose to utilize the fixation sequence distribution for the sake of a scaling function, detailed in the following text of this section.

There exist a series of fixations $F_S = \langle f_1, \cdots, f_n, \cdots, f_{N_f+1} \rangle$, then we have N_f fixation sequences $S = \{f_{\langle n,n+1\rangle} = \langle f_n, f_{n+1}\rangle \,|\, \langle n, n+1\rangle \in C_S\}$, $C_S = \{\langle 1,2\rangle \cdots, \langle N_F, N_F+1\rangle\}$. A set of *AOI* sequences is constituted as $X_S = \{\chi_{\langle i,j\rangle} = \langle \chi_i, \chi_j\rangle \,|\, \langle i,j\rangle \in A_S\}$, where A_S contains the indices of all the *AOI* sequences $\chi_{\langle i,j\rangle}$ hit by all the fixation sequences in S, note that a fixation sequence $f_{\langle n,n+1\rangle}$ hitting *AOI* sequence $\chi_{\langle i,j\rangle}$ indicates a transition from f_n on *AOI* χ_i to f_{n+1} on *AOI* χ_j. This means that the performer of visual scanning voluntarily exerts some cognitive efforts to switch the fixations from f_n to f_{n+1}. And in this case, the optical flow I_n induced by f_n, which is taken for representing the cognitive efforts exerted, is considered lost for the purpose of completing the transition of fixations from f_n to f_{n+1}. So in this paper, for each *AOI* sequence $\chi_{\langle i,j\rangle} \in X_S$, all its corresponding optical flows induced by $f_{\langle n,n+1\rangle}$ satisfying that $f_{\langle n,n+1\rangle}$ hits $\chi_{\langle i,j\rangle}$ are obtained, here each optical flow takes I_n and is denoted as $I_{\langle n,n+1\rangle}$. Then for all the *AOI* sequences in X_S, their corresponding optical flows are embedded into the fixation sequence distribution to construct the corresponding optical flow distribution. Note, the *SRJSD* between these two distributions takes as the scaling function for defining the cognitive effort measure.

For an *AOI* sequence $\chi_{\langle i,j\rangle} \in X_S$, all the indices of the fixation sequences $f_{\langle l,l+1\rangle} \in S$ hitting $\chi_{\langle i,j\rangle}$ constitute a set

$$\Gamma_S(\chi_{\langle i,j\rangle}) = \{\langle l, l+1\rangle\}. \tag{6}$$

Then the distribution of fixation sequences $P_{fs} = \{p_{fs}(\chi_{\langle i,j\rangle}) | \chi_{\langle i,j\rangle} \in X_S\}$ and its corresponding optical flow distribution $P_{os} = \{p_{os}(\chi_{\langle i,j\rangle} | \chi_{\langle i,j\rangle} \in X_S\}$ are obtained as

$$\begin{cases} p_{fs}(\chi_{\langle i,j\rangle}) = \dfrac{|\Gamma_S(\chi_{\langle i,j\rangle})|}{|C_S|} \\ \\ p_{os}(\chi_{\langle i,j\rangle}) = \dfrac{\sum_{\langle n,n+1\rangle \in \Gamma_S(\chi_{\langle i,j\rangle})} I_{\langle n,n+1\rangle}}{\sum_{\langle m,m+1\rangle \in C_S} I_{\langle m,m+1\rangle}} \end{cases}. \tag{7}$$

It is worth pointing out that if an *AOI* sequence $\chi_{\langle i,j\rangle}$ is not hit by any fixation sequence, then this sequence will not be included in X_S. Therefore, it is easy to see that

$$p_{fs}(\chi_{\langle i,j\rangle}), p_{os}(\chi_{\langle i,j\rangle}) > 0, \chi_{\langle i,j\rangle} \in X_S, \tag{8}$$

this means that the probability mass functions of fixation sequence distribution and its corresponding optical flow distribution are strictly positive.

For a certain *AOI* $\chi_i \in X_F$, we have the quantitative relationships that

$$\begin{cases} p_f(\chi_i) = \sum_{\langle i,j\rangle \in A_S} p_{fs}(\chi_{\langle i,j\rangle}) \\ \\ p_o(\chi_i) = \sum_{\langle i,j\rangle \in A_S} p_{os}(\chi_{\langle i,j\rangle}) \end{cases}. \tag{9}$$

Notice that *SRJSD* between two probability distributions P and Q is not smaller than that between the distributions constructed by the sums of the corresponding subsets of P and Q respectively, as proved in Theorem 1.

Theorem *1:* There are two probability distributions $P = \{p(x)|x \in \Psi\}$ and $Q = \{q(x)|x \in \Psi\}$ containing non-empty and non-overlapping subsets μ_j and ν_j $(j = 1, \cdots, M)$ respectively. Here $\mu_j = \{p(x)|x \in \psi_j, \psi_j \subseteq \Psi\}$ and $\nu_j = \{q(x)|x \in \psi_j, \psi_j \subseteq \Psi\}$. Then two sets can be constituted as $P_S = \{\sum_{x\in\psi_j} p(x)|j = 1, \cdots, M\}$ and $Q_S = \{\sum_{x\in\psi_j} q(x)|j = 1, \cdots, M\}$. Then we have

$$SRJSD(P_S||Q_S) \leq SRJSD(P||Q). \tag{10}$$

Proof. The log sum inequality [9] says that, for nonnegative numbers $a_1, \cdots, a_n$ and $b_1, \cdots, b_n$,

$$\sum_{i=1}^{n}(a_i \log \frac{a_i}{b_i}) \geq (\sum_{i=1}^{n} a_i) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{11}$$

holds, here equality happens if and only if $\frac{a_i}{b_i}$ is a constant.

Setting $a_i = p(x), x \in \psi_j$ and $b_i = \dfrac{p(x)+q(x)}{2}, x \in \psi_j$, we obtain

$$\sum_{x\in\psi_j}(p(x) \log \frac{p(x)}{\frac{p(x)+q(x)}{2}}) \geq (\sum_{x\in\psi_j} p(x)) \log \frac{\sum_{x\in\psi_j} p(x)}{\sum_{x\in\psi_j} \frac{p(x)+q(x)}{2}}. \tag{12}$$

Similarly,

$$\sum_{x\in\psi_j}(q(x) \log \frac{q(x)}{\frac{p(x)+q(x)}{2}}) \geq (\sum_{x\in\psi_j} q(x)) \log \frac{\sum_{x\in\psi_j} q(x)}{\sum_{x\in\psi_j} \frac{p(x)+q(x)}{2}} \tag{13}$$

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

can be obtained. Thus,

$$\begin{aligned} JSD(P_S||Q_S) &= \sum_{j=1}^{M}((\sum_{x\in\psi_j} p(x)) \log \frac{\sum_{x\in\psi_j} p(x)}{\sum_{x\in\psi_j} \frac{p(x)+q(x)}{2}} \\ &+ (\sum_{x\in\psi_j} q(x)) \log \frac{\sum_{x\in\psi_j} q(x)}{\sum_{x\in\psi_j} \frac{p(x)+q(x)}{2}}) \\ &\leq \sum_{j=1}^{M}(\sum_{x\in\psi_j} (p(x) \log \frac{p(x)}{\frac{p(x)+q(x)}{2}}) \\ &+ \sum_{x\in\psi_j} (q(x) \log \frac{q(x)}{\frac{p(x)+q(x)}{2}})) \\ &= JSD(P||Q). \end{aligned} \tag{14}$$

Therefore we have

$$SRJSD(P_S||Q_S) \leq SRJSD(P||Q). \tag{15}$$

□

When the fixation distribution and the optical flow distribution, as shown in (9), are applied to Theorem 1, we obtain

$$VSE = SRJSD(P_f||P_o) \geq SRJSD(P_{fs}||P_{os})$$

and

$$\frac{VSE}{SRJSD(P_{fs}||P_{os})} \leq 1. \tag{16}$$

Therefore, $SRJSD(P_{fs}||P_{os})$ is considered as a normalization factor for *VSE* in this paper.

As discussed in Section III-A, the central part of *CEM* is $-\log_2 VSE$. Inspired by (16), we propose the hypothesis that the cognitive effort measure (*CEM*) takes a scaled form

$$CEM = \frac{-\log_2 VSE}{-\log_2 SRJSD(P_{fs}||P_{os})}. \tag{17}$$

We set *CEM*$= 0$ when $JSD(P_{fs}||P_{os}) = 0$.

Notice that *JSD* between two probability distributions with strictly positive probability mass functions is always lower than 1, and this is proved in Theorem 2.

Theorem *2:* There are two probability distributions $P = \{p(x)|x \in \Psi\}$ and $Q = \{q(x)|x \in \Psi\}$, where $p(x), q(x) > 0$, then we have $JSD(P||Q) < 1$.

Proof.

$$\begin{aligned} &JSD(P||Q) \\ =&\frac{1}{2}(\sum_{\Psi} p(x) \log_2 \frac{2p(x)}{p(x)+q(x)} + \sum_{\Psi} q(x) \log_2 \frac{2q(x)}{p(x)+q(x)}) \\ <&\frac{1}{2}(\sum_{\Psi} p(x) + \sum_{\Psi} q(x)) = 1. \end{aligned}$$

□

Because of the positive property of the two distributions P_{fs} and P_{os} given by (8), according to Theorem 2, we have $\log_2 SRJSD(P_{fs}||P_{os}) \neq 0$. And thus, *CEM* always makes sense.

Fig. 3. A participant is performing the driving task in the *VE*.

IV. Experiment

A. Participants

Fourteen Master/Phd students (5 females; age range: 21-29, Mean = 21.3, SD = 2.37) with driving experience (they hold their driver license at least one and a half years) from Tianjin University volunteer to participate in the psychophysical studies. All of the participants have normal/corrected-to-normal visual acuity and normal color vision. There is no participant having adverse reaction to the virtual environment we set up for the studies.

B. Virtual Reality Environment

In order to deeply understand visuo-motor behavior based on eye movement and driving data in an effective and controllable way, virtual reality is used in this paper, as has been done in many works [22]. A virtual environment (*VE*), including common straight, curved roads and buildings, is designed for the experimental study. In this paper, visual scanning happened for sensorimotor tasks is considered as a task directed behavior, as has been widely done in many relevant works [16] [3]. Therefore, there are no visual stimuli based distractors, such as pedestrians, traffic light signals and similar, are included in the *VE*.

C. Apparatus

HTC Vive headset [23] is used to display the *VE* for participants. The eye-tracking equipment is *7INVENSUN Instrument aGlass DKII* [24], which is embedded into the *HTC Vive* display to capture visual scanning data in a frequency of 90 Hz and in an accuracy of gaze position of $0.5°$. The driving device is a *Logitech G29* steering wheel [25]. Participants listen the ambient traffic and car engine sounds in *VE* by speakers. The visual scanning and driving behaviors of participants are displayed on desktop monitor to observe the experimental procedure.

D. Driving Task

Driving is taken as the sensorimotor task, because it is typically happened in everyday life. As discussed in Section IV-B and, as well as in relevant research [16] [3], the task directed focus on visual scanning and driving is taken. And as a result, participants are required to keep driving at a target speed of 40 km/h, for speed control. This paper takes the inverse of the mean acceleration of vehicle to denote the driving performance. The smaller the mean acceleration, the higher the driving performance becomes, and *vice versa*. In fact, this kind of performance measure has been used a lot in literature [26]. An example of performing driving tasks is presented in Fig. 3.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E. Procedure

Each participant completes four test sessions with the same task requirements and the same driving routes, with an interval of one week between every two sessions. All the sessions for each participant begin with a 9-point calibration for the eye tracker. In this paper, a trial represents a test session, and there are $14 * 4 = 56$ valid trials in all. Data for visual scanning and driving behaviors are recorded during test sessions.

A preparation session is applied to participants before each test session to let them know the purpose and procedure about the psychological studies.

V. Results and Discussions

A. Correlation between VSE and Driving Performance

We verify the hypothesis that the proposed *VSE* takes as an indicator for visual scanning efficiency. Because visual scanning efficiency is usually expected to reflect the performance of sensorimotor tasks [2], correlation analysis based on three kinds of widely used correlation coefficients (*CC*) [27], namely Pearson linear correlation coefficient (*PLCC*), Spearman rank order correlation coefficient (*SROCC*) and Kendall rank order correlation coefficient (*KROCC*), is employed to validate the proposed hypothesis. As listed in Table I, the *CC* results show a statistically significant correlation between our proposed efficiency measure and the task performance. But in contrast, the entropy based measures and eye tracking indices do not correlate with the driving performance. These findings suggest that the new proposed efficiency metric may be a proxy assessment for the driving performance in virtual environment.

An example of eye tracking and driving performance data for two trials, Trials A and B, from the second and fourth test sessions conducted by the 4th and 13th participants, respectively, are presented in Fig. 4. This figure has upper and bottom components, the probability distribution histograms and the fixation-acceleration plot. In the probability distribution histograms of Trials A and B, blue and yellow bars indicate the fixation distributions and optical flow distributions, respectively. Almost all of the differences between the correspondence blue and yellow bars in Trial A are larger than those in Trial B; in particular, for example, this happens for *AOI*s 2 and 5. Note *JSD* is calculated as a sum of the differences between the correspondence probabilities of two distributions. Obviously, in this case, the divergence between the fixation and optical flow distributions from Trial A ($SRJSD$=0.23) is largely bigger than that from Trial B ($SRJSD$=0.01).

TABLE I
CC between indicators and driving performance

Indicators	Pearson		Kendall		Spearman	
	CC	p-value	CC	p-value	CC	p-value
VSE	**0.49**	**p<0.0005**	**0.34**	**p<0.0005**	**0.48**	**p<0.0005**
GTE	0.15	p>0.05	0.04	p > 0.05	0.06	p > 0.05
SGE	0.11	p > 0.05	0.05	p > 0.05	0.08	p > 0.05
EoFS	0.14	p > 0.05	0.10	p >0.05	0.12	p > 0.05
Entropy rate	0.05	p > 0.05	0.05	p > 0.05	0.08	p >0.05
Fixation rate	-0.14	p > 0.05	-0.05	p > 0.05	-0.06	p > 0.05
Saccade amplitude	0.12	p > 0.05	-0.05	p > 0.05	-0.08	p > 0.05

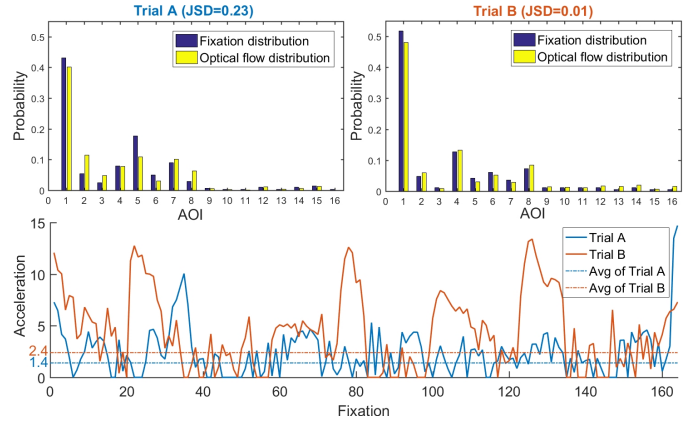

Fig. 4. Illustrations of the fixation and optical flow distributions from Trials A and B, $SRJSD = 0.23$ of Trial A is much higher than $SRJSD = 0.01$ of Trial B (upper). The driving performance of Trial A is largely better than that of Trial B, as shown by their average mean accelerations 1.4 and 2.4, respectively (bottom). The *JSD* visual scanning efficiency corresponds very well to the driving performance.

In the fixation-acceleration plot, the horizontal and vertical axes represent the indices of fixations and the accelerations, respectively. Average accelerations of Trials A and B are 1.4 and 2.4, which are indicated by the blue and orange dashed lines, respectively. Therefore, task performance of Trials A is much better than that of Trial B. This means that, a higher $SRJSD$ corresponds well to a lower acceleration and thus, to a better driving performance, and *vice versa*, validating the effectiveness of the proposed metric of visual scanning efficiency.

In addition, a linear regression analysis is performed for the further confirmation of the correlation, as shown in Fig. 5, where the horizontal and vertical axes represent the proposed *VSE* and the driving performance (inverse of mean acceleration) respectively. The regression result shows a significant linear relationship between the proposed efficiency metric and driving performance.

B. Correlation between CEM and Pupil Size Change

Pupil size change, as a kind of pupillary response, is an autonomic and reflexive indicator of cognitive effort [8] [28]. In this paper, we utilize the standard deviation of pupil size during each trial to represent the pupil size change because of its simplicity and effectiveness [29]. A correlation is revealed between the cognitive effort measure and the pupil size change, as listed in Table II, which validates the hypothesis that the proposed *CEM* can be an indicator for cognitive effort.

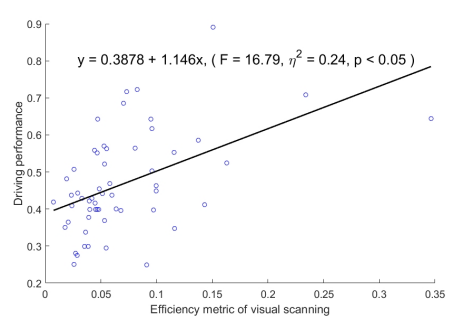

Fig. 5. Linear regression analysis between the *VSE* and driving performance.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLE II
CC BETWEEN CEM AND PUPIL SIZE CHANGE

	Pearson		Kendall		Spearman	
	CC	p-value	CC	p-value	CC	p-value
Pupil size change	0.27	p<0.05	0.20	p < 0.05	0.27	p < 0.05

C. Discussions

This paper discovers a behaviometric based on the exploitation of eye tracking data, for investigating and understanding the efficiency of visual scanning and the cognitive effort, in the procedure of driving. To examine these findings, participants are asked to drive at a target speed in a *3D* virtual environment. The psychophysical results have clearly suggested that the proposed visual scanning efficiency (*VSE*) and cognitive effort measure (*CEM*) can take good proxies for driving task performance and for pupil size change, respectively.

The *PLCC*, *KROCC* and *SROCC* between the proposed *VSE* and driving performance are 0.49, 0.34 and 0.48 respectively, with significant *p-value*s ($p < 0.0005$), as listed in Table I. The success of our proposed metric of visual scanning efficiency benefits from the characterization of how many optical flows (visual motions) are perceived in driving. The central consideration of this is to place the most importance on the perceived visual motions employed for the completion of visual scanning and of driving. That is, visual motion, which always and naturally happens in driving environments [18], is taken as a motivation force for visuo-motor behavior and for driving activity. And, the visual motion perceived for visual scanning during the task of driving indicates the top-down modulation exerted by driver to fulfill the task. As a matter of fact, our idea resides in that the visual scanning is a behavior directed by the driving task, as a combination of top-down and bottom-up influences; this has been widely accepted in literature [3]. As discussed above and, as commonly known in relevant research [1], the visual scanning behavior plays a very important role in performing a sensorimotor task (such as driving), this paper calculates how much correlation there exists between visual scanning efficiency and driving performance, as an evaluation of the proposed *VSE*.

The entropy based measures under comparison do not have significant correlation with task performance, and the direct use of eye tracking indices (fixation rate and saccade amplitude) does not either. Indeed, the entropy based measures express the concept of complexity, which is essentially different from our proposed *VSE*. Concretely, *GTE* is a powerful estimator of the complexity for transition pattern [1] [2], but not of the efficiency for the whole visual scanning procedure. Similarly, *SGE* [13] and *EoFS* [12] give the complexity for spatial patterns of fixations and of fixation sequences, respectively. We believe that the complexity of visuo-motor behavior just plays a marginal role for the efficiency measure of visual scanning in task directed sensorimotor tasks. In fact, visual motion, which is critical for performing sensorimotor tasks like driving, is not explicitly considered by this complexity measurement. And thus this is the reason why the entropy based measures under comparison cannot relate enough to the driving task performance. The direct eye tracking indices, which are generated based on fundamental fixation and saccade, cannot be a proxy for the driving task performance. This is because that these indices are basically simple statistics and only applicable for specific applications, so as not to behave generally and robustly as an estimator of visual scanning efficiency [7].

A significant correlation ($p < 0.05$) is revealed between the proposed *CEM* and the pupil size change by the three correlation analyses, having the *PLCC*, *KROCC* and *SROCC* of 0.27, 0.20, 0.27 respectively, as shown in Table II. This finding verifies the power of the cognitive effort measure, according to the close association between cognitive effort and pupil size change [8] [28]. Considering that we have made progress on the exploitation of eye tracking data, as a behaviometric, for the measurements of visual scanning efficiency and cognitive effort, a further investigation into the relationship between these two proposed measures will be performed in future work. And actually, this could be a working path illuminated based on the exploitation of Yerkes-Dodson law [30].

In this paper, visual scanning in sensorimotor tasks is mainly considered as a task directed behavior, as has been largely accepted in literature [16] [3]. And in this case, two corresponding treatments are performed in our psychophysical studies. Firstly, we especially design the virtual environment not to include visual stimuli based distractors (such as pedestrians, traffic light signals and similar). The purpose of this is to use the visual motion, which is resulted from the dynamic nature of driving environment, as the single bottom-up contribution. Secondly, we simply ask participants to travel at a fixed speed as the requirement of driving task. The objective of this is to take the speed control as the single contributor of top-down modulation. As a result of these two treatments, mean acceleration is taken as the driving task performance. Note that mean acceleration is a common measure of driving performance in related research [26]. Obviously, all in all, we emphasize that the proposed *VSE* and *CEM* are investigated and assessed more easily and more clearly in the current two treatments. Complex virtual environment and more requirements of the driving task will be our future work.

Notice that the findings of this paper may not be applicable for all cases, but it does work in the context of our topic. Due to that visual scanning is exceptional important in virtual and real-world sensorimotor tasks, what we have achieved on efficiency metric and on cognitive effort measure in virtual driving should be potentially helpful for ergonomic evaluation pragmatically, in many practical and relevant applications.

VI. Conclusions and Future Works

In this paper, we take an important step for thorough understanding the mechanism of visual scanning in virtual driving. This paper has established, in an objective and quantitative way, a new metric for visual scanning efficiency through a methodology that exploits the perceived visual motions in a

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.17.386185; this version posted April 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sensorimotor task by the divergence between fixation distribution and optical flow distribution. The efficiency measure provides a window into how to evaluate the performance of visual scanning, and also behaves as a strong proxy for objectively assessing the performance of driving tasks. As far as we can know, no research up to now has reported this kind of finding to shed light on the issue of efficiency metric for visual scanning and for performing sensorimotor tasks. Additionally, the proposed cognitive effort measure, which is significantly correlates with pupil size change, may offer a new perspective on the inherent relationship between task directed visual scanning and eye tracking data. The two measures we have proposed have been verified by the psychophysical studies conducted in this paper. In summary, this paper proposes a new methodology to measure visuomotor behavior by using eye tracking data, so as to help the development of behaviometric discovery, from both theoretical and practical perspectives.

In the near future, we will investigate our proposed methodology and measures for real-life driving scenarios, for instance, for crash risk problem [31]. In consideration of the critical role of illumination conditions for driving, we will exploit the manipulation of illumination levels in a detailed quantitative way, to comprehensively understand the mechanism of visual scanning efficiency. The association between the two proposed measures will be studied, by the exploitation of Yerkes-Dodson law [30], as has been mentioned in Section V-C. Also, physiological signals such as heartbeat [32] will be investigated for understanding the relationships and interplays between these signals and eye tracking data, for the sake of visual scanning efficiency and cognitive effort.

References

[1] B. Shiferaw, L. A. Downey, J. Westlake, B. Stevens, S. M. W. Rajaratnam, D. J. Berlowitz, P. Swann, and M. E. Howard, "Stationary gaze entropy predicts lane departure events in sleep-deprived drivers," *Scientific Reports*, vol. 8, no. 1, pp. 2220–2220, 2018.

[2] B. Shiferaw, L. Downey, and D. Crewther, "A review of gaze entropy as a measure of visual scanning efficiency," *Neuroscience & Biobehavioral Reviews*, vol. 96, pp. 353–366, 2019.

[3] J. Kurz, M. Hegele, and J. Munzert, "Gaze behavior in a natural environment with a task-relevant distractor: How the presence of a goalkeeper distracts the penalty taker," *Frontiers in Psychology*, vol. 9, 2018.

[4] A. C. Hayley, B. Shiferaw, and L. A. Downey, "Amphetamine-induced alteration to gaze parameters: A novel conceptual pathway and implications for naturalistic behavior," *Progress in Neurobiology*, 2020, In Press.

[5] A. Constant, A. Clark, M. Kirchhoff, and K. J. Friston, "Extended active inference: constructing predictive cognition beyond skulls," *Mind and Language*, 2019.

[6] J. Lin, "Divergence measures based on the shannon entropy," *IEEE Transactions on Information Theory*, vol. 37, no. 1, pp. 145–151, 1991.

[7] H. Jeong, Z. Kang, and Y. Liu, "Driver glance behaviors and scanning patterns: Applying static and dynamic glance measures to the analysis of curve driving with secondary tasks," *Human Factors and Ergonomics in Manufacturing & Service Industries*, vol. 29, no. 6, pp. 437–446, 2019.

[8] M. K. Eckstein, B. Guerra-Carrillo, A. T. M. Singley, and S. A. Bunge, "Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development," *Developmental cognitive neuroscience*, vol. 25, pp. 69–91, 2017.

[9] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[10] Y. Itoh, Y. Hayashi, I. Tsukui, and S. Saito, "The ergonomic evaluation of eye movement and mental workload in aircraft pilots," *Ergonomics*, vol. 33, no. 6, pp. 719–732, 1990.

[11] S. S. M. Chanijani, P. Klein, S. S. Bukhari, J. Kuhn, and A. Dengel, "Entropy based transition analysis of eye movement on physics representational competence," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 2016, pp. 1027–1034.

[12] I. T. Hooge and G. Camps, "Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers," *Frontiers in Psychology*, vol. 4, p. 996, 2013.

[13] K. Krejtz, A. Duchowski, T. Szmidt, I. Krejtz, F. González Perilli, A. Pires, A. Vilaro, and N. Villalobos, "Gaze transition entropy," *ACM Transactions on Applied Perception (TAP)*, vol. 13, no. 1, pp. 1–20, 2015.

[14] J. Allsop, R. Gray, H. H. Bulthoff, and L. Chuang, "Effects of anxiety and cognitive load on instrument scanning behavior in a flight simulation," in *2016 IEEE Second Workshop on Eye Tracking and Visualization (ETVIS)*. IEEE, 2016, pp. 55–59.

[15] K. Kaspar, T.-M. Hloucal, J. Kriz, S. Canzler, R. R. Gameiro, V. Krapp, and P. König, "Emotions' impact on viewing behavior under natural conditions," *PloS one*, vol. 8, no. 1, 2013.

[16] O. Lappi, "Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference," *Neuroscience & Biobehavioral Reviews*, vol. 69, pp. 49–68, 2016.

[17] S. Negahdaripour, "Revised definition of optical flow: integration of radiometric and geometric cues for dynamic scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 9, pp. 961–979, 1998.

[18] Y. Okafuji, C. Mole, N. Merat, T. Fukao, Y. Yokokohji, H. Inou, and R. M. Wilkie, "Steering bends and changing lanes: the impact of optic flow and road edges on two point steering control," *Journal of Vision*, vol. 18, no. 9, pp. 14–14, 2018.

[19] D. M. Endres and J. E. Schindelin, "A new metric for probability distributions," *IEEE Transactions on Information Theory It*, 2003.

[20] K. Holmqvist, M. Nystrom, R. Andersson, R. Dewhurst, H. Jarodzka, and J. V. Weijer, *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2011.

[21] J. Y. Chew, K. Ohtomi, and H. Suzuki, "Glance behavior as design indices of in-vehicle visual support system: A study using crane simulators," *Applied ergonomics*, vol. 73, pp. 183–193, 2018.

[22] L. Sidenmark and H. Gellersen, "Eye, head and torso coordination during gaze shifts in virtual reality," *ACM Transactions on Computer-Human Interaction*, vol. 27, no. 1, pp. 1–40, 2019.

[23] "Htc corporation," https://www.htcvive.com.

[24] "7invensun instrument aglass," https://www.7invensun.com/xnxsxqy.

[25] "Logitech g29," https://www.logitechg.com/en-us/products/driving/driving-force-racing-wheel.html.

[26] A. K. Yadav and N. R. Velaga, "Effect of alcohol use on accelerating and braking behaviors of drivers," *Traffic Injury Prevention*, vol. 20, no. 4, pp. 353–358, 2019.

[27] "Correlation and dependence," https://en.wikipedia.org/wiki/Correlation and dependence.

[28] S. Joshi, Y. Li, R. M. Kalwani, and J. I. Gold, "Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex," *Neuron*, vol. 89, no. 1, pp. 221–234, 2016.

[29] S. Chen, J. Epps, N. Ruiz, and F. Chen, "Eye activity as a measure of human mental effort in hci," *intelligent user interfaces*, pp. 315–318, 2011.

[30] P. A. Watters, F. Martin, and Z. Schreter, "Caffeine and cognitive performance: The nonlinear yerkes-dodson law," *Human Psychopharmacology-clinical and Experimental*, vol. 12, no. 3, pp. 249–257, 1997.

[31] B. J. Victor T, Dozza M, "Analysis of naturalistic driving study data: Safer glances, driver inattention, and crash risk," Tech. Rep., 2015.

[32] A. Galvez-Pol, R. Mcconnell, and J. M. Kilner, "Active sampling in visual search is coupled to the cardiac cycle," *Cognition*, vol. 196, 2020.