Detecting expert's eye using a multiple-kernel Relevance Vector Machine

Journal of Eye Movement Research
7(2):3, 1-15

Detecting expert's eye using a multiple-kernel Relevance Vector Machine

Giuseppe Boccignone
Dip. Informatica, Università di Milano, Italy

Mario Ferraro
Dip. Fisica, Università di Torino, Italy

Sofia Crespi
LAPCO - Università Vita-Salute San Raffaele, Italy
Neuroradiology Unit & CERMAC, Ospedale San Raffaele, Italy

Carlo Robino
LAPCO - Università Vita-Salute San Raffaele, Italy
2Easy s.r.l., Milano, Italy

Claudio de'Sperati
LAPCO - Università Vita-Salute San Raffaele, Italy

Decoding mental states from the pattern of neural activity or overt behavior is an intensely pursued goal. Here we applied machine learning to detect expertise from the oculomotor behavior of novice and expert billiard players during free viewing of a filmed billiard match with no specific task, and in a dynamic trajectory prediction task involving ad-hoc, occluded billiard shots. We have adopted a ground framework for feature space fusion and a Bayesian sparse classifier, namely, a Relevance Vector Machine. By testing different combinations of simple oculomotor features (gaze shifts amplitude and direction, and fixation duration), we could classify on an individual basis which group - novice or expert - the observers belonged to with an accuracy of 82% and 87%, respectively for the match and the shots. These results provide evidence that, at least in the particular domain of billiard sport, a signature of expertise is hidden in very basic aspects of oculomotor behavior, and that expertise can be detected at the individual level both with ad-hoc testing conditions and under naturalistic conditions - and suitable data mining. Our procedure paves the way for the development of a test for the "expert's eye", and promotes the use of eye movements as an additional signal source in Brain-Computer-Interface (BCI) systems.

Keywords: eye movements, expertise, billiards, mind reading, machine learning, feature fusion, relevance vector machine

Introduction

Eye movements can be a useful source of information to infer cognitive processes (Buswell, 1935; Yarbus, 1967; Rayner, 1998; Viviani, 1990; Henderson, 2003). Among the various top-down factors that guide our gaze, expertise plays a prominent role, and can effectively drive the ocular exploratory behavior. The scanpath, i.e., the sequence of saccades and fixations, (Noton & Stark, 1971), of expert and novice observers differs when they look at pictures or art pieces (Nodine, Locher, & Krupinski, 1993; Zangemeister, Sherman, & Stark, 1995; Vogt & Magnussen, 2007; Humphrey & Underwood, 2009; Pihko et al., 2011), interpret medical images (Nodine, Kundel, Lauver, & Toto, 1996; Donovan & Manning, 2007), drive (Underwood, 1998), read music (Waters, Underwood, & Findlay, 1997), play chess (Chase & Simon, 1973; Reingold & Sheridan, 2011), practice or watch sports (Vickers, 2007). Thus, from the characteristics of eye movements it is possible to extrapolate important information about expertise in several knowledge and activity domains.

We have recently provided evidence that the eye movements of novice and expert billiard players differ when they have to predict the outcome of partially-occluded single shots (Crespi, Robino, Silva, & de'Sperati, 2012). Specifically, in order to solve the visual prediction task, novices tended to adopt a strategy based on mental extrapolation of the ball trajectory, whereas experts monitored certain diagnostic points along the trajectory. By exploiting the eye movements differences of novices and experts, we could also identify the temporal boundaries of the single billiard shots contained in a videoclip, thus in fact realizing a sort of physiologically-based video parser (Robino, Crespi,

This work received partial financial support from ERC (Stanib program) and the Italian Ministry of Instruction, University and Research (grant PRIN 2008RBFNLH 002 to CdS)

Silva, & de'Sperati, 2012).

In the present study we extend our previous work and ask whether the differences in eye movements of novices and experts are robust enough to detect expertise i) at the individual level, and ii) under not only ad-hoc, controlled conditions but also naturalistic, unconstrained conditions i.e., during free viewing of a billiard match without a specific task. Also, iii) we aim to detect the "expert's eye" by analyzing the data regardless of the visual stimulus, that is, relying only on the oculomotor behaviour. Meeting these three conditions would be an important step towards automatic expertise detection.

Quantifying reliably and uniquely a complex behavior such as a sequence of exploratory eye movements (the so-called scanpath) is a non-trivial challenge. The existing methods can be classified into two broad classes, both pioneered by Larry Stark (see Hacisalihzade, Stark, & Allen, 1992, for a combined use of both). The first approach aims at characterizing the spatial distribution of fixations on the scene (spatio-temporal, in case of dynamic scenes) and to provide some similarity metrics (Brandt & Stark, 1997). Methods following this approach can be further distinguished as *content-driven* or *data-driven* (Grindinger et al., 2011).

The *content-driven* approach largely relies upon Regions Of Interest (ROIs), identified a priori in the stimulus and analyzed in terms of fixations falling inside them. The *data-driven* approach, in contrast, directly exploits scanpaths, or features extracted from them, independent of whatever was presented as the stimulus. An important advantage of the latter approach is that it obviates the need of arbitrary ROI definition.

The similarity of two scanpaths can be measured in principle by using ROI-based methods followed by coding of the sequence in which ROIs are visually inspected. A common method is the string edit, in which a string is defined by assigning each ROI a discrete symbol (e.g., a character), so that each scanpath is transformed in a string of symbols. Then the editing cost of transforming one string into another one is computed (e.g., by computing the Levenshtein distance, which measures the editing cost of transforming one string into another one, Brandt & Stark, 1997; Choi, Mosley, & Stark, 1995; Hacisalihzade et al., 1992; Foulsham & Underwood, 2008). Other methods are also used, such as the Needleman- Wunsch algorithm borrowed from bioinformatics (Cristino, Mathôt, Theeuwes, & Gilchrist, 2010). However, ROI based method suffer from well-known limitations, mostly related to how to cluster and regionalize fixations (Hacisalihzade et al., 1992; Privitera & Stark, 2000). For instance many methods rely upon dividing the image into a regular grid, but this way of operating loses any reference to the content of the image, and introduces quantization errors; in this limit case string edit techniques turns into a *data-driven* approach, while exploiting an oversimplified representation of data. Semantic ROIs could be used instead (Privitera & Stark, 2000; Josephson & Holmes, 2006), but these have by definition different sizes, and therefore the approximation of fixation position can be very coarse and subtle differences in oculo-motor behavior cancelled. In the last few years, heatmaps have become a very popular, *data-driven* , tool: heatmaps are plots in which a given oculomotor quantity (typically, the fixation dwell-time) is coded as colored, semi-transparent "bubbles" superimposed to the bi-dimensional image. This graphical representation is very appealing, but it is mostly used to convey an immediate, qualitative impression of the attended regions within a figure (see, however, Caldara & Miellet, 2011; Crespi et al., 2012). Other methods have also been proposed, based on the construction of an average scanpath (Hembrooke, Feusner, & Gay, 2006), or that minimize an energy function (Dempere-Marco, Hu, Ellis, Hansell, & Yang, 2006), or that end up with a multidimensional vector rather than a single scalar quantity (Jarodzka, Holmqvist, & Nyström, 2010). A main concern of these approaches is to quantify the similarity between scanpaths, which is a crucial issue in certain applications where an average observer is needed (Boccignone et al., 2008).

The second approach, again pioneered by Stark, takes straightforwardly into account the very stochastic nature of scanpaths. Indeed, gaze-shift processes, and especially saccadic eye movements, exhibit noisy, idiosyncratic variation of visual exploration by different observers viewing the same scene, or even by the same subject along different trials; this is a well-known issue debated since the early eye tracking studies by Ellis and Stark (1986), who modeled sequences using Markov transition probability matrices identified from experimental sequences (see Hayes, Petrov, & Sederberg, 2011 for a detailed discussion on methods aiming at capturing statistical regularities in temporally extended eye movement sequences). Here we follow this second approach or, more precisely, the very rationale behind such approach: namely, we consider the gaze shift behavior as a realization of a stochastic process (Feng, 2006; Brockmann & Geisel, 2000; Boccignone & Ferraro, 2014, 2013b, 2013a). In other terms, the distribution functions and the temporal dynamics of eye movements are specified by the stochastic process. In this perspective the visual exploratory features we can measure (saccade amplitude and direction, fixation duration) can be thought of as random variables generated by such a process, however complex it may be (Tatler & Vincent, 2008, 2009).

In order to discriminate between different oculomotor behavior exhibited by novices and experts, there are two options: to provide a model for the generating process, or to exploit the generated oculomotor pattern. For what concerns the first option, investigating expertise differences in dynamic tasks, such as a billiard match, is a complex modeling issue, and involves

aspects far beyond the limits of current computational models (Borji & Itti, 2013). The second option, i.e., analyzing the generated oculomotor pattern, relies upon the rationale that the key requirements of expertise are discriminability and consistency across different stimuli (Shanteau, Weiss, Thomas, & Pounds, 2002), properties that should be reflected in the generated pattern.

In the present study we applied machine learning techniques to discriminate eye movements of experts and novices at the individual level. Among machine learning techniques, the Support Vector Machine (SVM, Cristianini & Shawe-Taylor, 2000) is widely used to classify noisy signals (see Murphy, 2012 for a general discussion), including eye movement data (Lagun, Manzanares, Zola, Buffalo, & Agichtein, 2011; Eivazi & Bednarik, 2011; Bednarik, Kinnunen, Mihaila, & Fränti, 2005; Vig, Dorr, & Barth, 2009). Methods simpler than SVM have also been used to classify eye movements (e.g., Henderson, 2003).

Specifically, in this study we have tried to deal with two problems. First, machine learning approaches as usually applied to the analysis of eye-movements tend to overlook the feature representation problem. In order to spot behavioral characteristics - expertise or cognitive impairments - in a *data-driven* way, a scanpath can be analyzed by using several features (e.g., Lagun et al., 2011). Each feature, in turn, might be differently related to a number of factors, from low-level biomechanics, to learnt knowledge of the structure of the world and the distribution of objects of interest (Tatler & Vincent, 2009). Thus, within a machine learning perspective, we are dealing with features from different sources and where there may be limited or no a priori knowledge of their significance and contribution to the classification task. Clearly, concatenating all the features into a single feature space does not guarantee an optimum performance, while facing the "curse of dimensionality" problem.

Second, though SVM methodology has proven to be a powerful one, it has a number of well-known limitations (Tipping, 2001; Murphy, 2012). Although relatively sparse, SVMs make unnecessarily liberal use of basis functions since the number of support vectors required typically grows linearly with the size of the training set; predictions are not probabilistic, which is particularly crucial in classification where posterior probabilities of class membership are necessary to adapt to varying class priors and asymmetric misclassification cost; the kernel function must satisfy Mercer's condition, namely, it must be the continuous symmetric kernel of a positive integral operator.

In order to cope with these problems, we have exploited a ground framework for feature space fusion followed by a Bayesian sparse classification technique (Tipping, 2001) with the ability of achieving sparse solutions that utilize only a subset of the basis functions. In particular, we have considered the basic oculomotor parameters of saccade amplitude, direction, and fixation duration as different information sources that are combined within a composite kernel space level and classified through a Relevance Vector Machine (RVM), namely a multiple-kernel RVM (mRVM, (Psorakis, Damoulas, & Girolami, 2010; Damoulas & Girolami, 2009a)). See Appendix A, for a detailed discussion of the RVM approach and its main differences with respect to SVMs.

To the best of our knowledge this approach has never been used with eye movement data.

Materials and Methods

The present analyses were performed on raw data acquired in the course of previous experiments (Crespi et al., 2012; Robino et al., 2012). The reader is referred to that work for details concerning stimuli and data acquisition.

Participants

Forty-two healthy participants volunteered for the experiment (all men but one, with normal or corrected-to-normal vision, aged between 27 and 70 years, naïve as to the purpose of the experiment). Half of them were elite billiard players, recruited on the basis of their national ranking, whereas the other half had no or occasional experience in billiard playing. The study was conducted in accordance with the recommendations of the Declaration of Helsinki and the local Ethical Committee. Before starting the experiments, all participants signed the informed consent.

Stimuli and procedure

The stimuli were movies of a billiard match or of individual shots, recorded from the top of the billiard table. The stimuli were subsequently presented on a computer screen. Whereas the former stimulus represented a real match without any experimental constraint, the shots were prepared by asking a professional player to execute a number of ad-hoc shots.

Match. This stimulus typology consisted of a piece of a billiard match (M), in which two professional players (the opponents) alternated in launching with the stick the cue ball (own ball) towards the target ball (opponent's ball) in such a way that the latter - but not the former - would knock down as many skittles as possible (there were 5 skittles in the central region of the table) and/or touch a third ball (a small red ball). The movie lasted 5 minutes and contained 11 shots, alternating naturally between the two opponents. The shots were obviously different for complexity, orientation, number of cushions, duration, ball velocity, and spin. The billiard match was always presented first.

Shots. The other stimulus typology consisted of 24 different shots with no spin, ultimately directed towards the central skittle. The shots were either short (2 cushions, SS) or long (5 cushions, LS). The initial direction of the shot (immediately after the contact with the stick) was either towards the right or the left, or towards the upper or the lower side of the table, in a balanced design. There were three versions of the shots, in one version the central skittle was knocked down, in the other two versions the ball passed just beside the skittle, to the right or to the left. In each shot, the final part of the trajectory was occluded 200 ms after the ball bounced on the second (SS) or the third (LS) cushion, because the observers' task was to tell whether or not the ball would strike the skittle (see below). There were 2 repetitions for each shot, for a total of 48 stimuli, presented in a pseudo-random sequence. The duration was 15 minutes. The shot trajectories, including the occluded portion, are illustrated in Figure 2.

Procedure. Observers watched the stimuli while seating about 57 cm in front of the computer screen, with the head resting on a forehead support. For the match stimulus, the observers were simply instructed to pay attention to the movie in order to answer to some general question afterwards. For the shots stimulus, their task was instead to predict, with a verbal response for each trial, whether or not the ball would strike the skittle. Eye movements were acquired through infrared video-oculography (Eyegaze System, LC Technologies; sampling frequency: 60 Hz; nominal precision: 0.18 deg). Monocular recordings were performed unobtrusively via a remote camera mounted below the computer screen. Gaze direction was determined by means of the pupil-center-corneal reflection method

Data Analyses

Ocular fixations were identified by means of a dispersion criterion: We defined gaze samples as belonging to a fixation if they were located within an area of 25 pixels (corresponding to 0.67 deg) for a minimum duration of 6 video frames (corresponding to 100 ms). Gaze shifts were defined as the transition from one fixation to the next.

The problem of distinguishing billiard experts from novice observers, by assessing their oculomotor behavior, can be recasted as a classification procedure in a supervised learning setting. A feature set should be defined in order to capture the oculomotor behavior of the observers. To this end, for each observer, given the sequence of fixations $\{\mathbf{r}_t\}_{t=1}^{N_T}$, where the vector $\mathbf{r}_t$ represents the fixation position (coordinates) at time t, we computed the amplitude and direction of each gaze shift $\{l_t, \theta_t\}_{t=1}^{N_T}$, where l_t is defined as the Euclidean distance between two successive fixations, and $\theta_t = \tan^{-1}\frac{\Delta y_t}{\Delta x_t}$ the direction of the gaze shift between successive fixations, $\Delta x_t, \Delta y_t$ being the horizontal and vertical components. These two features are good descriptors of the exploratory oculomotor activity (Tatler & Vincent, 2008, 2009; Boccignone & Ferraro, 2013b, 2013a). As a third feature we used the fixation duration $\{f_t\}_{t=1}^{N_T}$, which is also a useful descriptor of the oculomotor behavior in terms of visual processing (Viviani, 1990).

Because we assume that the scanpath is the result of an underlying stochastic process (Boccignone & Ferraro, 2014), we summarize the random sample $\{l_t, \theta_t, f_n\}_{t=1}^{N_T}$ through the empirical distribution functions (histograms), which we denote as the random vectors $\mathbf{x}^l = \left[x_1^l \cdots x_D^l\right]^T$, $\mathbf{x}^\theta = \left[x_1^\theta \cdots x_D^\theta\right]^T$ and $\mathbf{x}^f = \left[x_1^f \cdots x_D^f\right]^T$, respectively, where the vector dimension D represents the number of bins of the histogram. In the following analyses $D = 6$ is used. The feature vector $\mathbf{x}^s$ is thus a summary of the behavior of a single observer with respect to a particular feature space or source of information $s = 1, \ldots S$, here $S = 3$.

In conclusion, each observer n, $n = 1, \cdots, N$ is represented in the dataset $\{\mathbf{X}, \mathbf{t}\}$, where the matrix $\mathbf{X}$ is the collection of features from all N observers, whose behaviour is summarized by the three feature vectors of dimension D, $\mathbf{x}_n^s \in \mathbb{R}^D$, $s = 1, \cdots, 3$. The target vector $\mathbf{t} = [t_1 \cdots, t_N]^T$ denotes the collection of random variables t_n, taking values (labels) in $\mathcal{C}$, a classification space of dimension C. In our case, $\mathcal{C} = \{expert, novice\}$, thus $C = 2$ (binary classification). Then, the posterior probability for observer n to be classified as expert or novice will be $P(t_n | \mathbf{x}_n^1, ..., \mathbf{x}_n^S)$ and according to Bayesian decision theory we would assign the observer n to the class that has the maximum a posteriori probability (MAP).

From a pattern recognition perspective, one could in principle use different classifiers trained on the different feature spaces, but classifier combination methodologies (product combination rule, mean combination rule, etc.) then would require specific assumptions such as independence of the feature spaces or, on the opposite, extreme correlation. Here we adopt the strategy of combining the feature spaces, and, in particular, we exploit the composite kernel construction technique (Damoulas, Ying, Girolami, & Campbell, 2008; Damoulas & Girolami, 2009a, 2009b), which is summarized at a glance in Figure 1.

First, the individual feature vectors were mapped into kernels (the *kernel trick*, Murphy, 2012) and thus embedded in Hilbert spaces via base kernels, that can be represented as the matrix $\mathbf{K}^s \in \mathbb{R}^{N \times N}$. Each element of $\mathbf{K}^s$ can be constructed through a suitable kernel function, which can be chosen based on prior knowledge, cross-validation or even via inference from a pool of kernel functions. Different choices are possible for the

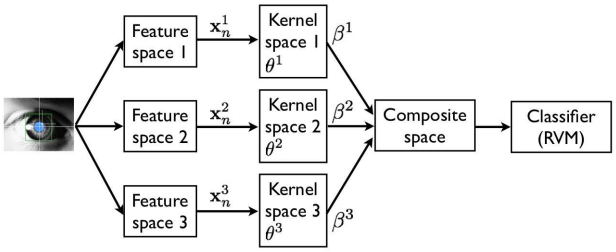


Figure 1. Data analysis in multiple-kernel representation. The fixation sequence is represented in different feature spaces $s = 1, \cdots, S$; each feature $\mathbf{x}^s$ is then separately mapped in a kernel space, each space being generated via kernel K^s of parameters θ^s. The separate kernel spaces are then combined in a composite kernel space, which is eventually used for classification

kernel functions, among which the most used are:

$$K^s(\mathbf{x}_i^s, \mathbf{x}_j^s) = \mathbf{x}_i^{sT}\mathbf{x}_j^s, \qquad K^s(\mathbf{x}_i^s, \mathbf{x}_j^s) = \exp\left(-\frac{\|\mathbf{x}_i^s - \mathbf{x}_j^s)\|^2}{2\rho^2}\right),$$

namely the linear and Gaussian kernel respectively.

In turn, base kernels can be combined into a composite kernel $\mathbf{K}^b \in \mathbb{R}^{N\times N}$ whose elements are:

$$K^\beta(\mathbf{x}_i, \mathbf{x}_j) = \sum_{s=1}^{S} \beta^s K^s(\mathbf{x}_i^s, \mathbf{x}_j^s) \tag{1}$$

This way, the composite kernel is a weighted summation of the base kernels with β^s as the corresponding weight for each one. Also, notice that in a multiple kernel setting we are free to choose different kernels for constructing the individual kernel spaces. As far as we employ at least two different feature spaces, even when the same kernel shape (e.g., Gaussian) is adopted for both spaces (cfr., Figure 1), nevertheless the multiple kernel learning (MKL) procedure permits to adapt individual kernel parameters so to capture the statistics of information source s as represented in the corresponding feature space (*data-driven* approach).

The detection of expertise in the eye movements of the n-th subject in terms of maximum a posteriori $P(t_n|\mathbf{x}_n^1, ..., \mathbf{x}_n^S)$, can be obtained at the most general level as:

$$P(t_n|\mathbf{x}_n^1, ..., \mathbf{x}_n^S) = P(t_n|\mathbf{W}, \mathbf{k}_n^\beta), \tag{2}$$

where the term on the r.h.s. is the Multinomial probit likelihood for the calculation of class membership probabilities (see Appendix A for a discussion and Damoulas & Girolami, 2009a; Psorakis et al., 2010 for further details). In Eq. 2. In the same equation, $\mathbf{W} \in \mathbb{R}^{N\times C}$ is the matrix of model parameters; the variable $\mathbf{k}_n^\beta$ is a row of the kernel matrix $\mathbf{K}^\beta \in \mathbb{R}^{N\times N}$ - whose elements are the $K^\beta(\mathbf{x}_i, \mathbf{x}_j)$ defined in Eq. 1 - and it expresses how related, based on the selected kernel function, observation $\mathbf{x}_n$ is to the others of the training set (Appendix A). Given the posterior $P(t_n|\mathbf{x}_n^1, ..., \mathbf{x}_n^S)$, classification $t_n = c$, $c \in \mathcal{C}$ is attained by using the MAP rule:

$$c = \arg\max_{t_n} P(t_n|\mathbf{x}_n^1, ..., \mathbf{x}_n^S). \tag{3}$$

The Multinomial probit likelihood $P(t_n|\mathbf{W}, \mathbf{k}_n^\beta)$ in Eq. 2 above can be computed provided that the parameters $\mathbf{W}, \mathbf{k}_n^\beta$ are known. In a Bayesian framework, the latter can be inferred (learned) from data by introducing a prior distribution for the regression parameters $\mathbf{W}$ (cfr. Appendix A), and to such end one suitable methodology is the Relevance Vector Machine (RVM, Tipping, 2001) framework in the variant proposed in (Psorakis et al., 2010). RVMs can be considered the Bayesian counterpart of SVMs. They are Bayesian sparse machines, that is they employ sparse Bayesian learning via an appropriate prior formulation. Not only do they overcome some of the limitations affecting SVMs (Appendix A), but also they achieve sparser solutions (and hence they are faster at test time) than SVM (Tipping, 2001; Murphy, 2012). In particular, we have exploited the multi class RVM (precisely, m-RVM1, Psorakis et al., 2010). Clearly, in our case the multi-class capability of the m-RVM1 (Psorakis et al., 2010) is redundant, since we are dealing with a binary classification problem ($C = 2$). However, essential in our case is the ability of achieving sparse solutions that utilize only a subset of the basis functions, the relevance vectors (Murphy, 2012), together with a ground framework for feature space fusion (Damoulas & Girolami, 2009a).

To sum up, the train and test procedure adopted has been the following. We have exploited a leave-one-out approach, where, for all observers, at each step, $N-1$ observers are enrolled for the training set and the N-th observer is used as one sample of the test set (Murphy, 2012) to be classified as in Eq. 3.

The input to the train and test procedure has been shaped in the form of all possible combinations of the feature vectors (histograms) $\{\mathbf{x}^s\}_{s=1}^S$ (single features, pairs, or the full set, see the Supplementary Table). Further, given the input, all possible mappings using either the linear and/or the Gaussian kernel have been considered. Since the Gaussian kernel has a free parameter, the scale ρ, at each learning step a 5-fold cross validation procedure was accomplished for tuning such parameter; validation has been performed by varying the scale parameter in the range $\rho \in \left[2^{-15}, \cdots, 2^3\right]$. Such interval has been discretised using a sampling step $\delta = 0.5$. The learning and classification steps accomplished in the leave-one-out schedule (see Appendix A for a general description) have been performed by using the MATLAB software implementation of the m-RVM1 available at `http://www.dcs.gla.ac.uk/inference/pMKL`, with standard parameter initialization.

In the following Section, results reported have been obtained after 5 classification runs for each kernel and

feature configuration taken into account, each run exploiting the leave-one-out procedure described above. At the beginning of each run the input data were randomly shuffled.

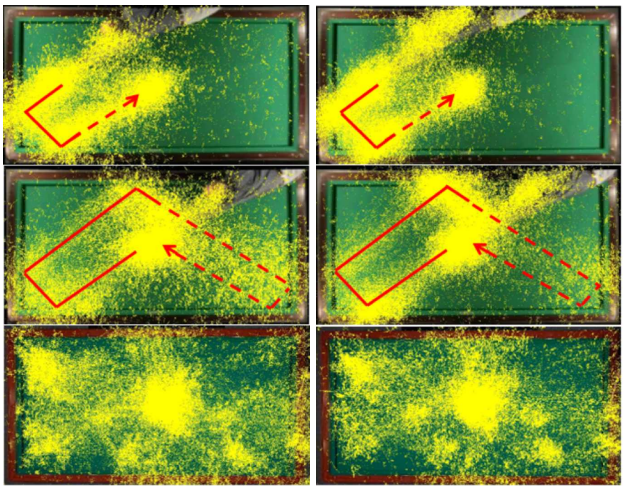
Results

Expert and novice observers exhibited rather similar exploratory eye movements when watching a given stimulus - at least this is the qualitative impression when observing the cumulative gaze position over time condensed in single snapshots (Figure 2). Examples of individual scanpaths are illustrated in Figure 3. Here too, as in the pooled data of Figure 2, a certain degree of similarity between experts and novices can be appreciated at visual inspection. For example, in the single shots the ball trajectories can be often glimpsed from the raw scanpaths. We quantified the scanpaths by means of three oculomotor features, namely, fixation duration, gaze shift amplitude and gaze shift direction, which were used as input to the classifier either as single features or concatenated in pairs or in a triplet.

The distributions of these basic oculomotor features looked very similar between experts and novices (Figure 4), with very close median values (fixation duration - novices vs. experts: 247 vs. 231 ms, 231 vs. 215 ms, 247 vs. 230 ms, respectively for SS, LS and Match; gaze shift amplitude - novices vs. experts: 2.219 vs. 2.458 deg, 2.383 vs. 2.525 deg, 2.076 vs. 2.150 deg, respectively for SS, LS and Match). Also the shapes of the gaze shift direction distributions looked rather similar (polar plots in Figure 4). Despite this apparent similarity, however, in all cases there were statistically significant differences between experts' and novices' distributions (2-samples Kolmogorov-Smirnov test for fixation duration and gaze shift amplitude, always $p < 0.01$; 2-samples Kuiper test for gaze shift direction, always $p < 0.01$). Indeed, across the 3 shots experts had on average slightly shorter fixations (-16 ms), and somewhat larger and more counterclockwise-rotated gaze shifts ($+0.15$ deg and $+0.336$ rad).

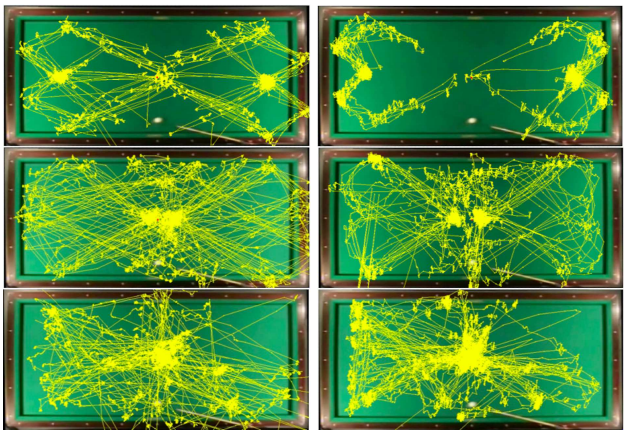
Such small differences, however, can be exploited to discriminate between novices and experts when raw features are processed by a suitable classifier. For this purposes a RVM has been chosen as classifier. We first used equal kernel functions (linear and Gaussian) for all feature channels (cfr., Figure 1), while taking into consideration different numbers of sources/feature spaces s. Analysis of the results showed that classifier performances for the features $\mathbf{x}^{\theta}$ derived from saccadic directions were worse in case of the Gaussian classifier: that lead us to use mixed functions kernels, namely a Gaussian kernel for the length of shifts and fixation times, and a linear one for directions.

The outcomes obtained from the different kernels were quite similar, as can be seen in Supplementary Table 1. Therefore, the following analysis is performed solely on the results obtained with the multiple kernel approach, because it is a more flexible and novel than single kernel methods. Moreover, except for the case of short shots, it was the only approach where the best performance was attained with more than one feature



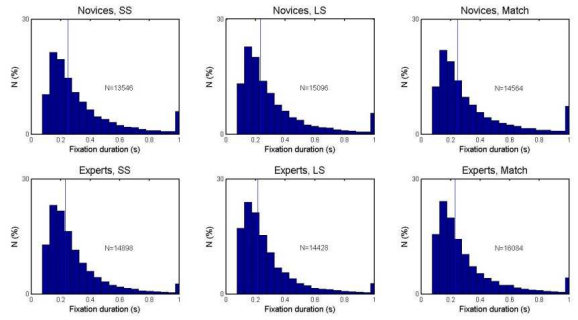
(a) Novices (b) Experts

Figure 2. Raw eye position (yellow) recorded during shot and match viewing, for both novices (2(a)) and experts (2(b)). All data from all participants are superimposed. The traces recorded during shot viewing have been re-oriented onto a single shot trajectory (red arrows, with the dashed part representing the occluded portion of the trajectory) for the clarity of the graphical presentation. Rows, from top to bottom: Short shots (SS), Long shots (LS), Match (M).

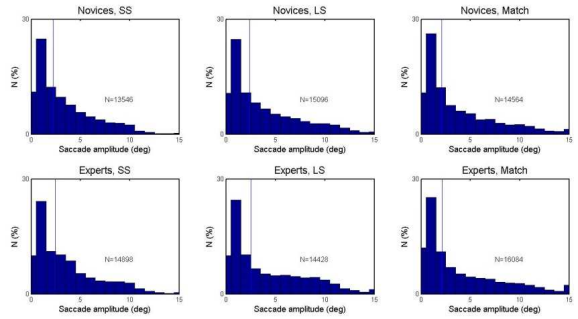


(a) Novices (b) Experts

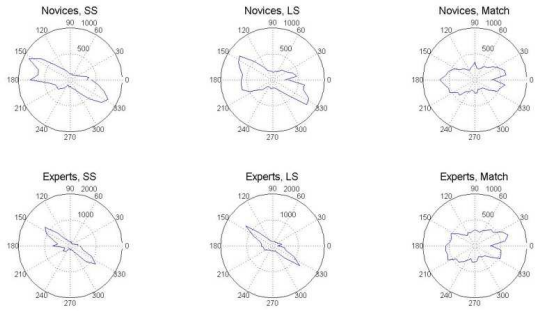
Figure 3. Examples of scanpaths of individual observers, for both novices (left panels) and experts (right panels), for the three typologies of shots (M, SS, LS). For both SS and LS, 24 scanpaths, lasting individually about ten seconds, are superimposed in each panel (one for each trial), whereas during match observation, there is only a single, 5 min long continuous scanpath. For simplicity, the same background table image has been used in all panels. Here the ocular traces during shot viewing are shown in their original orientation.



(a) Fixation duration



(b) Gaze shift amplitude



(c) Gaze shift direction

Figure 4. Distributions of the three oculomotor features used to classify expertise. Top panels (4(a)), fixation duration; middle panels (4(b)), gaze shift amplitude; bottom panels (4(c)), gaze shift direction. Vertical solid lines, median values. SS=Short Shots, LS=Long Shots.

or combination of features - actually three for the long shots and two on the match - thus indicating a higher efficiency than the other approaches.

Tables 1 and 2 report results in terms of the accuracy (percent correct) and discriminability ($d^{'}$), respectively. Accuracy was defined as N_c/N_{tot}, where N_c is the number of trials in which correct classification was attained, regardless of the stimulus (novice or expert). Discriminability was computed as $Z_H - Z_F$, where Z_H is the z-transformed hit rate (a hit being a "novice" classification given a "novice" stimulus) and Z_F is the

Table 1
Mean classification accuracy with the Multiple Kernel analysis. Base features: gaze shift amplitude (A), gaze shift direction (D), fixation duration (F). Best and worst performances are marked in green and in red, respectively

	Mean accuracy (%)		
Features	Short Shots	Long Shots	Match
A	79.52	83.33	80.95
D	70.95	86.19	71.42
F	88.09	80.47	81.90
A + D	68.57	86.19	72.85
A + F	68.09	80.00	81.90
D + F	68.57	86.19	63.80
A + D + F	67.61	85.23	77.61

Table 2
Classification discriminability ($d^{'}$) with the Multiple Kernel analysis. Base features: gaze shift amplitude (A), gaze shift direction (D), fixation duration (F). Best and worst performances are marked in green and in red, respectively

	Discriminability ($d^{'}$)		
Features	Short Shots	Long Shots	Match
A	1.709	1.948	1.766
D	1.077	2.220	1.137
F	2.399	1.726	1.836
A + D	0.991	2.206	1.198
A + F	0.959	1.700	1.852
D + F	0.992	2.211	0.717
A + D + F	0.935	2.120	1.512

z-transformed false alarms rate (a false alarm being a "novice" classification given an "expert" stimulus). Discriminability represents the capability of the classifier to separate novices and experts, regardless of the decision criterion.

For both accuracy and discriminability the reported tables represent the mean values across the 5 classifier repetitions, separately for each feature or feature combination and for each stimulus typology. We define the best performance as the highest classification score reported within each stimulus typology (short shots, long shots, match), regardless of which feature, or combination thereof, contributed to it. In case of ties, the best performance was stipulated to be the one in which both accuracy and discriminability were highest. From Table 1 it can be seen that the classification rate was rather good (range: $63.80\% - 88.09\%$) and always above chance ($p < 0.01$ even for the lowest classification rate, one-tail binomial test), with a rather high best performance within each stimulus typology (88.09%, 86.19% and 81.90%, marked in green; red denotes the worst performances within each stimulus typology).

In the best case (88.09%) this amounts to saying that the RVM correctly distinguished as being a novice or an expert 37 out of 42 observers, with a moderate bias

to classify correctly novices better than experts (predictive value for novices: 0.917; predictive value for experts: 0.851). By considering the best performances, which show the achievement of the classifier, accuracy was higher with the short shots (88.09%) than the match (81.90%), with the performance with the long shots being somewhat intermediate (86.19%). A one-way ANOVA among the 3 best performances showed a marginally significant effect of classification conditions (either stimulus type or feature; $F(2,12) = 3.547$, $p = 0.062$). Post-hoc LSD pairwise tests indicated that, whereas the two former figures (88.09% and 86.19%) did not differ significantly from each other ($p > 0.4$), the difference with the accuracy measure obtained with the match stimulus (81.90%) was statistically significant or marginally significant ($p = 0.023$ and $p = 0.097$, respectively).

No clear tendency could be appreciated as to which feature, or combination of features, best contributed to the classification. From Table 1 it can be seen that in no case the same feature, or combination thereof, determined the best accuracy across the three stimulus typologies. In terms of mean performance, using single features provided a somewhat better result (80.31%) than combining them in pairs (75.13%) or triplet (76.82%). The best classification performance within each stimulus category was never obtained with the triplet of features, though only in one case the triplet determined the worst performance (67.61%). An almost identical pattern of results was obtained by computing d' as index of performance (Table 2). Again, the best performance within each stimulus category was higher with the shots than with the match. Interestingly, also the three worst performances (marked in red in the Tables) were coincident for accuracy and discriminability, and were higher for the long shots than the short shots.

Discussion

In this study we have applied machine learning techniques (MKL-based feature combination and RVM) to analyze the oculomotor behavior of individual observers engaged in a visual task, with the aim of classifying them as experts or novices. To this end, we have administered to 42 subjects, half novices and half expert billiard players, various visual stimuli and tasks. As stimuli we used a portion of a real match, videorecorded from the top, containing several shots of variable length and complexity, as well as a number of ad-hoc individual shots, also videorecorded from the top in a real setting. The match stimulus was associated to a free-viewing observation condition, while for the individual shots, which were occluded in the final part of the trajectory, observers were asked to predict the outcome of the shot, which placed implicitly a significant constraint on the deployment of visuospatial attention, and, consequently, on the overt scanpath. Thus, we demonstrated that, in both constrained and unconstrained naturalistic viewing conditions, eye movements contain enough information to detect an internal state such as expertise.

To our knowledge this is the first time that MKL-based feature combination and RVM techniques are applied to eye movement data. A very recent study by Henderson, Shinkareva, Wang, Luke, and Olejarczyk (2013) inferred successfully the observers' cognitive task (search, memorizing, reading) through classification. However, for the purpose of that study, a dedicated classifier was trained for each observer, and a simple baseline technique as the Naïve Bayes' classifier was sufficient. Clearly, when addressing a scenario in which individual observers are classified as belonging to one or another population, more sophisticated machine learning tools are needed. Many studies used an approach based on SVM classification (e.g., Lagun et al., 2011; Eivazi & Bednarik, 2011; Bednarik et al., 2005; Vig et al., 2009; Tseng et al., 2013; Bulling, Ward, Gellersen, & Trster, 2011; Bednarik, Vrzakova, & Hradis, 2012). Beyond some limitations inherent to SVM (Tipping, 2001; Murphy, 2012), it is worth pointing out that the final classification step is just one side of the problem when spotting expertise from scanpaths in a *data-driven* way, the other side being how features are best combined and exploited. As anticipated in the Introduction, to address these issues we have adopted a feature fusion strategy relying on multiple kernel combination.

A comment is due on the choice of the features. The feature we have used are typical basic parameters that characterize saccadic exploration of static scenes. However, our stimuli contained also moving elements (e.g., the ball motion) capable of eliciting smooth pursuit eye movements, which are characterized by different parameters. Thus, it may be argued that using saccade parameters is not too appropriate. Let us firstly note that in our experiment smooth pursuit eye movements were in fact not frequent. Although this may sound surprising, consider that our observers were not instructed to follow the moving target; also, the ball motion occupied only a minor part of the overall stimulus duration, and furthermore its motion was not continuous but interrupted by bounces, which implied rather frequent catchup/anticipatory saccades. To take specific figures, consider the shot trials (Crespi et al., 2012): the ball was in motion for about 2.1 seconds in each trial, on average. During this short time window, the eyes spent on average only 63% of the time in slow motion (tangential velocity between 0.5 and 40 deg/s with a minimum duration of 100 ms), which amounts to about 1.3 seconds per trial. Considering that the mean recording window within a trial was 12.4 seconds, this indicates that smooth pursuit eye movements contributed to the overall eye movements pattern for only about 10% of the time. We did not measure all these parameters in the match task, but we can

assume comparable figures. Secondly, much of the difference between experts and novices was found when the ball was not moving (ROI analysis, figs. 5 and 6 in Crespi et al., 2012; VDA peaks, fig. 2 in Robino et al., 2012). Thirdly, and more importantly, from the perspective of machine learning, segmenting a gradually changing signal into discrete elements and using them as features for the classifier is perfectly legitimate. Using virtual fixations or whatever other signal preprocessing of the oculomotor traces before the classification step is just a matter of convenience, as it is well known that machine learning techniques are blind as to the nature of the underlying processes. To the extent that features bring information, they work (features do not introduce new information).

Indeed, by combining only three basic parameters of visual exploration, the overall classification accuracy, expressed as percent correct and averaged across stimulus types and oculomotor features, scored a respectable 78%. More interesting is to consider the best performance for each stimulus type, which testifies the achievement of the classifier, and which depends on the features used. The best performance ranged between 81.90% and 88.09% - 1.852 to 2.399 in terms of d', which is a quite remarkable result, especially considering that a naturalistic, unconstrained viewing condition was included (M). Beside confirming that eye movements contain a signature of billiard expertise (Crespi et al., 2012), this finding demonstrates that, even ignoring "where" the gaze is directed, i.e., to which objects or events overt visuospatial attention is allocated (*content-driven* approach), the "expert's eye" can be identified at the individual level from "how" the gaze is shifted, i.e, from basic oculomotor features such as saccade amplitude and direction and fixation duration (*data-driven* approach). Clearly, this does not amount to saying that the physiology of eye movements is modified by expertise, nor that expertise in a given field could be detected by using whatever visual stimulus, but simply that there is not always the need to match the oculomotor features with the visual features, a common approach that we also used in our past work (Crespi et al., 2012; Robino et al., 2012). Notably, expertise detection was successful at the level of individual observers (see below).

The classification accuracy was higher with the shots than the match. This difference, despite being small, is in keeping with the idea that the individual scanpath provides an indication about the degree of "expertise allocation", that is, how much an observer is actually using knowledge: The more expertise is used, the larger the systematic differences in visual exploration between a novice and an expert, hence the higher the classification performance. For example, the prediction task in which participants had to make a rapid guess as to the outcome of the shots ("will the ball hit the central skittle?") would seem to leave little room for free ocular exploration, especially for the short shots, thus reducing the idiosyncratic component of ocular exploration. As a consequence, the systematic differences between novices and experts emerge more clearly. Conversely, the fact that during match observation observers had no specific task, and that the pace of the shots was relatively relaxed, allowed more free eye movements, especially after the shots. In other words, the difference between the classification accuracy when the shots rather than the match stimulus is used might depend on the different degree of "expertise allocation" in the two conditions, being higher in the shot prediction task than in the relatively unconstrained match observation task. Indeed, we had previously proposed that, during billiard match observation, it is precisely the alternation between the focusing of attention on the upcoming shot and the post-shot relaxation that allowed us to successfully parse the shot alternation exclusively on the basis of the scanpath differences between novices and experts (Robino et al., 2012).

The above considerations underscore the importance of selecting a proper test setting in order to detect expertise from the scanpath. On the one hand, it is clearly better to find the conditions (i.e., stimuli and tasks) that best elicit the use of expertise. These should be as stringent and controlled as possible, such as for example the ad-hoc shots coupled with the prediction task that we have used, where the highest classification performance was attained. On the other hand, it is intriguing that the RVM yielded a high accuracy, though not the highest, also with the match stimulus (81.90%). Considering the uncontrolled variability of a real billiard match, coupled with the lack of a specific task for the observers, we think this is a remarkable achievement in terms of capability to extract information from eye movements in naturalistic conditions. Pervasive behavioural monitoring of real-life visual exploration through wearable eye trackers may take advantage of high-performance classification methods such as RVM (Schumann et al., 2008; Hart, Onceanu, Sohn, Wightman, & Vertegaal, 2009; Noris, Nadel, Barker, Hadjikhani, & Billard, 2012; Vidal, Turner, Bulling, & Gellersen, 2012). Furthermore, especially for real-life conditions, it is crucial that the scanpath analysis can be *data-driven*, at least as much as possible, as a *content-driven* approach would inevitably require manual labeling of each video frame in terms of semantically-identified regions or visual elements. Indeed, this would preclude an automatic analysis of real-life scanpaths, and even more so for a real-time analysis.

Besides confirming that top-down cognitive processes are an important factor in gaze guidance (Buswell, 1935; Yarbus, 1967; Rayner, 1998; Viviani, 1990; Henderson, 2003), our study has an applicative potential.

Firstly, our findings suggest that a number of low-level physiological parameters of visual exploration be-

havior could be suitably used to automatically decode inner cognitive processes to the benefit of BCI systems. In the field of neuro-rehabilitation, for example, many efforts are directed at decoding motor imagery and covert motor commands from brain signals with the goal of driving prosthetic devices and boosting motor improvement through neurofeedback training (Silvoni et al., 2011). Central to this endeavor is the capability to extract in the simplest possible way useful neural information from subjects engaged in some sort of mental imagery tasks. For this, brain activity is recorded via amplifiers and decoded using on-line classification algorithms. Brain signals are not the only physiological correlate of mental imagery, however. Eye movements have been shown to tag in a precise way an elusive covert process such as mental imagery (Brandt & Stark, 1997; Johansson, Holsanova, Dewhurst, & Holmqvist, 2012), and, more specifically, dynamic motion imagery (de'Sperati, 1999, 2003; de'Sperati & Santandrea, 2005; Jonikaitis, Deubel, & de'Sperati, 2009; Crespi et al., 2012). Thus, the methodological approach that we have described in the present study might be profitably applied to extract eye movements information to drive BCI external devices. For example, automatically classifying good and bad imagery performance could help to refine the mental training procedures until expertise is achieved, or to avoid that incorrect signals are erroneously sent to the BCI device. Also, a classifier could detect spurious eye movements - or their absence - that might mean that visuospatial attention has been drawn from the current imagery task. In sum, an oculomotor-based channel with efficient classification capabilities could be suitably paired to EEG-based or fMRI-based channels to improve mind reading performance in hybrid, multiple input signal sources BCI systems (Amiri, Fazel-Rezai, & Asadpour, 2013).

Another potential application of our approach is the development of an expertise test based on the "expert's eye". Clearly, a general expertise test cannot exist. Expertise is specific to particular domains, and it can be of various types and qualities (e.g., declarative-conceptual, procedural, strategic; (De Jong & Ferguson-Hessler, 1996). Although expertise is ultimately established by directly measuring performance (e.g., through questionnaire scores, as in school grades, or with official rankings, as in sports), an indirect assessment of the visual exploratory behaviour may uncover subtle aspects underlying expertise in all those cases where visual information is crucial (e.g. understanding the working of a mechanical apparatus, or providing legal authentication of a painting, or playing chess, or detecting faults in sports). For example, in our previous study on billiard expertise we have documented, through eye movement recording, the passage from intuitive, procedural knowledge based on mental imagery, a strategy typical of novices, to rule-based, conceptual knowledge, which was expressed only in experts (Crespi et al., 2012). Incidentally, this may explain the small bias that we have found with the best performance towards a higher misclassification of experts than novices: because experts can adopt a novice's strategy but a novice cannot adopt an expert's strategy, a classifier can be fooled by an expert but not by a novice.

The capability to detect expertise automatically, that is, without the need of semantically analyzing which particular objects and events of a visual scene the gaze of an observer is directed to, will enhance "mind reading" methods. However, it should be borne in mind that a psychophysiological test for the expert's eye would not substitute direct measures of expertise, but rather complement them. Thus, finding a mismatch between the output of an automatic "ocular expertmeter" and the outcome of direct evaluation of expertise obtained with classical methods (e.g., testing, questionnaires) could raise issues as to what strategy or what evidence has actually been used. For example, assuming that the scanpath is indicative of expertise, the finding of an anomalous scanpath in inspecting the figures of a difficult geometry exam would perhaps question what mental procedure was used by a student who nonetheless provided all correct answers; An alternative interpretation could be that the student answered correctly by chance.

The automatic recognition of individual traits through behavioral analyses is an intensely pursued goal. Biometrics is a field of study aimed at identifying individuals through their unique biological characteristics or behavioral patterns. Biological methods in biometrics include for example fingerprint, face, or iris verification, whereas behavioral methods include voice, signature, typing or gait analysis. Recently, behavioral biometrics has been applied to eye movements, with the goal of identifying individuals through their oculomotor patterns (Holland & Komogortsev, 2011), even in a task-independent way (Bednarik et al., 2005; Kinnunen, Sedlak, & Bednarik, 2010). In these studies various methods to analyze eye movements have been used, with an ensuing performance however still short of the accepted standards for biometrics systems. Our work was aimed at distinguishing a novice from an expert, that is, two classes of individuals rather than a given individual as in biometrics. Though, the high classification rates that we obtained, even in a poorly constrained scenario such as match observation, suggests that our approach based on feature space fusion and a Bayesian sparse classifier could be profitably applied to personal identification as well. It is interesting that a similar set of eye movements features (e.g., duration and amplitude of saccades) can be used successfully for both individual and categorical classification (personal identity or expertise). This seems to confirm that these basic features are more than just oculomotor traits.

Appendix
RVM and SVM

Denote $\{\mathbf{x}_n, t_n\}_{n=1}^N$ a training data set of N samples with $\mathbf{x}_n \in \mathbb{R}^D$ and $t_n \in \mathcal{C}$, $\mathcal{C}$ being the classification space of dimension C. The SVM approach (Cristianini & Shawe-Taylor, 2000) relies on building a classifier of the form $sign[f(\mathbf{x};\mathbf{w})]$ where $t_n \in \{-1, 1\}$ (binary classification, $C = 2$) and

$$f(\mathbf{x};\mathbf{w}) = \sum_{i=1}^{M} w_i \phi_i(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}), \tag{4}$$

is a linear regression model $\hat{\mathbf{t}} = f(\mathbf{x};\mathbf{w})$ that approximates the true mapping function $\mathbf{t}$. In Eq. 4, $\phi(\cdot)$ represent a generally nonlinear and fixed basis functions, mapping the input space in a higher dimensional space, and $\mathbf{w} \in \mathbb{R}^M$ are adjustable parameters (or weights) that appear linearly in (4). Note that, though the model is linear in the parameters, it may still be highly flexible as the size of the basis set, M, may be very large. The objective of training is to estimate good values for those parameters, which in the SVM framework is accomplished through an optimization technique (Cristianini & Shawe-Taylor, 2000). Also, in the SVM, the model is implicitly defined such that $M = N$, i.e. designining one basis function for each example in the training set. A particular kind of function known as *kernel function* is employed, which provides an implicit calculation of the product between $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$, i.e., $K(\mathbf{x}_i, \mathbf{x}_j) =< \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) >$; thus, predictions are based on the function:

$$f(\mathbf{x};\mathbf{w}) = \sum_{i=1}^{M} w_i K(\mathbf{x}, \mathbf{x}_i). \tag{5}$$

The key feature of the SVM is that, in the classification case, its target function attempts to minimise a measure of error on the training set while simultaneously maximising the *margin* between the two classes (in the feature space implicitly defined by the kernel). This is a highly effective mechanism for avoiding over-fitting, which leads to good generalisation. It furthermore results in a sparse model dependent only on a subset of kernel functions: those associated with training examples $\mathbf{x}_n$ that lie either on the margin or on the "wrong" side of it, namely the *support vectors* (Cristianini & Shawe-Taylor, 2000).

The RVM has the same functional form as SVMs, but is conceived in a Bayesian framework (Tipping, 2001). Following the standard probabilistic formulation, the targets are assumed to be samples generated from the model (4) perturbed with a noise process ε:

$$t_n = f(\mathbf{x}_n;\mathbf{w}) + \varepsilon_n. \tag{6}$$

Here ε represents the error between the estimated targets $\hat{\mathbf{t}}$ and the true ones $\mathbf{t}$, which is assumed to be normally distributed with zero mean and unknown variance σ^2, i.e. $t_n \sim P(t_n|\mathbf{x}_n, \mathbf{w}, \sigma^2) = \mathcal{N}(t_n|f(x_n;\mathbf{w}), \sigma^2)$, where the latter notation specifies a Gaussian distribution $\mathcal{N}$ over the target labels with mean $f(x_n;\mathbf{w})$ and variance σ^2. Under independent and identical distribution generation of the observations, the data likelihood can be written as:

$$P(\mathbf{t}|\mathbf{w}, \sigma^2) = \prod_{n=1}^{N} \mathcal{N}(t_n|f(x_n;\mathbf{w}), \sigma^2). \tag{7}$$

From now on, we will write terms such as $P(\mathbf{t}|\mathbf{x}, \mathbf{w}, \sigma^2)$ as $P(\mathbf{t}|\mathbf{w}, \sigma^2)$. Omitting to include $\mathbf{x}$ variables is purely for notational convenience and it implies no further model assumptions.

In a Bayesian framework, model parameters $\mathbf{w}$ and σ^2 are considered as random variables. These are estimated by first assigning prior distributions and then estimating their posterior distribution using the likelihood of the observed data (Eq. 7). The key of the RVM approach (Tipping, 2001) is to define a prior conditional distribution on each coefficient w_i, such that, according to the *Automatic Relevance Determination* (ARD) mechanism (MacKay, 1992), all coefficient which are unnecessary are pruned:

$$P(\mathbf{w}|\alpha) = \prod_{n=1}^{N} \mathcal{N}(\mathbf{w}_n|0, \alpha_n^{-1}), \tag{8}$$

where $\alpha = [\alpha_1, \cdots, \alpha_N]^T$ is the vector of RVM hyper-parameters. Since many of such hyper-parameters usually assume elevated values, their associated weights will be sharply peaked around zero. This has the effect of switching off basis functions for which there is no evidence in the data, yielding sparse prediction models. Thus, unlike the SVM, the RVM explicitly encodes the criterion of model sparsity as a prior over the model weights. Whilst in SVM regression/classification a desirable level of sparsity has to be brought about indirectly by determining an error or margin parameter via a cross-validation scheme, the Bayesian formulation of the regression problem in the RVM allows for a prior structure that explicitly encodes the desirability of sparse representations (Tipping, 2001).

As a practical consequence, for SVM the support vectors are typically formed by "borderline" , difficult-to-classify samples in the training set, which are located near the decision boundary of the classifier; in contrast, for RVM the *relevance vectors* are formed by samples appearing to be more representative of the two classes, which are located away from the decision boundary of the classifier.

To sum up, from a Bayesian standpoint, the goal is to estimate from data the model parameters $\mathbf{w}, \alpha, \sigma^2$; unfortunately an analytical expression for the posterior distribution $P(\mathbf{w}, \alpha, \sigma^2|\mathbf{t})$ is not available. However, the posterior can be factorised as $P(\mathbf{w}, \alpha, \sigma^2|\mathbf{t}) =$

$P(\mathbf{w}|\mathbf{t},\alpha,\sigma^2)P(\alpha,\sigma^2|\mathbf{t})$. The first term $P(\mathbf{w}|\mathbf{t},\alpha,\sigma^2)$ is the posterior probability of the weights given α,σ^2, namely via Baye' rule,

$$P(\mathbf{w}|\mathbf{t},\alpha,\sigma^2) = \frac{P(\mathbf{t}|\mathbf{w},\sigma^2)P(\mathbf{w}|\alpha)}{P(\mathbf{t}|\alpha,\sigma^2)}, \tag{9}$$

which, by considering Eqs. 7 and 8, is normally distributed. The second term $P(\alpha,\sigma^2|\mathbf{t})$ is the posterior probability of α and σ^2. For an in-depth discussion about the calculus of these probabilities, the reader should refer to (Tipping, 2001).

The RVM classifier based on Multiple Kernels (Damoulas & Girolami, 2009a; Psorakis et al., 2010) can be obtained by generalizing Eq. (5) as follows. A base kernel can be combined into an $N \times N$ composite kernel as $K^{\beta}(\mathbf{x}_i,\mathbf{x}_j) = \sum_{s=1}^{S} \beta^s K^s(\mathbf{x}_i^s,\mathbf{x}_j^s)$ (Eq. 1, Data analyses Section).

More precisely, denote with the matrix $\mathbf{X} \in \mathbb{R}^{N\times D}$ the input data from which the kernel matrix $\mathbf{K}^{\beta} \in \mathbb{R}^{N\times N}$ is derived, where each row $\mathbf{k}_n^{\beta}$ expresses how related, based on the selected kernel function, observation $\mathbf{x}_n$ is to the others of the training set. The learning process involves the learning of model parameters $\mathbf{W} \in \mathbb{R}^{N\times C}$, which by the quantity $\mathbf{W}^T\mathbf{K}^{\beta}$ act as a voting system to express which relationships of the data are important in order for our model to have appropriate discriminative properties

By introducing the auxiliary variables $\mathbf{Y} \in \mathbb{R}^{N\times C}$, we regress on $\mathbf{Y}$ with a standardized noise model; thus, for a sample n and a class c, Eq.7 can be written as:

$$y_{nc}|\mathbf{w}_c,\mathbf{k}_n^{\beta} \sim \mathcal{N}_{y_{nc}}(\mathbf{k}_n^{\beta}\mathbf{w}_c,1), \tag{10}$$

where the vector $\mathbf{w}_c$ defines the c-th column of the model parameters matrix $\mathbf{W}$. The regression target is linked to the classification label by setting $t_n = 1$ if $y_{in} > y_{jn}\ \forall j \neq i$.

This way, the posterior class membership distribution is the multinomial probit likelihood (details in Damoulas & Girolami, 2009a)

$$P(t_n = i|\mathbf{W},\mathbf{k}_n^{\beta}) = E_{p(u)}\left\{\prod_{j\neq i}\Phi(u+\mathbf{k}_n^{\beta}(\mathbf{w}_i-\mathbf{w}_j))\right\} \tag{11}$$

where $u \sim \mathcal{N}(0,1)$ and Φ is the Gaussian cumulative distribution function. Following the RVM approach, the elements w_{nc} of matrix $\mathbf{W}$ follow a standard normal distribution with zero mean and variance α_{nc}^{-1}, where the latter are the elements of the hyper-parameter matrix $\mathbf{A} \in \mathbb{R}^{N\times C}$, $P(\mathbf{W}|\mathbf{A}) = \prod_{n=1}^{N}\prod_{c=1}^{C}\mathcal{N}(\mathbf{w}_n|0,\alpha_{nc}^{-1})$, while α_{nc} follow a Gamma distribution, thus $P(\mathbf{A}|a,b) = \prod_{n=1}^{N}\prod_{c=1}^{C}\mathcal{G}(a,b)$. With sufficiently small hyper-parameters $a,b(<10^{-5})$ the scales $\mathbf{A}$ restrict $\mathbf{W}$ around its zero mean due to small variance.

The learning procedure for latent variables $\mathbf{Y}$ and parameters $\mathbf{W},\mathbf{A},\beta$ is a generalised Expectation-Maximization algorithm, which can be summarised as follows.

Step 1: use a type-II Maximum Likelihood (ML) procedure, which maximises the log of the marginal likelihood $\log P(\mathbf{Y}|\mathbf{K}^{\beta},\mathbf{A}) = \log\int P(\mathbf{Y}|\mathbf{K}^{\beta},\mathbf{W})P(\mathbf{W}|\mathbf{A})d\mathbf{W}$ with respect to $\mathbf{A}$, and boils down to either add a sample or update its associated hyper-parameter α_{nc}; thus, the model can start with a single sample and proceed in a constructive manner.

Step 2: perform an M-step for obtaining $\mathbf{W}$.

Step 3: perform an E-step for $\mathbf{Y}$.

Step 4: obtain β_s weights via constrained Quadratic Programming.

Step 1 to 4 are iterated using as a convergence measure the % mean change of $(\mathbf{Y}-\mathbf{K}^{\beta}\mathbf{W})^2$. Once the parameters of the model have been learned, then the Multinomial probit likelihood for the calculation of class membership probabilities $P(t_n = i|\mathbf{W},\mathbf{k}_n^{\beta})$ (Eq. 11) is computed by resorting to Quadrature approximation for solving the expectation integral. For details, see (Damoulas & Girolami, 2009a; Psorakis et al., 2010).

References

Amiri, S., Fazel-Rezai, R., & Asadpour, V. (2013). A review of hybrid brain-computer interface systems. *Advances in Human-Computer Interaction*, *2013*, 1–8. doi: 10.1155/2013/187024

Bednarik, R., Kinnunen, T., Mihaila, A., & Fränti, P. (2005). Eye-movements as a biometric. In H. Kalviainen, J. Parkkinen, & A. Kaarna (Eds.), *Image analysis* (Vol. 3540, p. 780-789). Springer Berlin Heidelberg.

Bednarik, R., Vrzakova, H., & Hradis, M. (2012). What do you want to do next: A novel approach for intent prediction in gaze-based interaction. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA '12)* (pp. 83–90). New York, NY, USA: ACM.

Boccignone, G., & Ferraro, M. (2013a). Feed and fly control of visual scanpaths for foveation image processing. *annals of telecommunications–annales des télécommunications*, *68*(3-4), 201-217.

Boccignone, G., & Ferraro, M. (2013b). Gaze shift behavior on video as composite information foraging. *Signal Processing: Image Communication*, *28*(8), 949 - 966.

Boccignone, G., & Ferraro, M. (2014, Feb). Ecological sampling of gaze shifts. *IEEE Transactions on Cybernetics*, *44*(2), 266-279.

Boccignone, G., Marcelli, A., Napoletano, P., Di Fiore, G., Iacovoni, G., & Morsa, S. (2008). Bayesian integration of face and low-level cues for foveated video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, *18*(12), 1727–1740.

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 135–207.

Brandt, S. A., & Stark, L. W. (1997). Spontaneous eye movements during visual imagery reflect the content of the visual scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38.

Brockmann, D., & Geisel, T. (2000). The ecology of gaze shifts. *Neurocomputing*, *32*(1), 643–650.

Bulling, A., Ward, J., Gellersen, H., & Trster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(4), 741–753.

Buswell, G. T. (1935). *How people look at pictures*. University of Chicago Press Chicago.

Caldara, R., & Miellet, S. (2011). imap: a novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*(3), 864–878.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Choi, Y. S., Mosley, A. D., & Stark, L. W. (1995). String editing analysis of human visual search. *Optometry & Vision Science*, *72*(7), 439–451.

Crespi, S., Robino, C., Silva, O., & de'Sperati, C. (2012). Spotting expertise in the eyes: Billiards knowledge as revealed by gaze shifts in a dynamic visual prediction task. *Journal of Vision*, *12*(11). doi: 10.1167/12.11.30

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). Scanmatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700.

Damoulas, T., & Girolami, M. A. (2009a). Combining feature spaces for classification. *Pattern Recognition*, *42*(11), 2671–2683.

Damoulas, T., & Girolami, M. A. (2009b). Pattern recognition with a bayesian kernel combination machine. *Pattern Recognition Letters*, *30*(1), 46–54.

Damoulas, T., Ying, Y., Girolami, M. A., & Campbell, C. (2008). Inferring sparse kernel combinations and relevance vectors: An application to subcellular localization of proteins. In *Seventh international conference on machine learning and applications, 2008 (ICMLA'08)* (pp. 577–582).

De Jong, T., & Ferguson-Hessler, M. G. (1996). Types and qualities of knowledge. *Educational Psychologist*, *31*(2), 105–113.

Dempere-Marco, L., Hu, X.-P., Ellis, S. M., Hansell, D. M., & Yang, G.-Z. (2006). Analysis of visual search patterns with emd metric in normalized anatomical space. *IEEE Transactions on Medical Imaging*, *25*(8), 1011–1021.

de'Sperati, C. (1999). Saccades to mentally rotated targets. *Experimental Brain Research*, *126*(4), 563–577.

de'Sperati, C. (2003). The inner working of dynamic visuospatial imagery as revealed by spontaneous eye movements. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: cognitive and applied aspects of eye movement research* (pp. 119–141). North-Holland.

de'Sperati, C., & Santandrea, E. (2005). Smooth pursuit-like eye movements during mental extrapolation of motion: The facilitatory effect of drowsiness. *Cognitive Brain Research*, *25*(1), 328–338.

Donovan, T., & Manning, D. (2007). The radiology task: Bayesian theory and perception. *British Journal of Radiology*, *80*(954), 389–391.

Eivazi, S., & Bednarik, R. (2011). Predicting problem-solving behavior and performance levels from visual attention data. In *Proceedings 2nd Workshop on Eye Gaze in Intelligent Human Machine Interaction at IUI* (p. 9-16). Palo Alto, California, USA.

Ellis, S., & Stark, L. (1986). Statistical dependency in visual scanning. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *28*(4), 421–438.

Feng, G. (2006). Eye movements as time-series random variables: A stochastic model of eye movement control in reading. *Cognitive Systems Research*, *7*(1), 70–95.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2). doi: 10.1167/8.2.6

Grindinger, T. J., Murali, V. N., Tetreault, S., Duchowski, A. T., Birchfield, S. T., & Orero, P. (2011). Algorithm for discriminating aggregate gaze points: Comparison with salient regions-of-interest. In *Proceedings of the 2010 International Conference on Computer Vision - Volume Part I* (pp. 390–399). Berlin, Heidelberg: Springer-Verlag.

Hacisalihzade, S., Stark, L., & Allen, J. (1992). Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Transactions on Systems, Man, Cybernetics B*, *22*(3), 474–481.

Hart, J., Onceanu, D., Sohn, C., Wightman, D., & Vertegaal, R. (2009). The attentive hearing aid: Eye selection of auditory sources for hearing impaired users. In T. Gross et al. (Eds.), *Human-Computer Interaction - INTERACT 2009* (Vol. 5726, p. 19-35). Springer Berlin Heidelberg.

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, *11*(10). doi: 10.1167/11.10.10

Hembrooke, H., Feusner, M., & Gay, G. (2006). Averaging scan patterns and what they can tell us. In *Proceedings of the 2006 Symposium on Eye Tracking Research &Amp; Applications (ETRA '06)* (pp. 41–41). New York, NY, USA: ACM.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, *7*(11), 498–504.

Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., & Olejarczyk, J. (2013). Predicting cognitive state from eye movements. *PLoS ONE*, *8*(5), e64937.

Holland, C., & Komogortsev, O. V. (2011). Biometric identification via eye movement scanpaths in reading. In *Proceedings of the 2011 International Joint Conference on Biometrics (IJCB '11)* (pp. 1–8). Washington, DC, USA: IEEE Computer Society.

Humphrey, K., & Underwood, G. (2009). Domain knowledge moderates the influence of visual saliency in scene recognition. *British Journal of Psychology*, *100*(2), 377–398.

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)* (pp. 211–218). New York, NY, USA: ACM.

Johansson, R., Holsanova, J., Dewhurst, R., & Holmqvist, K. (2012). Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1289–1314.

Jonikaitis, D., Deubel, H., & de'Sperati, C. (2009). Time gaps in mental imagery introduced by competing saccadic tasks. *Vision Research*, *49*(17), 2164–2175.

Josephson, S., & Holmes, M. E. (2006). Clutter or content?: How on-screen enhancements affect how tv viewers scan and what they learn. In *Proceedings of the 2006 Sympo-*

sium on Eye Tracking Research & Applications (ETRA '06) (pp. 155–162). New York, NY, USA: ACM.

Kinnunen, T., Sedlak, F., & Bednarik, R. (2010). Towards task-independent person authentication using eye movement signals. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)* (pp. 187–190). New York, NY, USA: ACM.

Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein, E. (2011). Detecting cognitive impairment by eye movement analysis using automatic classification algorithms. *Journal of Neuroscience Methods*, *201*(1), 196–203.

MacKay, D. J. (1992). Bayesian interpolation. *Neural Computation*, *4*(3), 415–447.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. The MIT Press.

Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Academic Radiology*, *3*(12), 1000–1006.

Nodine, C. F., Locher, P. J., & Krupinski, E. A. (1993). The role of formal art training on perception and aesthetic judgment of art compositions. *Leonardo*, 219–227.

Noris, B., Nadel, J., Barker, M., Hadjikhani, N., & Billard, A. (2012). Investigating gaze of children with asd in naturalistic settings. *PLoS ONE*, *7*(9), e44144.

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, *171*(968), 308-311.

Pihko, E., Virtanen, A., Saarinen, V.-M., Pannasch, S., Hirvenkari, L., Tossavainen, T., … Hari, R. (2011). Experiencing art: the influence of expertise and painting abstraction level. *Frontiers in Human Neuroscience*, *5*(94). doi: 10.3389/fnhum.2011.00094

Privitera, C. M., & Stark, L. W. (2000, September). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEETransactions on Pattern Analysis and Machine Intelligence*, *22*(9), 970–982.

Psorakis, I., Damoulas, T., & Girolami, M. A. (2010). Multiclass relevance vector machines: sparsity and accuracy. *IEEE Transactions on Neural Networks*, *21*(10), 1588–1598.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372.

Reingold, E., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. In S. Liversedge, I. Gilchrist, & S. Everlingl (Eds.), *Oxford handbook on eye movements* (pp. 528–550). Oxford University Press.

Robino, C., Crespi, S., Silva, O., & de'Sperati, C. (2012). Parsing visual stimuli into temporal units through eye movements. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA '12)* (pp. 181–184). New York, NY, USA: ACM.

Schumann, F., Einhäuser, W., Vockeroth, J., Bartl, K., Schneider, E., & König, P. (2008). Salient features in gaze-aligned recordings of human visual input during free exploration of natural environments. *Journal of Vision*, *8*(14). doi: 10.1167/8.14.12

Shanteau, J., Weiss, D. J., Thomas, R. P., & Pounds, J. C. (2002). Performance-based assessment of expertise: How to decide if someone is an expert or not. *European Journal of Operational Research*, *136*(2), 253–263.

Silvoni, S., Ramos-Murguialday, A., Cavinato, M., Volpato, C., Cisotto, G., Turolla, A., … Birbaumer, N. (2011). Brain-computer interface in stroke: a review of progress. *Clinical EEG and Neuroscience*, *42*(4), 245–252.

Tatler, B., & Vincent, B. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2), 1–18.

Tatler, B., & Vincent, B. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*(6-7), 1029–1054.

Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, *1*, 211–244.

Tseng, P.-H., Cameron, I., Pari, G., Reynolds, J., Munoz, D., & Itti, L. (2013). High-throughput classification of clinical populations from natural viewing eye movements. *Journal of Neurology*, *260*(1), 275-284.

Underwood, G. (1998). *Eye guidance in reading and scene perception*. Amsterdam: Elsevier Science.

Vickers, J. N. (2007). *Perception, cognition, and decision training: The quiet eye in action*. Champaign, IL: Human Kinetics 1.

Vidal, M., Turner, J., Bulling, A., & Gellersen, H. (2012). Wearable eye tracking for mental health monitoring. *Computer Communications*, *35*(11), 1306–1311.

Vig, E., Dorr, M., & Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. *Spatial Vision*, *22*(5), 397–408.

Viviani, P. (1990). Eye movements in visual search: cognitive, perceptual and motor control aspects. In E. Kowler (Ed.), *Reviews of oculomotor research* (Vol. 4, pp. 353–393). Elsevier.

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: eye-movement patterns and visual memory in artists and laymen. *Perception*, *36*(1), 91.

Waters, A. J., Underwood, G., & Findlay, J. M. (1997). Studying expertise in music reading: Use of a pattern-matching paradigm. *Perception & Psychophysics*, *59*(4), 477–488.

Yarbus, A. (1967). *Eye movements and vision*. New York: Plenum Press.

Zangemeister, W., Sherman, K., & Stark, L. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia*, *33*(8), 1009–1025.

FEATURE FUSION LINEAR KERNEL	Accuracy (%)						
Features	**Short Shots**	**Long shots**	**Match**	*Best performance*		*Averages*	
Gaze shift amplitude	60.00	60.95	82.85	*82.85*		*67.93*	
Gaze shift direction	70.95	86.19	71.42	*86.19*	*86.19*	*76.19*	*68.20*
Fixation duration	47.61	51.42	82.38	*82.38*		*60.47*	
Gaze shift amplitude + Gaze shift direction	70.47	81.42	72.85	*81.42*		*74.91*	
Gaze shift amplitude + Fixation duration	56.66	53.33	54.76	*56.66*	*81.42*	*54.92*	*67.32*
Gaze shift direction + Fixation duration	66.66	81.14	68.57	*81.14*		*72.12*	
Gaze shift amplitude + Gaze shift direction + Fixation duration	68.09	81.42	70.47	*81.42*	*81.42*	*73.33*	*73.33*
Best performance	*70.95*	*86.19*	*82.85*				**80.00**
Averages	*62.92*	*70.84*	*71.90*				**68.55**

FEATURE FUSION LINEAR KERNEL	Discriminability (d')						
Features	**Short Shots**	**Long shots**	**Match**	*Best performance*		*Averages*	
Gaze shift amplitude	0.508	0.568	1.914	*1.914*		*0.997*	
Gaze shift direction	1.077	2.220	1.137	*2.220*	*2.220*	*1.478*	*1.030*
Fixation duration	-0.100	0.077	1.872	*1.872*		*0.616*	
Gaze shift amplitude + Gaze shift direction	1.080	1.832	1.196	*1.832*		*1.369*	
Gaze shift amplitude + Fixation duration	0.327	0.182	0.227	*0.327*	*2.306*	*0.245*	*1.000*
Gaze shift direction + Fixation duration	0.881	2.306	0.965	*2.306*		*1.384*	
Gaze shift amplitude + Gaze shift direction + Fixation duration	0.971	1.793	1.077	*1.793*	*1.793*	*1.280*	*1.280*
Best performance	*1.080*	*2.306*	*1.914*				**1.767**
Averages	*0.678*	*1.283*	*1.198*				**1.053**

FEATURE FUSION GAUSSIAN KERNEL	Accuracy (%)						
Features	**Short Shots**	**Long shots**	**Match**	*Best performance*		*Averages*	
Gaze shift amplitude	79.52	83.33	80.95	*83.33*		*81.27*	
Gaze shift direction	71.42	76.66	67.14	*76.66*	*88.09*	*71.74*	*78.83*
Fixation duration	88.09	80.47	81.90	*88.09*		*83.49*	
Gaze shift amplitude + Gaze shift direction	72.38	77.61	68.57	*77.61*		*72.85*	
Gaze shift amplitude + Fixation duration	68.09	80.00	81.90	*81.90*	*81.90*	*76.66*	*72.43*
Gaze shift direction + Fixation duration	66.19	75.23	61.90	*75.23*		*67.77*	
Gaze shift amplitude + Gaze shift direction + Fixation duration	62.38	68.09	65.23	*68.09*	*68.09*	*65.23*	*65.23*
Best performance	*88.09*	*83.33*	*81.90*				**84.44**
Averages	*72.58*	*77.34*	*72.51*				**74.15**

FEATURE FUSION GAUSSIAN KERNEL	Discriminability (d')						
Features	**Short Shots**	**Long shots**	**Match**	*Best performance*		*Averages*	
Gaze shift amplitude	1.709	1.948	1.766	*1.948*		*1.808*	
Gaze shift direction	1.137	1.445	0.909	*1.445*	*2.399*	*1.164*	*1.653*
Fixation duration	2.399	1.726	1.836	*2.399*		*1.987*	
Gaze shift amplitude + Gaze shift direction	1.212	1.516	0.965	*1.516*		*1.231*	
Gaze shift amplitude + Fixation duration	0.959	1.700	1.852	*1.852*	*1.852*	*1.504*	*1.224*
Gaze shift direction + Fixation duration	0.852	1.349	0.614	*1.349*		*0.938*	
Gaze shift amplitude + Gaze shift direction + Fixation duration	0.638	0.938	0.797	*0.938*	*0.938*	*0.791*	*0.791*
Best performance	*2.399*	*1.948*	*1.852*				**2.066**
Averages	*1.272*	*1.517*	*1.248*				**1.346**

FEATURE FUSION MIXED KERNEL	Accuracy (%)						
Features	**Short Shots**	**Long shots**	**Match**	*Best performance*		*Averages*	
Gaze shift amplitude	79.52	83.33	80.95	*83.33*		*81.27*	
Gaze shift direction	70.95	86.19	71.42	*86.19*	*88.09*	*76.19*	*80.31*
Fixation duration	88.09	80.47	81.90	*88.09*		*83.49*	
Gaze shift amplitude + Gaze shift direction	68.57	86.19	72.85	*86.19*		*75.87*	
Gaze shift amplitude + Fixation duration	68.09	80.00	81.90	*81.90*	*86.19*	*76.66*	*75.13*
Gaze shift direction + Fixation duration	68.57	86.19	63.80	*86.19*		*72.85*	
Gaze shift amplitude + Gaze shift direction + Fixation duration	67.61	85.23	77.61	*85.23*	*85.23*	*76.82*	*76.82*
Best performance	*88.09*	*86.19*	*81.90*				**85.39**
Averages	*73.06*	*83.94*	*75.78*				**77.59**

FEATURE FUSION MIXED KERNEL	Discriminability (d')						
Features	**Short Shots**	**Long shots**	**Match**	*Best performance*		*Averages*	
Gaze shift amplitude	1.709	1.948	1.766	*1.948*		*1.808*	
Gaze shift direction	1.077	2.220	1.137	*2.220*	*2.399*	*1.478*	*1.758*
Fixation duration	2.399	1.726	1.836	*2.399*		*1.987*	
Gaze shift amplitude + Gaze shift direction	0.991	2.206	1.198	*2.206*		*1.465*	
Gaze shift amplitude + Fixation duration	0.959	1.700	1.852	*1.852*	*2.211*	*1.504*	*1.425*
Gaze shift direction + Fixation duration	0.992	2.211	0.717	*2.211*		*1.307*	
Gaze shift amplitude + Gaze shift direction + Fixation duration	0.935	2.120	1.512	*2.120*	*2.120*	*1.522*	*1.522*
Best performance	*2.399*	*2.220*	*1.852*				**2.157**
Averages	*1.295*	*2.019*	*1.431*				**1.581**

Table 1 Supplementary. Comparison of the three RVM analyses: linear kernel, gaussian kernel, multiple (mixed) kernel. Both mean classifier accuracy (left panels) and discriminability (right panels) are shown.