Recognition of Visual Memory Recall Processes Using Eye Movement Analysis

Andreas Bulling
Computer Laboratory
University of Cambridge
andreas.bulling@acm.org

Daniel Roggen
Wearable Computing Laboratory
ETH Zurich
droggen@ife.ee.ethz.ch

ABSTRACT

Physical activity, location, as well as a person's psychophysiological and affective state are common dimensions for developing context-aware systems in ubiquitous computing. An important yet missing contextual dimension is the cognitive context that comprises all aspects related to mental information processing, such as perception, memory, knowledge, or learning. In this work we investigate the feasibility of recognising visual memory recall. We use a recognition methodology that combines minimum redundancy maximum relevance feature selection (mRMR) with a support vector machine (SVM) classifier. We validate the methodology in a dual user study with a total of fourteen participants looking at familiar and unfamiliar pictures from four picture categories: abstract, landscapes, faces, and buildings. Using person-independent training, we are able to discriminate between familiar and unfamiliar abstract pictures with a top recognition rate of 84.3% (89.3% recall, 21.0% false positive rate) over all participants. We show that eye movement analysis is a promising approach to infer the cognitive context of a person and discuss the key challenges for the real-world implementation of eye-based cognition-aware systems.

Author Keywords

Visual Memory Recall, Eye Movement Analysis, Cognition-Awareness, Cognitive Context, Electrooculography (EOG)

ACM Classification Keywords

H.1.2 Models and Principles: User/Machine Systems–Human information processing; I.5.2 Pattern Recognition: Design Methodology–Pattern analysis

General Terms

Algorithms, Experimentation, Human Factors

INTRODUCTION

Context-awareness has emerged as a key area of research in ubiquitous computing [12]. Considerable advances in sensing, inferring, and using context information were achieved by investigating different dimensions of context, such as physical activity [10], location [37], or the psychophysiological and affective state of a person [20]. These common contextual dimensions do not provide a complete picture of the context of a person. An important yet not explicitly considered dimension of context is the *cognitive context* of a person. According to the major research fields in experimental psychology, we define the cognitive context to comprise all aspects related to mental processing, such as perception, memory, knowledge, and learning. We define a computing system as *cognition-aware* if it is able to sense, infer, and adapt to the cognitive context of its user.

Current context-aware systems have a hard time assessing the cognitive context in an unobtrusive manner. This is due to the fact that the cognitive context is encoded in complex neural dynamics inside the brain and few obvious cues are accessible by non-invasive measurement techniques. Cognitive neuroscience uses techniques such as functional magnetic resonance imaging (fMRI, [8]) that are not suited for real-world applications. More light-weight techniques potentially useful to get at the cognitive context, such as electroencephalography (EEG, [2]), are not (yet) unobtrusive and robust enough for use in mobile daily life settings.

In earlier work we introduced eye movement analysis as a new modality for activity and context recognition [5, 6]. We showed that the movement patterns the eyes perform during different activities carry information that allows to recognise the activities themselves [6]. A large body of research in experimental psychology has evidenced that, in addition to physical activity, visual behaviour is tightly linked to cognitive processes, such as attention [26], relational memory [18], learning [21], or saliency determination [22]. This link to cognition makes eye movements a particularly promising source of information on the cognitive context of a person - beyond mere physical or visual activities.

To illustrate the vision of eye-based cognition-awareness consider the following scenario: Attendees of a business reception wear eye trackers that are unobtrusively embedded into their goggles. By analysing their eye movement patterns during conversations, cognition-aware memory assistants running on their mobile phones assess whether the involved speakers have met before and still remember each other. Using this information, the systems then automatically provide real-time memory assistance about people fallen into oblivion to prevent from embarrassing situations.

UbiComp'11, September 17–21, 2011, Beijing, China.

Although this example scenario is not novel per se, current state-of-the-art approaches use image processing techniques to detect whether two people have met before. The assistant envisioned here goes beyond mere detection of conversations or matching of faces to a database. Instead, it needs to detect whether a person actually *remembers* having seen somebody else before. For instance, let's assume that an attendee has met another person before. Consequently, a memory assistant using image processing would indicate that the other person is "known". The attendee, however, may still not remember that other person. Thus, in our scenario, the assistant is required to identify the *process of visual memory retrieval*. This can only be accomplished by extending the current notion of context with a cognitive dimension that reflects the attendee's subjective appraisal of the situation.

Paper Scope and Contributions

As a first step towards our vision of eye-based cognition-awareness, in this work we investigate the feasibility of using eye movement analysis to recognise visual memory recall processes of people looking at familiar pictures (that they have seen before) or unfamiliar pictures (that they see for the first time). The specific contributions are: (1) the introduction of cognition-awareness and the cognitive context as new paradigms in ubiquitous computing; (2) the introduction of eye movement analysis as a promising modality to infer processes of visual cognition, here memory recall, of a person; (3) the concept of a dual user study to investigate automatic recognition of one example cognitive process from eye movements, namely visual memory recall; and (4) the identification and discussion of the key challenges for the real-world implementation of cognition-aware systems.

RELATED WORK

Eye Movement Analysis

Eye movement analysis has long been used as a tool to investigate visual behaviour. In an early study, Hacisalihzade *et al.* used Markov processes to model visual fixations of observers looking at an object [17]. They transformed fixation sequences into character strings and used the string edit distance to quantify the similarity of eye movements. Elhelw *et al.* used discrete time Markov chains on sequences of temporal fixations to identify salient image features that affect the perception of visual realism [15]. They found that fixation clusters were able to uncover the features that most attract an observer's attention. Dempere-Marco *et al.* presented a method for training novices in assessing tomography images [11]. They modelled the assessment behaviour of domain experts based on the dynamics of their saccadic eye movements. Salvucci *et al.* evaluated means for automated analysis of eye movements [32]. They described three methods based on sequence-matching and hidden Markov models that interpreted eye movements as accurately as human experts but in significantly less time.

Eye Movements and Cognition

A growing number of researchers study eye movements in natural environments to better understand the role the visual system plays in the execution of everyday tasks [19]. Human vision research has shown that unconscious eye movements are strongly related to the underlying cognitive and perceptive processes. For example, it has been shown that visual behaviour is a good measure of visual engagement [34], drowsiness [33], and cognitive load [35]. Heisz *et al.* investigated changes in eye movement behaviour across several exposures to pictures of faces [21]. They found that as a face became more familiar, observers looked longer and more often at the eyes and less often at the nose, mouth, or forehead.

Differences in eye movement patterns are also linked to a number of mental disorders. It is for this reason that eye tracking has been investigated for the diagnosis of disorders on the autism spectrum (see [3] for a review). For example, Klin *et al.* showed that people with autism tend to show fewer fixations to the eyes but more fixations to the mouth [24]. Similar links were found for schizophrenia [16] as well as Parkinson's [28] and Alzheimer's disease [9].

All of these studies demonstrate the close link between visual behaviour and cognition and underline the potential of eye movement analysis for assessing the cognitive context of a person. While these studies analysed eye movements they were purely descriptive in nature. They did not attempt to automatically predict from the eye movements whether the object of attention, such as a face, was previously seen and remembered.

EYE MOVEMENT ANALYSIS

Wearable Eye Tracking

Developing sensors to track eye movements in daily life is an active topic of research. Portable video-based eye trackers - such as the Dikablis from Ergoneers or the iView X HED from SensoMotoric Instruments - require auxiliary equipment for the demanding video processing. The size of latest systems, such as the Glasses from Tobii Technology, is more appropriate for mobile settings, however, these eye trackers only allow for recordings over a couple of hours and do not (yet) provide real-time processing and output. Electrooculography (EOG) - the measurement technique used in this work - is an inexpensive method for mobile eye movement recordings; it is computationally light-weight and can be implemented using on-body sensors [27, 4]. These characteristics are crucial with a view to long-term eye movement recordings with real-time feedback in daily life.

Electrooculography

The eye can be modelled as a dipole with its positive pole at the cornea and its negative pole at the retina. Assuming a stable corneo-retinal potential difference, the eye is the origin of a steady electric potential field. The process of measuring changes in this field is called electrooculography. Using two pairs of skin electrodes placed at opposite sides of the eye and an additional reference electrode, two signal components (EOG_h and EOG_v), corresponding to two movement components - a horizontal and a vertical - can be identified. If the eye moves away from the centre position, the retina approaches one electrode while the cornea approaches the opposing one. This change in dipole orientation causes a change in the electric potential field and thus the measured

EOG signal amplitude. By analysing these changes, eye movements can be tracked.

Eye Movement Characteristics

To use eye movement analysis for context-awareness, it is important to understand the different types of eye movements. In earlier work we identified three types that can be robustly detected using EOG: saccades, fixations, and blinks.

Saccades

The eyes do not remain still when viewing a visual scene. Instead, they have to move constantly to build up a mental "map" from interesting parts of that scene. The main reason for this is that only a small central region of the retina, the fovea, is able to perceive with high acuity. The fast movement of the eyes is called a saccade. The duration of a saccade depends on the angular distance the eyes travel during this movement: the so-called saccade amplitude.

Fixations

Fixations are stationary states of the eyes during which gaze is held upon a specific location in the visual scene. Fixations can also be defined as the time between each two saccades.

Blinks

The frontal part of the cornea is coated with a thin tear film. To spread this fluid across the corneal surface, regular opening and closing of the eyelids, or blinking, is required. The average blink rate is influenced by environmental factors such as relative humidity, temperature or brightness, but also by physical activity [6], cognitive load [35], or fatigue [33].

EXPERIMENT

The experiment consisted of user studies: A main study to record natural eye movements during visual memory recall and a validation study to evaluate visual memory recall performance of a second group of participants. The main study was designed with two objectives in mind: (1) to elicit distinct eye movements by using a large screen and well-defined visual stimuli, and (2) to record natural visual behaviour without any active visual search or memory task by not asking participants for real-time feedback. The validation study followed the same protocol as the main study but participants were asked to provide real-time feedback on whether each image had previously been shown. While the main focus of the current work was on visual memory recall of faces, we included additional picture categories for comparison and to investigate the general applicability of our approach.

Apparatus

We created four picture sets, one set for each of the following picture categories: abstract images, landscapes, faces, and buildings. For the faces category, we manually selected 20 mixed-gender pictures (11 male, 9 female) from the CVL Face Database[1]. We chose pictures with frontal view and neutral expression; the face of the person was centred in the picture (see Figure 1). Pictures in the other categories were randomly selected from the Internet. We ensured, however, that these pictures had similar visual features. For example, we selected landscape photographs that showed a lake as their main feature; the building photographs always showed skyscrapers centred in the picture. The pictures were shown on a screen using a beamer resulting in a picture dimension of between 1x1 m and 1.5x1.5 m. A MATLAB script was used to control the display of the pictures as well as to ground truth annotate the experimental procedure.

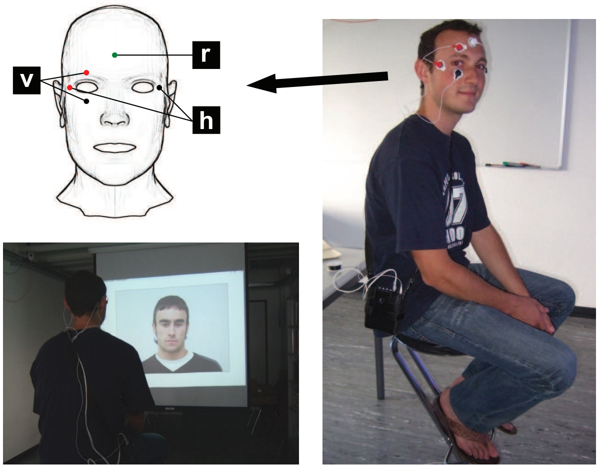

Figure 1: Experimental setup consisting of five electrodes for EOG data collection (h: horizontal, v: vertical, r: reference). The participants' eye movements were recorded while seated in front of a screen in a dimmed office room. No constraints with respect to movements of the upper body or head were imposed.

For EOG data collection we used a commercial system, the Mobi from Twente Medical Systems International (TMSI). The device records a four-channel EOG with a joint sampling rate of 128 Hz. The device was worn on a belt around each participant's waist and transmitted aggregated data via Bluetooth to a laptop placed behind the participant.

EOG signals were picked up using an array of five 24 mm Ag/AgCl wet electrodes from Tyco Healthcare placed around the right eye (see upper left picture in Figure 1). The horizontal signal was collected using one electrode on the nose and another directly across from this on the edge of the right eye socket. The vertical signal was collected using one electrode above the right eyebrow and another on the lower edge of the right eye socket. The fifth electrode, the signal reference, was placed away from the other electrodes in the middle of the forehead. Five participants (three male, two female) had to wear spectacles during the experiment. For these participants, the nose electrode was moved to the edge of the left eye socket to not interfere with the glasses frame. Data recording and synchronisation was handled by the Context Recognition Network (CRN) Toolbox [1].

[1] Database is available online: `http://lrv.fri.uni-lj.si/facedb.html`

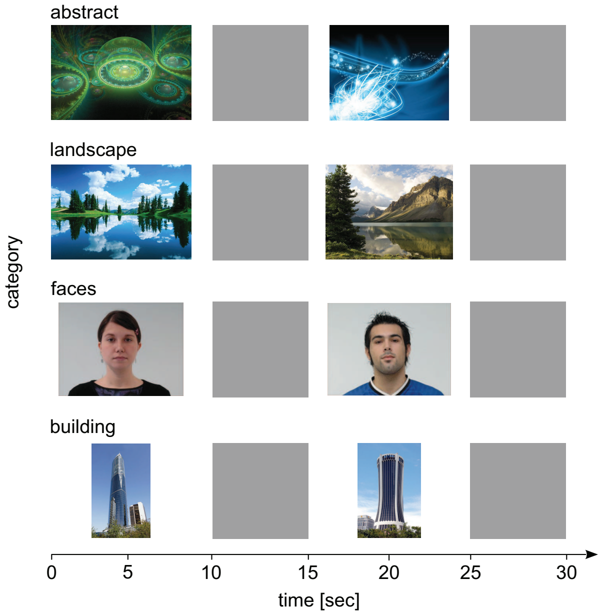

Figure 2: Example pictures from the four categories (abstract, landscape, faces, and buildings) and their sequence of display used in the experiment. Each picture was shown for 10 seconds; pictures with Gaussian noise were shown in between for five seconds.

Participants

We collected eye movement data from seven participants - four male and three female - recruited from the lab. Participants were between 25 and 29 years old (*M* = 26.4, *SD* = 1.6). Originally there were eight participants but one was withdrawn due to bad EOG signal quality that prevented robust detection of eye movements. We made sure all participants were well rested and we confirmed that none of them had seen pictures from the CVL database before.

Setup and Procedure

In contrast to [21], participants were only informed about the measurement equipment before the experiment; neither the purpose of the study nor any task were given to them. This was to not influence their visual behaviour by engaging them in an active visual search task. For the same reason, participants were also not asked to provide real-time feedback on whether they remembered each picture during the experiment. Participants were seated in a dimmed office room about 2 m in front of a screen facing its centre. Movements of the upper body were allowed at any time but we encouraged the participants to sit still throughout the experiment (see right picture in Figure 1).

Participants were asked to look at four continuous sequences of pictures. Each sequence was randomly created from pictures of a single picture category (see Figure 2). Within each sequence, 12 pictures were presented only once; five others were presented four times at regular intervals. This resulted in a total number of 32 pictures for each sequence. We randomised both the overall sequence as well as the selection of the repeated pictures across participants. In contrast to [21] we limited the exposure time for each picture to 10 seconds. In between each exposure, a picture with Gaussian noise was shown for five seconds as a baseline measurement and to allow the participants to relax. The total experiment time for each participant was about one hour. At the end of each experiment, the participants were asked on their experiences on the procedure in a short questionnaire.

Validation Study

The validation study used the same picture sets and experimental procedure as the main study. In contrast to the main study, the pictures were shown on a laptop and no eye movements were recorded from the participants. Instead, participants were asked for real-time feedback on whether or not each picture had been shown before by pressing two buttons on the laptop's keyboard. We collected feedback from seven participants disjunct to those of the main study - four male and three female - aged between 23 and 30 years (*M* = 26.9, *SD* = 2.9).

METHOD

All methods used in this work were implemented offline using MATLAB and C++. In this section we first describe the EOG signal processing algorithms and provide an overview of the extracted eye movement features (an in-depth description of these algorithms and features is outside the scope of this paper but can be found in [6]). We then introduce the feature selection and classification algorithms, as well as the parameter selection and training procedures.

Noise and Baseline Drift Removal

EOG_h and EOG_v were first stripped of high frequency noise using a median filter. For baseline drift removal, we then performed an approximated multilevel 1-D wavelet decomposition at level nine using Daubechies wavelets on each signal component. The reconstructed decomposition coefficients gave a baseline drift estimation. Subtracting this estimation from the original signals yielded the corrected signals with reduced drift offset (see [36] for further details).

Saccade Detection

In an earlier work we introduced the *Continuous Wavelet Transform - Saccade Detection* (CWT-SD) algorithm [6]. Briefly, CWT-SD detects saccades by thresholding on the continuous 1-D wavelet coefficient vector computed from the de-noised and baseline drift removed EOG_h and EOG_v. CWT-SD takes physiological saccade characteristics into account to increase the robustness of detection [14].

Fixation Detection

Our algorithm for fixation detection exploits the fact that fixation points tend to cluster together closely in time. Thus, by thresholding on the dispersion of these points, fixations can be detected [38]. Based on the output of the CWT-SD algorithm, dispersion and duration values are calculated for each non-saccadic segment. If the dispersion is below a maximum threshold, and the duration above a minimum threshold, a fixation is detected.

Blink Detection

Similar to the algorithm for saccade detection, the *Continuous Wavelet Transform - Blink Detection* (CWT-BD) algorithm uses thresholding of wavelet coefficients to detect blinks in EOG_v. In contrast to saccades, a blink is characterised by a short sequence of two large peaks in the coefficient vector: one positive, the other negative. The time between these peaks is much smaller than for saccades. Thus, blinks are distinguished from saccades by applying a maximum threshold on this time difference.

Analysis of Saccade Sequences

Visual attention while looking at faces is strongly attracted by internal face features (eyes, nose, and mouth) as these convey crucial information about face identity. When viewing unfamiliar faces observers were found to scan more diverse regions of a face compared to familiar faces [21]. This suggests that eye movement features that capture the sequential nature of visual scanning behaviour contain useful information for recognising memory recall processes.

To extract information about saccade sequences we used a wordbook analysis. First, each saccade was encoded into a discrete, character-based representation. A sliding window of length l and a step size of one was used to scan the stream of encoded saccades for saccade sequences. A saccade sequence is given by l successive characters. As an example with $l = 4$, the sequence "LrBd" translates to large left (L) → small right (r) → large diagonal right (B) → small down (d). These sequences were then collected in wordbooks and analysed statistically. Each new sequence was added to the corresponding wordbook Wb_l; for a sequence already included in Wb_l its occurrence count was increased by one.

Feature Extraction and Selection

We considered the two-class recognition problem of discriminating between pictures that were only seen once (class "non-repeated") and pictures that were seen several times (class "repeated") by the participants. We first removed all eye movement data that belonged to Gaussian noise pictures. We then assigned all picture instances (picture and corresponding eye movement data) of all first and single exposures to the "non-repeated" class (17 picture instances), and picture instances of exposures two, three, and four to the "repeated" class (15 picture instances).

Feature extraction was run on all picture instances separately using a sliding window with window size W_{fe} and step size S_{fe}. We extracted four groups of features from the detected saccades, fixations, blinks, and wordbooks (see Table 1). The features were calculated on both EOG_h and EOG_v. Features calculated from saccadic eye movements made up the largest proportion of extracted features. In total, there were 62 such features comprising the mean, variance and maximum EOG signal amplitudes of saccades, and the normalised saccade rates. These were calculated for both EOG_h and EOG_v; for small and large saccades; for saccades in positive or negative direction; and for all possible combinations of these. We calculated five different fixation features: the mean and variance of the EOG signal amplitude within a fixation; the mean and the variance of fixation duration; and the fixation rate over window W_{fe}. For blinks, we extracted three features: blink rate, as well as the mean and variance of the blink duration. We used four wordbooks. This allowed us to account for all possible eye movement patterns up to a length of four ($l = 4$), with each wordbook containing the type and occurrence count of all patterns found. For each wordbook we extracted five features: the wordbook size, the maximum occurrence count, the difference between the maximum and minimum occurrence counts, and the variance and mean of all occurrence counts. The resulting feature matrices were combined to one large feature matrix per participant, each comprising 32 picture instances.

Group	Features
saccade (*S*-)	mean (mean), variance (var) or maximum (max) EOG signal amplitudes (Amp) or rate (rate) of small (S) or large (L), positive (P) or negative (N) saccades in horizontal (Hor) or vertical (Ver) direction
fixation (*F*-)	mean (mean) and/or variance (var) of the horizontal (Hor) or vertical (Ver) EOG signal amplitude (Amp) within or length (Length) of a fixation or rate of fixations
blink (*B*-)	mean (mean) or variance (var) of the blink duration or blink rate (rate)
wordbook (*W*-)	wordbook size (size) or maximum (max), difference (diff) between maximum and minimum, mean (mean) or variance (var) of all occurrence counts (Count) in the wordbook of length (-lx)

Table 1: Naming scheme for the features used in this work. For a particular feature, e.g. *S-rateSPHor*, the capital letter represents the group - saccadic (S), blink (B), fixation (F) or wordbook(W) - and the combination of abbreviations after the dash describes the particular type of feature and the characteristics it covers.

For feature selection we chose a filter scheme over the commonly used wrapper approaches because of the lower computational costs and thus shorter runtime. We use minimum redundancy maximum relevance feature selection (mRMR, [31]). The mRMR algorithm selects a feature subset of arbitrary size S that best characterises the statistical properties of the given target classes based on the ground truth labelling (see [30] for the MATLAB implementation we used).

Classification and Performance Evaluation

For classification we chose a linear support vector machine (SVM, see [25] for the C implementation we used). All parameters of the saccade, fixation, and blink detection algorithms were fixed to values common to all participants. For evaluation we followed a leave-one-person-out scheme: the datasets of all but one participant were combined and used for training (the "training set"); the dataset of the remaining participant was used for testing (the "test set"). This was repeated for each participant. Feature selection was performed

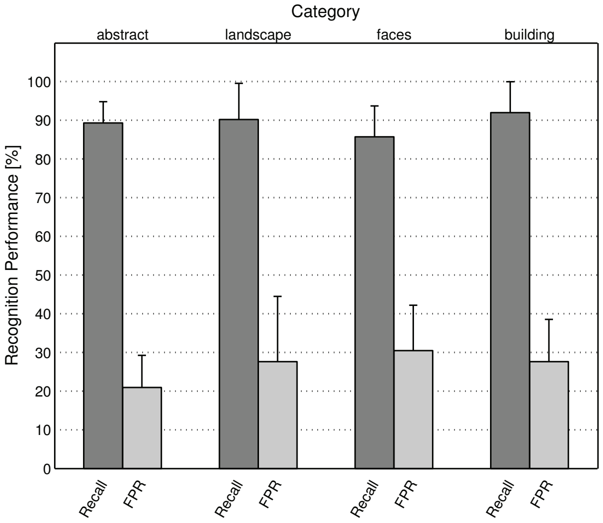

Figure 3: Mean recall and false positive rate (FPR) for the different picture categories using an exposure time $T_{exp} = 10s$. Error bars denote 95% confidence intervals.

solely on the training set. During the classification process the size of the feature set for each leave-one-person-out iteration was optimised with respect to accuracy by sweeping over S, W_{fe}, and S_{fe}. In addition, the prediction vector returned by the classifier for each picture instance was reduced to a single class label ("repeated" or "non-repeated") using majority voting.

RESULTS

Results for Each Picture Category

On average, participants from the validation study were able to correctly identify pictures that had previously been shown with an accuracy of 97.3% for abstract, 97.8% for landscape, 96.4% for faces, and 97.3% for building pictures. Given this high accuracy, in the following analysis we assume that participants in the main study perfectly remember a picture that was shown before.

Based on the data recorded in the main study, Figure 3 summarises the overall recognition performance using person-independent parameters and training for each picture category. The bars contrast true positive rate (recall) ($\frac{TP}{TP+FN}$) to false positive rate (FPR) ($\frac{FP}{FP+TN}$), where TP, FP, TN and FN represent true positive, false positive, true negative and false positive counts, respectively. The figure shows that the recognition system consistently achieves recall values above 85% with the highest recall of 92% for the building (FPR: 27.6%) and the lowest recall of 85.7% for the faces picture category. The FPR is above 20% for all picture categories with the highest FPR of 30.5% for the faces category and the lowest FPR of 21.0% for abstract pictures (recall: 89.3%).

The results for each individual participant show a range of differences in recognition performance (see Table 2). For example, the highest recall result for the faces category is 93.8% (participants 3 and 4) but with a FPR of 46.7%. The worst result was for participant 6, with 68.8% recall but a FPR of only 13.3%. What can be seen from the table, however, is that for all categories the differences do not seem to correlate to the gender of the person.

Further Analysis of the Faces Picture Category

With a view to the cognition-aware memory assistant outlined in the introduction we then analysed the results for the faces picture category in more detail.

Eye Movement Features

We first analysed how mRMR ranked the features on each of the seven leave-one-person-out training sets for the faces category. The rank of a feature is the position at which mRMR selected it within a set. The position corresponds to the importance with which mRMR assesses a feature's ability to discriminate between classes in combination with the features selected before it. Figure 4 shows the top 15 features according to the median rank over all sets (see Table 1 for a description of the type and name of the features). For each feature the vertical bar represents the spread of mRMR ranks for the seven training sets. The most useful features are those found with the highest rank (close to one) for most training sets, indicated by shorter bars. Note that some features are not always included in the final result (e.g. feature 64 only appears in four sets). Equally, a useful feature that is ranked lowly might still improve a classification (e.g. feature 89 is spread between rank two and 21, but is included in six sets).

This analysis reveals that most features are based on horizontal and vertical saccades, e.g. 39 (*sacc-EMRt*), 24 (*sacc-varAmpLPHor*), and 62 (*sacc-propHorVer*). Feature 67 (*fix-varLength*, variance of fixation duration) is used by six sets, four of which rank it highly. Feature 65 (*bl-varLength*, variance of blink duration) is selected for six out of the seven sets, all of which give it a high rank. Three wordbook features are in the top ranks, all of which were selected at least four times and describe eye movements sequences of length three and four: 89 (*str-varCount-l4*), 83 (*str-diffCount-l3*), and 88 (*str-diffCount-l4*).

Analysis of Different Exposure Times

The analysis so far assumed a fixed exposure time $T_{exp} = 10s$. A short exposure time is desired as this directly translates to a low latency of the recognition system. In addition, 10 seconds may be regarded as rather long for real-world environments considering that looking at the faces of others, for example while walking down a street, is a subtle visual activity and may occur on a smaller time scale. In a laboratory setting, Hsiao *et al.* found that the best performance for face recognition was achieved with only two fixations and that performance did not improve with additional fixations [23]. Based on the fixation durations reported there and typical physiological saccade characteristics, two fixations correspond to about one second of exposure.

To analyse the influence of T_{exp} on the recognition performance we sweeped $T_{exp} = 10s, 5s, 3s, 2s, 1s$. For example for $T_{exp} = 3s$, we only used the eye movement data

picture category		P1 (m)	P2 (m)	P3 (f)	P4 (m)	P5 (f)	P6 (m)	P7 (f)	mean	std
abstract	Recall [%]	*81.3*	87.5	87.5	93.8	**100.0**	87.5	87.5	89.3	5.9
	FPR [%]	**6.7**	20.0	26.7	20.0	26.7	*33.3*	13.3	21.0	9.0
landscape	Recall [%]	93.8	**100.0**	93.8	93.8	93.8	87.5	*68.8*	90.2	10.1
	FPR [%]	40.0	33.3	*46.7*	13.3	**6.7**	*46.7*	**6.7**	27.6	18.2
faces	Recall [%]	87.5	87.5	**93.8**	**93.8**	87.5	*68.8*	81.3	85.7	8.6
	FPR [%]	26.7	20.0	*46.7*	*46.7*	26.7	**13.3**	33.3	30.5	12.7
building	Recall [%]	93.8	*75.0*	93.8	93.8	**100.0**	87.5	**100.0**	92.0	8.6
	FPR [%]	20.0	**13.3**	20.0	26.7	26.7	40.0	*46.7*	27.6	11.8

Table 2: Recall and false positive rate (FPR) for each participant and the mean and standard deviation over all. The table also shows the participants' gender (f: female, m: male). Best case results for each picture category are indicated in bold, worst case results in italic.

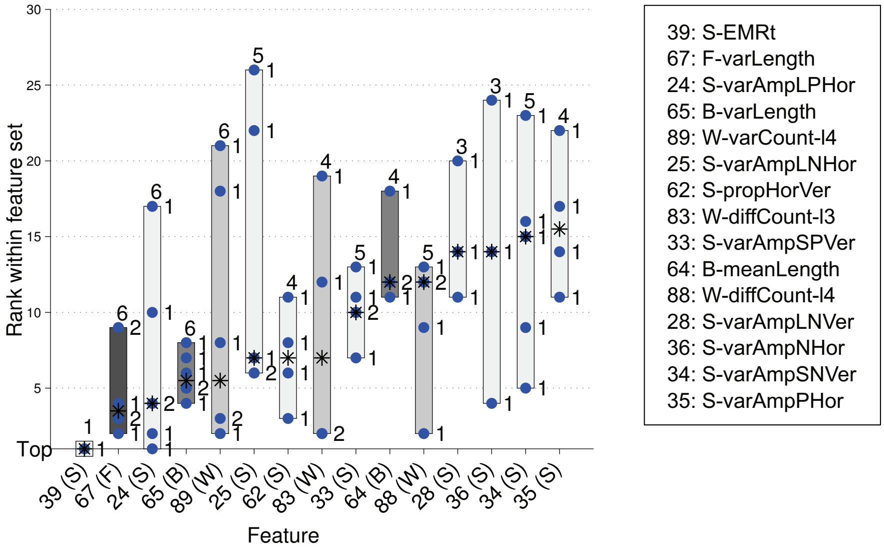

Figure 4: Top 15 eye movement features selected by mRMR for all seven training sets for the faces picture category. X-axis shows feature number and group; the key on the right shows the corresponding feature names as described in Table 1; Y-axis shows the rank (top = 1). For each feature, the bars show: the total number of training sets for which the feature was chosen (bold number at the top), the rank of the feature within each set (dots, with a number representing the set count), and the median rank over all sets (black star). For example, a useful feature is 67 (F) - a fixation feature selected for six sets, in four of which it is ranked four or below; less useful is 28 (S) - a saccade feature used in only three sets and ranked between 11 and 20.

recorded during the first three seconds of each exposure for classification. We calculated the recognition performance again using person-independent parameters and training. As can be seen from Figure 5 the exposure time considerably influences the recognition performance. For $T_{exp} = 10s$ the mean recall is 85.7% (FPR: 30.5%), whereas for $T_{exp} = 1s$ the recall drops to 65.2% with a FPR of 34.3%. It is interesting to note that $T_{exp} = 5s$ yields the best FPR of 21.9% (recall: 77.7%) and while $T_{exp} = 3s$ yields the worst FPR of 38.1% the recall only decreases to 83.9% compared to an exposure time of 10 seconds.

DISCUSSION

Eye Movement Features

The mRMR-based feature ranking provides interesting insights into the type of eye movement features that are useful for assessing visual memory recall of faces (see Figure 4). The saccade and wordbook feature groups were particularly well represented in the study. This result confirms that in addition to eye movement characteristics well-known in experimental psychology, such as the mean fixation duration, saccades and saccade sequences carry useful information on a person's visual behaviour during memory recall tasks. Similar to earlier results for eye-based activity recognition [6] the best recognition performance was only achieved using a mixture of features from different feature groups.

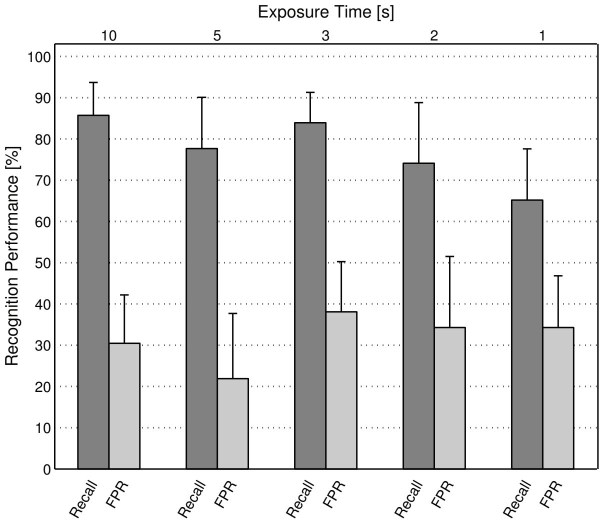

Figure 5: Mean recall and false positive rate (FPR) for different exposure times for the faces picture category. Error bars denote 95% confidence intervals.

Blink features were less well represented in the top ranks. This is most likely due to the short exposure time of only 10 seconds for which participants could not show much blink rate variations as for example while looking at unpleasant visual stimuli [29]. Blink features may be found to be more discriminative for longer duration activities, such as prolonged visual search. In combination with the ease by which they were calculated, we believe that blink features are still promising for future work on cognition-aware systems.

The same applies to the group of fixation features from which only one feature, the variance of the fixation duration, was selected in top ranks for six of the seven sets. It is interesting to note that the mean fixation count - a feature found in [21] to significantly change across exposures - was not among the top 15 features in this study.

Recognition Performance

All parameters of the saccade, fixation and blink detection algorithms were fixed to values common to all participants; the same applies to the parameters of the feature selection and classification algorithms. Using person-independent training, our recognition system achieved best average recall values of between 85.7% and 92%, and FPR of between 21% and 30.5% for all four picture categories (see Figure 3). While such a high FPR may not be problematic for certain use cases, lower FPR are generally preferable and could, for example, be achieved by using person-dependent parameters and training.

The analysis of different exposure times revealed that while a one second exposure yielded a recall well above chance, the recall for three seconds was close to that of a full 10 second exposure (see Figure 5). This lower bound is important for a potential real-world implementation of a cognition-aware memory assistant and may be further lowered by using eye movement features and analysis methods particularly geared towards recognition of visual memory recall.

Additional eye movement characteristics - such as pupil dilation or microsaccades - that are potentially useful for recognising visual memory recall or the other cognitive processes mentioned before were not used here because of the difficulty in measuring them with EOG. These characteristics are still worth investigating in the future as they may carry information that complements that available in the current work.

Experiment

To investigate the feasibility of inferring cognitive processes from eye movements requires an experimental methodology that is more similar to that used in experimental psychology rather than in ubiquitous computing. Human vision experiments typically involve controlled tasks with carefully designed and timed visual stimuli. In contrast, ubiquitous computing aims for real-world applications that involve unconstrained natural behaviour. Techniques from experimental psychology can therefore not directly be adopted for ubiquitous computing. As demonstrated here, a viable approach is to first evoke and infer specific cognitive processes in a similar but less controlled laboratory setting. The experimental design and procedures can then be transferred and gradually extended to cover more complex daily life situations.

In the experiment we faced a trade-off between natural visual behaviour and user-annotated ground truth. We addressed this trade-off by implementing a dual study design with two different groups of participants. In the main study we did not ask participants for real-time feedback on whether they actually remembered each picture. This was to not involve them in an active visual memory task that would have influenced their visual behaviour. In terms of evaluation, this lack of user-annotated ground truth required the assumption that the ground truth labels defined by the experimental procedure reflected the participants' subjective experience. The validation study showed that this assumption may not always hold, i.e. some pictures had been shown before but were not remembered by the participants. Participants' consistently high memory recall performance for all picture categories in the validation study also showed, however, that this occurred in only about 3% of the cases and had negligible influence on the participants' performance in the main study. We plan to compare the dual study design introduced in the current work with other validation techniques such as post-experiment questionnaires or analysis of video footage.

Toward a Real-World Implementation

While the initial results presented here are promising, there are several challenges that we aim to address in future work for a real-world implementation of the envisioned cognition-aware memory assistant.

One of the key challenges for a real-world implementation is the co-influence of (visual) task, situation, and cognitive processes on a person's eye movements. In laboratory settings these influences can be minimised by using a constrained experimental setup and well-defined visual stimuli. Everyday settings can typically not be controlled in a similar fashion. It is therefore crucial to identify and separate these different sources of influence for robust recognition of visual memory recall and other cognitive processes. This problem could be addressed by using a multi-modal approach for context recognition and annotation that incorporates additional modalities to eye tracking, such as proximity sensors, GPS for localisation, inertial measurement units for head movements, or eye contact sensors [13].

This leads to a second challenge. Personal encounters in daily life differ considerably from the situation investigated here. In these settings, facial expressions of conversational partners change continuously, the viewpoint is dynamic, and other visual stimuli may attract attention and lead to "random" saccades to other entities in the surrounding environment. In addition, personal encounters may range from longer face-to-face discussions between two people, over glances to faces of others while in transit, to looking at several faces of a group of people in succession. This will require advanced methods for robust detection of when and how people look at each other's face. One part of a solution to this problem is to augment the analysis of eye movement dynamics - as presented here - with a computer vision system for face detection and a wearable gaze tracker to identify the points in time the person has looked at a face.

The current experiment involved participants to look at a large screen to provoke distinct eye movements that could easily be measured using EOG. It remains to be investigated whether current wearable eye trackers - whether EOG- or video-based - are accurate and robust enough to capture eye movement characteristics that reflect visual memory recall processes on smaller screens (e.g. on a mobile phone) or with the person being in transit.

In the questionnaire participants one and two reported that they got slightly bored and participants one, two, and five reported of getting slightly tired while looking at the pictures. Participants one, three, four, and seven also reported of having lost concentration towards the end of the experiment. These phenomena did not seem to have influenced the recognition performance in the current experiment (see Table 2). Given that blink duration was found to be a reliable indicator of fatigue [7] and that its variance was a high-ranked features for six of the seven training sets (see Figure 4) we still believe that a fatigue detector may prove beneficial for increasing the performance of a real-world implementation of a cognition-aware memory assistant.

Finally, it will be interesting to see how social conventions, such as not to look into the eyes of others for too long or not to scrutinise the faces of others, will influence the available time to analyse eye movements and therefore the recognition performance of the memory assistant.

CONCLUSION

In this work we proposed cognition-awareness as a novel paradigm to describe context-aware computing systems that are able to sense and adapt to a user's cognitive context. We introduced eye movement analysis as a promising method to assess the cognitive context in an unobtrusive manner. Using a dual user study we showed that visual memory recall processes while looking at familiar and unfamiliar pictures can be recognised from eye movements of seven participants with decent performance. These initial results are promising as the described approach may soon be applicable to other stationary real-world setups, e.g. in a cognition-aware picture browser on a desktop computer. They also open up the discussion on the wider applicability of the approach to other cognitive processes and mobile daily life settings.

REFERENCES

1. D. Bannach, P. Lukowicz, and O. Amft. Rapid Prototyping of Activity Recognition Applications. *IEEE Pervasive Computing*, 7(2):22–31, 2008.

2. N. Bigdely-Shamlo, A. Vankov, R. R. Ramirez, and S. Makeig. Brain activity-based image classification from rapid serial visual presentation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 16(5):432 –441, 2008.

3. Z. Boraston and S.-J. Blakemore. The application of eye-tracking technology in the study of autism. *Journal of Physiology*, 581(3):893 –898, 2007.

4. A. Bulling, D. Roggen, and G. Tröster. Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments. *Journal of Ambient Intelligence and Smart Environments*, 1(2):157–171, 2009.

5. A. Bulling, J. A. Ward, and H. Gellersen. Multi-Modal Recognition of Reading Activity in Transit Using Body-Worn Sensors. *ACM Transactions on Applied Perception*, 2011, to appear.

6. A. Bulling, J. A. Ward, H. Gellersen, and G. Tröster. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):741–753, 2011.

7. P. P. Caffier, U. Erdmann, and P. Ullsperger. Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89(3-4):319–325, 2003.

8. M. J. Chadwick, D. Hassabis, N. Weiskopf, and E. A. Maguire. Decoding individual episodic memory traces in the human hippocampus. *Current Biology*, 20(6):544 – 547, 2010.

9. T. J. Crawford, S. Higham, T. Renvoize, J. Patel, M. Dale, A. Suriya, and S. Tetley. Inhibitory control of saccadic eye movements and cognitive impairment in alzheimer's disease. *Biological Psychiatry*, 57(9):1052–1060, 2005.

10. N. Davies, D. P. Siewiorek, and R. Sukthankar. Special issue: Activity-based computing. *IEEE Pervasive Computing*, 7(2), 2008.

11. L. Dempere-Marco, X. Hu, S. L. S. MacDonald, S. M. Ellis, D. M. Hansell, and G.-Z. Yang. The use of visual search for knowledge gathering in image decision support. *IEEE Transactions on Medical Imaging*, 21(7):741–754, 2002.

12. A. K. Dey. Understanding and using context. *Personal and Ubiquitous Computing*, 5(1):4–7, 2001.

13. C. Dickie, R. Vertegaal, J. S. Shell, C. Sohn, D. Cheng, and O. Aoudeh. Eye contact sensing glasses for attention-sensitive wearable video blogging. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 769–770, 2004.

14. A. T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

15. M. Elhelw, M. Nicolaou, A. Chung, G.-Z. Yang, and M. S. Atkins. A gaze-based study for investigating the perception of visual realism in simulated scenes. *ACM Transactions on Applied Perception*, 5(1):1–20, 2008.

16. U. Ettinger, M. Picchioni, M.-H. Hall, K. Schulze, T. Touloupoulou, S. Landau, T. J. Crawford, and R. M. Murray. Antisaccade performance in monozygotic twins discordant for schizophrenia: The maudsley twin study. *American Journal of Psychiatry*, 163(3):543–545, 2006.

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

18. D. E. Hannula and C. Ranganath. The eyes have it: Hippocampal activity predicts expression of memory in eye movements. *Neuron*, 63(5):592 – 599, 2009.

19. M. M. Hayhoe and D. H. Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4):188–194, 2005.

20. J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris. Out of the lab and into the fray: Towards modeling emotion in everyday life. In *Proc. of the 8th International Conference on Pervasive Computing*, pages 156–173, 2010.

21. J. J. Heisz and D. I. Shore. More efficient scanning for familiar faces. *Journal of Vision*, 8(1):1–10, 2008.

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

23. J. H.-w. Hsiao and G. Cottrell. Two fixations suffice in face recognition. *Psychological Science*, 19(10):998–1006, 2008.

24. A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen. Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, 59(9):809–816, 2002.

25. C.-J. Lin. LIBLINEAR - a library for large linear classification, 2008. `http://www.csie.ntu.edu.tw/~cjlin/liblinear/`.

26. S. P. Liversedge and J. M. Findlay. Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4:6–14, 2000.

27. H. Manabe and M. Fukumoto. Full-time wearable headphone-type gaze detector. In *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems*, pages 1073–1078, 2006.

28. U. P. Mosimann, R. M. Müri, D. J. Burn, J. Felblinger, J. T. O'Brien, and I. G. McKeith. Saccadic eye movement changes in Parkinson's disease dementia and dementia with Lewy bodies. *Brain*, 128(6):1267–1276, 2005.

29. D. Palomba, M. Sarlo, A. Angrilli, A. Mini, and L. Stegagno. Cardiac responses associated with affective processing of unpleasant film stimuli. *International Journal of Psychophysiology*, 36(1):45 – 57, 2000.

30. H. Peng. mRMR Feature Selection Toolbox for MATLAB, 2007. `http://research.janelia.org/peng/proj/mRMR/`.

31. H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005.

32. D. D. Salvucci and J. R. Anderson. Automated eye-movement protocol analysis. *Human-Computer Interaction*, 16(1):39–86, 2001.

33. R. Schleicher, N. Galley, S. Briest, and L. Galley. Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? *Ergonomics*, 51(7):982 – 1010, 2008.

34. J. Skotte, J. Nøjgaard, L. Jørgensen, K. Christensen, and G. Sjøgaard. Eye blink frequency during different computer tasks quantified by electrooculography. *European Journal of Applied Physiology*, 99(2):113–119, 2007.

35. E. Stuyven, K. V. der Goten, A. Vandierendonck, K. Claeys, and L. Crevits. The effect of cognitive load on saccadic eye movements. *Acta Psychologica*, 104(1):69 – 85, 2000.

36. M. A. Tinati and B. Mozaffary. A wavelet packets approach to electrocardiograph baseline drift cancellation. *International Journal of Biomedical Imaging*, Article ID 97157, 2006.

37. R. Want, A. Hopper, V. Falcão, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems*, 10:91–102, 1992.

38. H. Widdel. *Theoretical and Applied Aspects of Eye Movement Research*, chapter Operational problems in analysing eye movements, pages 21–29. Elsevier, 1984.