Review of eye tracking metrics involved in emotional and cognitive processes

Vasileios Skaramagkas¹, Giorgos Giannakakis¹, Member, IEEE, Emmanouil Ktistakis¹,³, Dimitris Manousos¹, Ioannis Karatzanis¹, Nikolaos S. Tachos⁴, Member, IEEE, Evangelia Tripoliti⁴, Member, IEEE, Kostas Marias¹,⁶, Member, IEEE, Dimitrios I. Fotiadis⁵, Fellow, IEEE, and Manolis Tsiknakis¹,⁶, Member, IEEE

Abstract—Eye behaviour provides valuable information revealing one's higher cognitive functions and state of affect. Although eye tracking is gaining ground in the research community, it is not yet a popular approach for the detection of emotional and cognitive states. In this paper, we present a review of eye and pupil tracking related metrics (such as gaze, fixations, saccades, blinks, pupil size variation, etc.) utilized towards the detection of emotional and cognitive processes, focusing on visual attention, emotional arousal and cognitive workload. Besides, we investigate their involvement as well as the computational recognition methods employed for the reliable emotional and cognitive assessment. The eye-tracking publicly available datasets employed in relevant research efforts were concentrated and described their specifications and details. The multimodal approaches which combine eye-tracking features with other modalities (e.g. biosignals), along with artificial intelligence and machine learning techniques were also surveyed in terms of their recognition/classification accuracy. The limitations, current open research problems and prospective future are discussed for the usage of eye tracking as the primary sensor modality. This study aims to comprehensively present the most robust and significant eye/pupil metrics based on available literature towards the development of a robust emotional or cognitive computational model.

Index Terms—eye tracking, gaze, pupil, fixations, saccades, smooth pursuit, blinks, stress, visual attention, emotional arousal, cognitive workload, emotional arousal datasets, cognitive workload datasets

I. INTRODUCTION

The investigation of emotional and cognitive processes that modulate human behaviour requires a comprehensive research approach. Various psychophysiological and psychophysical modalities (such as electroencephalography (EEG), event-related potentials (ERP), electrodermal activity (EDA), electrocardiography (ECG), facial expressions, body posture, etc.) have been employed in the relevant emotion-related literature. The eye and pupillary response in relation to emotional or cognitive processing provides valuable information for one's higher cognitive function and state of affect [1], [2], [3]. However, there is not a concrete comprehensive guide of the utilization of eye and pupil behaviour towards this objective.

Over the last years, there is an increasing research effort in the area of emotional and cognitive functions. A significant portion of this research is based on neurophysiological data, investigating the pattern and behaviour of the implicated neural networks. Another approach is the investigation of the human body’s physiological and physical measures (biosignals). Eye and pupil behavioural measures, although been mediated by the autonomic nervous system (ANS) just like biosignals, have remained in the research background [4].

Lately, the evolution of eye tracking hardware/software enables their usage on convenient wearable devices boosted the related eye tracking research. Robust eye trackers, in terms of accuracy, portability and ease of use, have been developed, that are able to unobtrusively monitor eye movements in real-time. Among the different eye tracking systems, the head-mounted type has become the most popular since they are used in daily indoor/outdoor activities. In addition, various computational algorithms have been developed to efficiently extract metrics associated with the behaviour of the eye.

To our knowledge, a comprehensive guide of specific eye patterns during cognitive and emotional processing does not exist. Most related studies apply machine learning techniques to the multivariate ocular feature set in order to categorize the data into predefined user’s states.

In this manuscript, the ocular features are investigated in the context of user’s emotional arousal, visual attention and cognitive workload. Emotional arousal is a fundamental concept in emotion theory, which is involved in the majority of affective experiences. It constitutes one of the two main axes in Russell’s two-dimensional circular space of emotions [5]. In this perspective, the level of arousal level is the structural el-
ment contributing, to a lesser or greater degree, to all emotions. Along the same line, visual attention and cognitive workload are significant indices of the human cognitive function and performance. Reduced performance may imply a deficiency of information processing. This may be caused by a limited pool of resource (cognitive capacity), or by a more conservative response under higher cognitive workloads (response caution) [6].

According to the literature, among the different eye movement metrics, there can be an initial categorization between those that are most correlated with visual attention, those that are more relevant with emotional arousal and those that are best indicators for the cognitive workload. However, there are no specific or unique eye features that provide a discrimination capability among visual, emotional and cognitive processes. For this reason, the collection of the various eye features and their positive or negative influence on an individual’s emotional or cognitive state is vital.

In the present review, we investigate eye and pupil behaviour related metrics that best describe and express the processes of emotional arousal, visual attention and cognitive workload. Firstly, we determine the scope of these three emotional/cognitive processes (section II) and we describe the nature and the corresponding underlying physiological functions of the eye movements and pupil behaviour features (section III). A short presentation of eye-tracker systems focusing on the most widely used, the video-based trackers and the techniques of gaze estimation (Section IV) is being performed. Then, this paper’s main scope is described which is the investigation and the association of eye/pupil metrics with the three emotional/cognitive processes in urban environments and daily life activities (section V). Besides, we provide comprehensive information on the emotional/cognitive recognition methods either only using eye metrics or using multimodal approaches (section VI). Finally, publicly available eye behaviour datasets are presented (section VII). The fields of research reviewed relate to eye metrics involved in emotional/cognitive processes (e.g. emotional arousal, visual attention, cognitive workload) including basic research in order for the reader to be able to recognize the most robust features that can be utilized on her/his research. The selection of the studies is based on the scope of our review and the investigation of the correlation of eye metrics with affective and cognitive processes in urban environments and daily life activities such as driving, reading, working on a PC etc.

The main aim of this review is to

- identify the eye metrics that have a significant relation to the investigated emotional arousal, cognitive workload, visual attention processes
- determine efficient recognition methods based on eye metrics for the investigated emotional arousal, cognitive workload, visual attention processes
- specify the most robust and relevant ocular combined feature set in a multimodal approach.

This analysis expects to aid related future experimental design and research towards an efficient selection of eye metrics.

II. EMOTIONAL AND COGNITIVE PROCESSES

In natural vision cognitive and affective factors influence an observer’s visual attention. Cognitive and emotional factors play a dominant role in active gaze control and can determine attention allocation in complex scenes [7]. It is known that emotional arousal modifies the allocation of attentional resources [8] but also conversely, how attention is allocated during emotional arousal can significantly alter an emotional state [9]. Both emotion and attention seem to modulate both early and late stages of visual processing [8].

In this section, the emotional and cognitive processes are defined in detail and their specific characteristics are presented. Visual attention is first presented and then, emotional arousal and cognitive workload as factors that affect visual attention.

A. Visual attention

The process by which a user selects a specific element from all the available information in order to further examine is called visual attention. In other words, the term “visual attention” refers to the collection of various cognitive operations that isolate the relevant from irrelevant information from cluttered visual scenes [10]. Attention remains a crucial area of investigation within education, psychology, cognitive neuroscience, and neuropsychology [11]. In recent years, active research is being carried out to determine the source of sensory and attention-triggering signals, the effects of these sensory points on the coordination properties of the sensory neurons as well as the relationship between attention and other behavioural and cognitive processes including memory and psychological vigilance.

B. Emotional arousal and stress

Emotional arousal is a state that describes the level of calmness (i.e., low arousal) or excitation (i.e., high arousal) elicited by a stimulus. In [12], arousal is defined as a global feeling of dynamism or lethargy that involves mental activity and physical preparedness to act. The most common manifestation of increased arousal and negative valence is stress. There are various dimensional models of affect, the most known of which is the circumplex model of Russell [5], which maps emotions along predefined axes. The physiological stress response modulates, among other body functions, eye behaviour and function [13]. Previous studies have investigated the relationship between the level of emotional arousal and response inhibition [14].

C. Cognitive workload

The cognitive workload is defined as the level of an individual’s measurable mental effort in order to cope with one or more cognitively demanding tasks [15]. According to cognitive load theory, there are three types of cognitive load: 1) intrinsic load 2) extraneous load and 3) germane load [16]. The intrinsic load comes from the complexity of the task and its association with the user, extraneous load is caused by the presentation style of the material and germane load refers to the ability of the user to fully understand the material [17].
Although the cognitive workload is considered to be a subjective personality property, cognitive load can be partially demarcated under three quantified measurement types:

- performance
- subjective
- physiological.

Performance measurements assess workload through the ability of a user to perform tasks or functions of a system. Subjective measurements are based on the judgments of the users regarding the workload associated with the execution of a task or a system function. Physiological measurements evaluate the physiological responses of the user during specific task demands [18].

III. FEATURES REPRESENTING EYE AND PUPIL BEHAVIOUR

In this section, some of the most important features that represent the behaviour of eye motion are presented and discussed.

A. Visual fixations

Ocular fixation is defined as the maintenance of eye gaze on a single location [19]. Human beings can fixate only when they possess a fovea in the anatomy of the eye. A fovea is typically located at the center of macula lutea of the retina and dictates the point of clearest vision [20], [21]. In recent years, following the development of eye trackers, the detection of fixation related metrics has become more robust and easier to implement. The most commonly used fixation metrics used are: number of fixations, number of fixations on each area of interest, total number of fixations, fixation duration, total fixation duration i.e. the cumulative duration of all of the fixations on a particular area of interest (AOI), time to first fixation on target, fixation density and repeat fixations [22].

![Eye anatomy responsible for the eye movements. The superior and inferior rectus muscles are responsible for the eye's vertical movements, whereas the lateral and medial rectus muscles control horizontal movements. (b) Function and implicated muscles of pupil dilation and contraction.](image)

B. Saccades

Saccadic movements or saccades are called the instantaneous and ballistic changes of the eyes between fixation points [23]. Their amplitude can be very small such as in reading situations or relatively large when, for example, a user is gazing around a room [24]. Saccades, just like fixations, can be discriminated into involuntary or voluntary depending on their duration. Usually, a saccadic movement appears with a frequency of 2 or 3 times every second. The most commonly used saccadic movement features are: number of saccades, saccadic velocity/amplitude, saccade rate and duration [25].

C. Microsaccades

Microsaccades form one of the three fixational eye movements, along with tremor and drift [26]. They serve as a counteraction to retinal adaptation by generating small random displacements of the retinal image in stationary viewing [27]. They move the retinal image in a distance of some hundreds of cones and they have a relatively constant duration of 25 msec and this is why a linear correlation of their peak velocity with their amplitude is being observed [26]. Microsaccades are most possibly conjugate eye movements.

D. Smooth pursuit eye movements

Smooth pursuit eye movements allow the gaze to be maintained on selected objects regardless of whether the subject or the objects are stationary or moving [28]. Smooth pursuit eye movements have a maximum velocity of about 100°/s and frequency of 2 or 3 times every second. The most commonly used saccadic movement features are: number of saccades, saccadic rate and duration [25].

E. Pupil

The size of the pupil is controlled by two sets of muscles, the constrictor and dilator pupillae, which are governed by the sympathetic (SNS) and parasympathetic (PNS) divisions of the autonomic nervous system (ANS) [29]. It reflects autonomic involuntary activity and it is associated with emotional, cognitive or sexual arousal [30]. The pupil size variation is a characteristic measure that is affected and is characteristic in the investigation of mental or cognitive processes [13]. However, pupil metrics are susceptible to issues affecting their reliability that need to be taken into consideration. It is known that pupil’s sensitivity to illumination conditions may affect the pupil size [31], [32] leading to the constriction of the pupil when the amount of light increases. Besides, pupil size is also affected by environmental conditions such as humidity and temperature [33]. Age is another significant parameter for pupil size variation, as it is referred a marked reduction of pupil size with the ageing [34]. Furthermore, pupil metrics can also be affected by the position of the camera and the angle of recording [35]. Therefore, placing the experimental stimuli at the center of the FOV or using well-established reference points or estimating pupil size in the normal state could be a good practice when investigating pupil metrics.

F. Eye blinks

Eye blinking is a semi-involuntary action of fast closing and reopening of the eyelid. It occurs as a result of the co-inhibition of eyelid’s protractors and retractors muscles. Blinking serves the spreading of the corneal tear film across the frontal surface of the cornea [36]. The average duration of a single eye blink is 0.1-0.4 sec [37]. The blink rate (BR), measured in blinks/min, is influenced by environmental factors...
(humidity, temperature, brightness), and physical activity [38]. Also, research evidence suggests that eye blink rate may be tied with emotional and cognitive processing, especially with attentional engagement and mental workload [39].

IV. VIDEO-BASED EYE AND PUPIL TRACKING SYSTEMS AND METHODS

The eye tracking technology is being used in a wide variety of disciplines, including psychology, human-computer interactions as well as in commercial and safety applications [40]. In the respective literature, various eye tracking systems have been proposed. Among them, the most common are electrooculography (EOG), photo/videooculography (POG), scleral contact lens/search coil and video-based. The EOG eye tracker was one of the first developed and rely on signal differences recorded from contact electrodes placed around the eyes. This technique is able to capture eye movements dynamics and reveal visual processing information [41], however in modern approaches video-based techniques appear to gain ground in the research community.

A. Video-based eye tracking

Video-based eye-gaze tracking is at the cutting edge of passive eye tracking techniques. It captures the eye movements non-intrusively utilizing a video camera alone [42]. Video-based eye tracking systems that use pupil and corneal reflection have been developed on the advances on hardware and software. The most common designs of video-based eye trackers use infrared/near-infrared light that creates corneal reflections [43], although there are also some webcam-based eye trackers which are much less accurate [44]. The eye tracker associates the corneal reflection and the centre of the pupil in order to compute vectors that relate eye position to locations in the perceived world. With appropriate calibration, these eye trackers are capable of measuring a viewer’s point of regard on a planar surface on which calibration points are displayed [45]. There is, however, evidence that video-based eye-trackers produce errors in the measurement of small eye movements [46]. The eye tracking setups are available as head-mounted [47], [48], desktop [49], [50] and mobile devices [51], [52]. A typical mobile eye tracker is presented in Fig. 1.

Head-mounted and mobile devices usually include one (monocular) or two (binocular) eye cameras and a scene/world camera. The eye camera monitors the pupil of the eye, while the scene camera captures the user’s field of view.

B. Eye gaze estimation

There are various eye gaze estimation algorithms that utilize Near InfraRed (NIR) illumination of the cornea. Among the most common are the 2D regression, 3D model and cross ratio-based methods. The 2D regression-based methods [53], [54] use the difference between the pupil centre and the corneal glint. Through a mapping function and a transformation matrix derived from a calibration of known gaze points, the gaze coordinates can be extracted. Various studies have evaluated the effect of head movements on system accuracy [55], some of them using neural networks [56], [57]. The implemented 2D regression methods use various combinations of cameras as well as NIR Light-Emitting Diode (LED). The accuracy of these methods varies according to the setup as well as the movement of the head and ranges from 0.8 to 8 degrees. The 3D model-based methods are divided into two subcategories using one [58], [59] or multiple cameras [60], [61]. These methods use a mathematical model of the human eye to reconstruct the centre of the cornea as well as the optical and visual axes. The accuracy of these methods depends on the number of cameras. Higher accuracy can be produced as compared to the corresponding two dimensional one [62] but requires multiple calibration procedures, including calibration of cameras for 3D measurements, estimation of the position with respect to the cameras and the geometry of the LED. There are however some calibration-free gaze estimation techniques which use either multiple cameras [63], [64] or multiple light sources [65]. Finally, regarding the cross-ratio methods the eye gaze points are calculated by projecting a known pattern of NIR light on a screen (four LEDs on four corners of a computer screen) and then by comparing two perspective projections on the camera [66], [67]. The first projection consists of the virtual images of the corneal reflections of the LEDs (scene plane), while the second projection is the camera projection [62].

C. Estimation of pupil area

Pupil area is an important feature of eye behaviour, thus its estimation is of great significance for many studies. Most of the times, the pupil and iris are darker than the surrounding eye area and therefore thresholds can be applied if the contrast is sufficiently large. Researchers have developed a repetitive threshold algorithm based on a skin-colour model where pupils can be identified by the search of dark areas that satisfy certain anthropometric constraints [68]. Unfortunately, the success
rates of the methods drop abruptly by the presence of dark areas around the eye, such as eyebrows or eyelashes [69]. Another disadvantage of these methods is that they can not model the eyes’ closure. In order to overcome this limitation, Tian et al. [70] proposed an eye tracking method that recovers the eye parameters through a dual state model (open/closed eyes). The method requires manual initialization of the eye model and uses a modified Lucas-Kanade tracking algorithm for the tracking of the inner corner of the eye eyelids [71].

Recently, a new method was proposed in [72], which extracts the pupil area through analysis of different eye images intensity levels. In [73], the pupil characteristics are detected using the Cascade Classifier algorithm based on Haar-like features with the help of a histogram equalization method to increase image contrast. Finally, a recent publication presents a method, which initially segments the pupil’s region through a convolutional neural network (CNN), and subsequently finds the center of mass of the region [74].

V. EYE AND PUPIL BEHAVIOUR METRICS

In this section, eye and pupil tracking metrics that are correlated with the 3 states investigated (visual attention, emotional arousal/stress and cognitive workload) along with their usage in relevant studies are presented. The most robust of them in terms of their discrimination ability are identified and discussed.

A. Metrics related to visual attention

When a user inspects an image, a video or a real-world scene, several eye inspection patterns and their computational algorithms can provide insights about the user’s visual attention and scene perception [75]. Some topics such as center bias, saliency and scan paths are key elements to understand eye movement behaviour related to visual attention.

When viewing an image there is a strong tendency of paying attention more at its center and that is translated to increased fixations at this region. This behavior is known as “center bias” and is well documented [76],[77],[78]. Some attributes or regions of the stimuli are more probable to attract the observer’s covert or overt attention, making them salient, such as distinctive color, motion, orientation, or size [79]. Although salience-based schemes are still widely used in computational models, they prove to be poor in predicting eye movement behaviour in natural tasks [80]. Eye movements and especially saccades are being used to investigate several processes, such as visual search. Models of visual search and attention often use scan paths and many studies are trying to quantify scan paths and their nonrandom component [81].

In visual attention studies, fixations form the most used eye movement metric. However, saccades and microsaccades, blinks and pupil size prove also to be helpful. In an attempt to distinguish focal and ambient attention, Krejtz et al. have introduced coefficient K as a metric that takes into account the relationship of fixation duration and its subsequent saccade amplitude. Positive values on the ordinate of coefficient K indicate focal patterns, while negative values suggest ambient visual scanning [82].

A plethora of stimuli and tasks have been used in order to study visual attention. In this section, we will follow a general categorisation of the stimuli (dynamic vs static) and we will present first the studies that used static and then those that used dynamic stimuli.

1) Fixations: In a study which compared visual attention in different kinds of videos and static pictures, fixations when viewing stop-motion movies and Hollywood trailers were longer than when viewing natural movies, and the shortest fixations occurred on static images. An explanation for this is that the abrupt image change in movies captures most of the visual attention of the viewer as compared to a simple static image [78].

In [83], participants had to decide on whether or not a specific person (target) appeared on an image. In doing so, they fixated less in the person presented images than in the person absent ones. Moreover, the duration of exploratory fixations (i.e., the fixations until the person was spotted or not) was larger when the person was in the images. On average, the observers spent 428 ms fixating on the target before responding. In [84], it was found that when observing illustrations, most of the fixations appear in saliency regions. When there was only one saliency object in the presented images, saliency rate of the first fixation (SRF) amounted up to 93.8%. When two or more saliency objects are present in a scene, it was found that the observer could not decide at which to focus first resulting in the SRF decrease. It is also reported that SRF and the saliency rate of the longest fixation (SRL), independent of the number of saliency objects present in a scene, were higher than the saliency rate. This may mean that people pay attention firstly on regions of interest which are likely more salient than other areas.

During image viewing, observers tended to pay more attention to emotionally salienced regions than on visual ones [7]. In [85], an experiment was conducted with random participants and orthodontists staring at images containing smiling faces. It was reported that people with no expertise in dentistry or orthodontic procedures exhibited larger fixation duration in the eyes than the nose or mouth. In contrast, the orthodontist group spent significantly more time looking at the eyes and mouth than the nose of facial images, determining that past experiences, as well as their education and work background, play a vital role in the visual attention motif.

In [86] sensual photographs of heterosexual couples depicting attractive women and men in sensual situations were presented to heterosexual women and men. Participants, irrespective of their gender, looked longer at the body (vs face) of stimuli. This result suggests that, in tasks related to sexual desire, the locus of spontaneous visual attention is preferentially directed toward the body. Moreover, both men and women looked longer at the bodies of women than of men, suggesting that automatic visual attention associated with sexual desire is prominently oriented toward women’s bodies, irrespective of gender.

In a study where images of paintings and faces were presented and participants were asked to grade how much they liked each picture, total fixation duration and number of fixations showed a strong positive correlation with liking ratings,
while no significant difference was found in mean fixation duration. The liking of paintings guided visual attention to the same extent as the liking of faces [87].

In a web task, participants fixated more on the text than on the illustration area. However, using successive web pages with less information seemed to promote more eye fixations on the illustration area [88]. In a similar study, participants made longer fixations during their navigation in the first compared to the second page of news and shopping websites. However, evidence showed that this behaviour was altered while interacting with business websites, in which the duration of fixations remained unchanged irrespective of the number of pages [89]. Furthermore, fixation duration of participants belonging to the millennial generation was shorter than the older participants in a web searching task, meaning that younger people needed less time to process the information of the image [90]. In the same study, when the main picture of the site was placed at the top of the web page, it received a larger number of fixations. Lastly, according to [91] and an experiment asking participants to browse through a website of products, the number of revisits revealed a tendency for the participants regardless of gender to return to a higher user experience more times, but without a main effect of gender, or interaction between gender and user experience.

Memory is associated with visual attention according to [92]. There is a firm correlation between fixation number and recall and this relationship is stronger for those with lower usage levels of a particular commercial brand when watching commercial videos.

Visual attention has also been linked with pain. In [93] participants were presented pictures showing a natural scene while painful electrical stimuli were applied to their left or right hand. Painful stimulation caused fewer and longer fixations. Moreover, painful stimulation on the right hand induced a rightward bias, i.e. increased initial saccades, total number and duration of fixations to the right hemifield of the screen, while pain applied to the left hand as well as no pain induced a leftward bias that was largest for the direction of first saccades.

During a driving task in [94] participants made two consecutive trips during one of which, they received a phone call on a hands-free device in the vehicle. During the phone call, road signs, other vehicles, and the speedometer were fixated less while no significant differences were observed in fixation duration.

There is an extensive study on visual attention on children. Language and visual cue conditions play a vital role in children’s gaze pattern according to [95]. Specifically, both conditions were responsible for the observed increased number of fixations on target landmarks and switches between target objects. Also, it is reported that infants tend to fixate on their mothers before the latter talked to them, they are more likely to look at mothers’ hands when mothers were holding objects, and they fixated at their mothers more quickly when mothers were already present in their field of view [96].

2) Saccades, microsaccades and smooth pursuit: Viewers looking at printed advertisements made longer saccades on the picture part of the ad compared to the text [97]. Painful stimulation while looking at images of natural scenes caused less and slower saccades suggesting reduced exploratory behavior [92].

It has been reported that when viewing natural movies, observers tend to make both more small and more large saccades (with amplitudes of less than 5 and more than 10 degrees, respectively) on natural movies, whereas saccades of intermediate amplitudes are less frequent than in Hollywood action movie trailers and stop-motion movies [78]. In contrast, saccades on Hollywood trailers show the smallest fraction of large amplitudes. During driving task in [94], saccade duration increased during the hands-free trip compared to the control trip, suggesting longer saccade lengths and thus a more disperse fixation pattern during handsfree phoning.

Microsaccades are widely studied in relation with covert attention and there is evidence of the effect of attentional cue presentation on rate and direction of microsaccades [26], [98], [99]. Microsaccades are shown to follow spatial attention during the cue-target interval in a high degree [100]. Mayberg et al. have proposed that microsaccades are related to both – the overt attentional selection of the task-relevant part of the cue stimulus and the subsequent covert attention shift [101]. According to [102] visual attention provides necessary control mechanisms for triggering smooth pursuit eye movements. After its onset, smooth pursuit also requires non-visual cognitive attention controls in order to achieve and maintain a high accuracy of eye tracking. In a clinical study, the relationship between smooth pursuit eye movements and visual attention in patients with schizophrenia and normal controls was evaluated. The magnitude of the correlation between smooth pursuit gain and visual attention measures was statistically compared to the magnitude of the correlation between smooth pursuit gain and motion perception threshold in the controls [103].

3) Blinks: As stated in [39], blink rate can provide useful information about the tendency of the viewer to pay more attention to a specific location in a picture, as the decrease in blink frequency signs the tendency in focusing to this exact location, meaning that the viewer attempts to keep his/her eyes open for a larger time period to observe the desired location.

4) Pupil: It should be noted that attention during work is sometimes modulated by the level of expertise. In [104], the percent of change in the pupil size of experienced workers is slightly lower than the novice workers. Expert’s eye movements may clearly exhibit a systematic pattern. Moreover, regarding the pupil diameter in [91], in parallel with the fixation count and total fixation duration there was also the important influence of user experience, independently of gender, or interaction between gender and user experience.

As shown on Table 1, the metrics that are mostly involved in increased visual attention are the fixation number and duration as well as total fixation time. The reasoning behind this finding lies in the fact that when we focus our attention on a specific object or person, we maintain our gaze more time on a specific AOI, thus more fixations and larger fixation duration on this AOI. Moreover, saccade amplitude, blink rate and microsaccade rate seem to be useful metrics. The rest of the metrics that are presented are not widely studied, thus no specific conclusion can be drawn.
TABLE I
SUMMARY OF EYE-RELATED MEASUREMENTS AND THEIR RELATIONSHIP TO VISUAL ATTENTION

| Eye feature      | Change | References |
|------------------|--------|------------|
| Fixation duration| ↑↑     | [78], [83], [85], [89], [90], [93], [87], [94] |
|                  | ns     |            |
| Number of fixations| ↑↑      | [83], [84], [87], [88], [90], [105], [92], [93], [94], [95] |
| Total fixation time| ↑↑      | [83], [85], [86], [87], [105] |
| Saccade amplitude| ↑↑     | [78], [94] |
| Saccade velocity  | ↑      | [92] |
| Number of saccades| ↑↑      | [92] |
| Microsaccade rate| ↑      | [26], [98], [99] |
| Smooth pursuit gain| ↑      | [102], [103] |
| Blink rate       | ↑      | [32] |
| Pupil size       | ↑      | [91], [104] |

↑↑↓↓: significant increase/decrease during visual attention at 0.01 level
↑↓/↓↑: significant increase/decrease during visual attention at 0.05 level
ns: non-significant difference

B. Metrics related to emotional arousal and stress

Emotional arousal and stress are known to affect eye function and behaviour [106]. In this section, their manifestations in eye metrics (such as fixations, saccades, blinks, pupil, etc) are presented.

1) Fixations: Researchers employed videos conveying both positive and negative emotions and report that the number and duration of fixations were significantly different between the two emotional states studied [107]. Concerning stress conditions, socially anxious participants did not exhibit initial orienting bias and they had a greater probability to fixate on angry faces (having also greater fixation duration) as compared to non-anxious participants [108]. On the other hand, non-anxious participants showed a higher probability of fixation at happy faces, during the two-second time-interval after stimulus onset [108]. Distribution of gaze points is also considered to be affected by arousal and stress [109], [110]. Gaze features such as gaze direction, gaze congruence and the size of the gaze-cuing effect have been employed in arousal studies [111], [112]. Besides, fixation instability has been associated with trait anxiety in both volitional and stimulus-driven conditions, but it is more pronounced in the presence of threat [110]. People with increased anxiety tend to have their first fixation on the emotional picture, contrary to the neutral one [113].

2) Saccades: Arousal has also been associated with saccadic duration, acceleration, and velocity. Saccadic velocity has been considered as an index of arousal/cognitive demand increasing in high arousal states [114]. Involuntary saccades were significantly increased under conditions of arousal with a specific time course for the increase in involuntary movements [115]. Specific inhibitory deficits related to arousal can be revealed through the antisaccade task, where saccadic control is disrupted [116].

3) Blinks: The frequency of spontaneous eye blinks increases and the blinks patterns are modified significantly during stress or other states of emotional arousal [117], [118]. This can be partially attributed to the redirection of blood peri-orbital eye musculature facilitating rapid eye movements [119]. However, eye blinks decrease during tasks that demand to pay more attention (e.g. reading a difficult text) [117]. According to [120], there is a significant correlation between eye blink frequency and stress level. Artificially triggering of emotional responses by billboards, and more natural emotional responses by simulating car crashes caused a temporary increase in the eye blink frequency.

4) Pupil: Pupil size reflects autonomic involuntary activity and it is associated with emotional, cognitive or sexual arousal. Various studies indicate the relation between pupil size variation and affective states [121], [122], [30], [123], [32]. The variation of pupil size has been employed efficiently as an index of stress and arousal [124], [125], [126],[127]. Pupil diameter increases during stress elicited by stressful stimuli in laboratory environment [123], [32], [128], [115].

The pupillary response appears to be greater when the visual stimuli are images conveying negative valence information and higher among persons reporting higher overall levels of stress [129]. Pupil size may also increase in response to positive/negative arousing sounds, as compared to emotionally neutral ones [122], [130]. Audience anxiety has been demonstrated to affect the pupil size [124]. The higher the anxiety level, the bigger the pupil becomes, presenting significant differences in expressions of contemning and surprise [131]. Other studies refer that under arousal conditions the pupil dilation increases [109]. In [127], pupil size was positively correlated with HR or EDA and their causal interaction during emotional processing was investigated. An interesting approach is the translation of pupil behaviour in arousal observed due to drug abuse. The pupil dilation among a cocaine-induced paranoia (CIP) group was significantly greater in response to a video image of crack cocaine than a non-CIP group [132] which can be attributed either to the recall of an event of cocaine intake or to the trait anxiety it is caused by it.

Available research data suggests that emotional arousal is a key element in modulating the pupil’s response. For instance, as early as 1960 authors [133] reported bi-directional effects of emotion on pupil change, specifically reporting that the pupil constricted when people viewed unpleasant pictures and dilated when they viewed pleasant pictures. Similar results were also reported in other studies [134]. In a recent study, high trait anxiety group’s pupil size was increased during tasks with facial expression processing (mainly expressions of condemn and surprise) in relation to low trait anxiety [131].

Table II summarizes the literature findings on the changes in eye and pupil metrics while subjects were under emotional arousal or stress state. As it can be observed, pupil size metric appears as the most common indicator for detecting emotional arousal as its increase clearly reflects the emotional charge in different scenarios. It is important to mention though that the detection or even more the quantification of emotional arousal only through measuring the pupil diameter is questionable. Besides, the limitations referred in Section II should make the researchers cautious on the proper usage of pupil metrics. The majority of the studies stated in Section V-B use other physiological signals outside of pupil size in order to determine among emotional arousal states, thus questioning the feasibility of solely using pupil size as an emotions indicator. Furthermore, except pupil dilation, blink
duration seems to increase with the increase of arousal level. Besides, an interesting finding is the fact that people with trait anxiety tend to have greater duration of the initial fixation and the duration of the fixations were longer on threat images (conveying negative valence) in relation to neutral images [138]. The other eye metrics demonstrated in Table II do not have a consistent correlation on the arousal.

C. Metrics related to cognitive workload

The study of mental workload, also known as cognitive workload, is a vital aspect in the areas of psychology, ergonomics, and human factors for understanding performance. [139]. Despite the multitudinous and extended research in this area, there is no single definition to describe cognitive workload. Often we refer to cognitive workload as taskload i.e. the effort needed to perform a certain procedure. However, defining workload can be a rather subjective task depending on how different people with different experience and abilities can handle the same task [140], [141]. So, a general definition of mental workload would be the product of factors that contribute to one’s workload efficiency for a given task.

Due to the majority of definitions for cognitive workload there is a plethora of ways to measure it. No sensor can handle the same task [140], [141]. So, a general definition of cognitive workload is demonstrated also through video game scenarios. In [146], the saccadic peak velocity decreased while the speed of the game slowed down and increased rapidly when the game speed raised in proportion to the increase of the difficulty level.

Average and maximum saccadic amplitude also show a moderate positive correlation with difficulty level [147]. On the other hand, [153] have shown that mean saccade amplitude decreases with increasing cognitive load, both in experts and novices map users.

During the performance of driving tasks, the saccade velocity and saccade frequency increased when time pressure got higher, and decreased when the subject was overload. The maximum workload was firmly correlated with the maxima of average saccadic velocity and saccade frequency [158]. When performing a secondary task during a driving scenario, a significant increase in drivers’ saccade rate was observed as the task difficulty level increased [159].

It has been shown that mean fixation duration has a significant negative correlation with the level of cognitive load in simulated flight [143], [144] and driving tasks [145], in video gaming [146] and in arithmetic task [147]. Max fixation duration also shows the same behaviour [147]. Fixation duration failed to reach significant difference with increasing cognitive load in a reading task [148]. In [149], fixation duration was found to be most sensitive to one of the three types of cognitive load; extraneous load.

Number of fixations has been shown to increase with increasing cognitive load in a surgical procedure [150], in chess playing [151], in video gaming [146], in simulated flight task [144] and in a complex task on a website [152].

As said before, the relationship between fixation parameters and cognitive workload in association with the level of expertise has been studied largely. Higher fixation duration has been found in novices compared to experts in several working tasks; surgical environment [150] and chess playing [151], while no significant difference was found between experts and novices in video gaming [146], in simulated flight task [144] and in a complex task on a website [152].

As far as the fixation rate is concerned, novices performed more fixations than experts in a surgical environment [150] and chess playing [151], while no significant difference was found between experts and novices in video gaming [146], in simulated flight task [144] and in a complex task on a website [152].

2) Saccades, microsaccades and smooth pursuit: Although saccades form the type of eye movements that are most evaluated in cognitive workload studies, smooth pursuit eye movements and microsaccades are also studied and they seem to correlate well with cognitive load [155], [156].

The average peak saccadic velocity is found to have a positive relationship with the increasing workload [157] and more specifically, it has been shown to be more sensitive to germane load [149]. The significance of saccadic velocity in determining the amount of cognitive load is demonstrated also through video game scenarios. In [146], the saccadic peak velocity decreased while the speed of the game slowed down and increased rapidly when the game speed raised in proportion to the increase of the difficulty level.

Average and maximum saccadic amplitude also show a moderate positive correlation with difficulty level [147]. On the other hand, [153] have shown that mean saccade amplitude decreases with increasing cognitive load, both in experts and novices map users.

| Eye feature                  | Change | References |
|------------------------------|--------|------------|
| Fixation duration           | ↑↑     | [108]      |
|                             | ns     | [107], [113] |
| Number of fixations         | ns     | [107]      |
| First fixation probability  | ↑↑     | [130]      |
| Last fixation probability   | ns     | [113]      |
| Time to first fixation      | ↑      | [107]      |
| Saccadic velocity           | ↑      | [115]      |
| Blink rate                  | ↑↑     | [117], [120] |
|                             | ns     | [109]      |
| Blink duration              | ns     | [109]      |
| Pupil size                  | ↑↑     | [129], [30], [122], [132], [130], [128], [135], [115] |
|                             | ↑      | [136], [131], [124], [137] |
| Gaze distribution           | ns     | [109]      |
| Dwell time                  | ns     | [129]      |

↑↑↓↓: significant increase/decrease of emotional arousal at 0.01 level
↑↓↓↓: significant increase/decrease of emotional arousal at 0.05 level
ns: non-significant difference
amplitude than experts [153]. In a low cognitive load task in which participants had to operate a training version of a military land platform, difference in saccade amplitude between novices and experts failed to reach significance [154].

Kosch et al. have shown that certain trajectories (fast and circular) cause increased gaze differences of smooth pursuit eye movements during the presence of cognitive workload [156]. In another study, eye-target synchronization during smooth pursuit eye movement improved under intermediate cognitive load in young normal subjects [160].

Microsaccades are often used to study cognitive processes. Microsaccade rate in mental arithmetic tasks while fixating a central target is found to decrease with increasing task difficulty [161], [162]. In a more recent study, microsaccade rate failed to reach significant change with increasing cognitive load [155]. Microsaccade amplitude seems to increase with increasing task difficulty [161], [155], while the behaviour of microsaccade peak velocity and amplitude is not clear [155].

4) Pupil: The pupil diameter changes have great subject variability leading to the result that the pupil size changed irregularly during various conditions along with the fact that the value of peak points was different under different difficulty tasks [175]. Another similar study in [169], proposed that pupil diameter distinguished differences in workload between task difficulty levels and increased as the task became more demanding.

Pupil area is strongly associated with the user’s ongoing task difficulty [167], [133], [173] and mean pupil diameter has been shown to have a positive correlation with the cognitive workload in several different tasks [17], [171], [155]. The standard deviation of the pupil size also increases with cognitive load [174]. In a task where the participants had to watch a multimedia lesson, Zu et al. trying to determine what type of cognitive load affects pupil size, concluded that mean ratio of pupil size change was most sensitive to extraneous and germane load [149]. In accordance with this study, another experimental study showed that pupil size increased when time pressure got higher [144]. This study also showed that pupil size decreased when the subject was overload. Pupil size reached to the maximum where the load was maximum.

In addition, pupil diameter increases analogous to the difficulty of the ongoing secondary task performed by the user [164]. During mentally demanding tasks pupil size appears to increase proportionally to the effort that someone has to make [172]. Moreover, pupil size showed the strongest relationship with changes in workload during the Tetris game [146] as it was positively correlated with the mental workload. In the arithmetic task of [147], max pupil dilation correlated strongly with the number of errors in high cognitive load. The increase of the mental workload in [170] caused the pupils to dilate and as the participants were close to overload the saccade rate increased too. Lastly, according to [148], the topic and difficulty level of a text did not significantly influence pupil size measures.

Average pupil size clearly reflected the effort of multitasking while driving as well as it exhibited the importance of habituation effect. After the IVIS (in-vehicle information systems) task was performed in easy and difficult situations, it was observed that average pupil size was significantly higher each time an IVIS condition was performed for the first time [166]. The pupil size of novice surgeons was larger than that of the intermediate surgeons while performing a surgical task as the inexperience led to increased mental and physical effort [150].

An overview of eye features correlated with cognition difficulty based on our literature review is demonstrated in Table III. As it can be observed, unlike to the case of emotional arousal, there is a greater variety of eye behaviour features

| Eye feature | Change | References |
|-------------|--------|------------|
| Fixation duration | ↑↑ [151] | [150] |
| Max fixation duration | ↑↑ [151] | [150] |
| Number of fixations/ Fixation frequency | ↑↑ [151] | [150] |
| Saccadic velocity | ↑↑ [146], [151] | [147], [144] |
| Saccade amplitude | ↑↑ [147] | [151] |
| Max saccade duration | ↑↑ [147] | [151] |
| Saccade rate / Number of saccades | ↑↑ [159], [163], [150] | |
| Microsaccade rate | ↓↓ [161], [162] | [155] |
| Microsaccade amplitude | ↑↑ [161], [155] | [155] |
| Blink rate | ↑↑ [164] | [17] |
| Blink duration | ↓↓ [146], [166] | [166] |
| Pupil size | ↑↑ [146], [167], [168], [169], [170], [17], [171], [155], [144], [166] | [167], [172], [153], [173], [164], [166], [148] |

↑↑: significant increase/decrease during cognitive workload at 0.01 level
↑↓: significant increase/decrease during cognitive workload at 0.05 level
ns: non-significant difference

3) Blinks: Blink duration is a sensitive indicator of cognitive workload, given the significantly higher presence of short blinks under high visual load conditions [166]. Moreover, a decreasing blink duration tendency as the task became more difficult was observed. In [146], blink duration and frequency were also shown to maximize at the lowest speed in the video game, and as the workload increased, blink frequency decreased. In [17] blink rate decreased from low to medium cognitive load but no further change was observed in high cognitive load. Borys et al. also showed that maximum blink duration correlated strongly with the number of errors in high cognitive load of an arithmetic task [147].

In another study, the blink rate was shown to drop to low levels compared to a resting-state rate and be sensitive to the phases of microsurgical suture [165]. In a driving scenario in [164], the eye tracker showed increasing blink rates along with the parallel increase in the difficulty of a -parallel with driving- performed secondary task.
that can witness an individual’s cognitive effort. Firstly, the fixations number increases when the workload is higher while the fixation duration seems to decrease. Saccade rate, saccade velocity and microsaccade amplitude increase with the increasing workload and so is the pupil size which proves to be a useful indicator of cognitive load conditions. Blink rate, microsaccade rate as well as blink duration seem to decrease with cognitive load.

VI. MULTIMODAL EMOTIONAL AND COGNITIVE PROCESSES RECOGNITION METHODS

Emotional arousal and cognitive workload detection are two important aspects of human behaviour. Automatic recognition of these processes through computerized techniques could possibly enhance the computers’ ability to respond and act more intelligently and drive the user away from negative emotional states or possible health issues without demanding high levels of cognitive effort [176],[177], [178][179].

Physiological signals are the most commonly used modality combined with eye-related metrics in order to estimate one’s emotional and cognitive load level [180]. Changes in physiological arousal during stressful conditions are quantifiable through skin conductance, thermal camera recordings, respiration and breath rate, heart activity, electrodermal activity (EDA), electroencephalogram (EEG), electromyogram (EMG), pupil dilation (PD), photoplethysmography (PPG) and body posture/movements [13], [181], [182], [183]. The cognitive workload is associated to a lesser extend with biosignals than in arousal state. Cardiac measurements [184] as well as blood pressure metrics [185] have shown evidence of quantifying the mental workload. These success rates increase by adding to the above metrics various eye movement data with the PD being the most commonly used eye movement metric followed by blink rate and fixation metrics.

In the following subsections we present studies that make use of eye features and biosignals in order to decide upon the levels of emotional arousal and cognitive workload using machine learning techniques.

A. Emotional arousal recognition

In Table IV, studies combining the above characteristics are presented in order to identify and quantify the emotional arousal and stress level. We can separate the studies presented in Table IV into the ones which make use only of eye features and into the ones which utilize various physiological signals for the recognition of the states under investigation. The nature of the tasks performed in the majority of the studies mainly concerns the viewing of emotion evoking images and videos. The researchers in all of the studies but one, make use of pupil size as a feature to their classifiers, thus confirming that pupil diameter is the most important eye metric for emotion recognition with blink related metrics coming at second place. The rest of the eye features such as fixations and saccades metrics as observed are complementary to pupil, blinks and the biosignals feature set in order to ensure the efficient discrimination among the emotional states.

For the studies relying solely on eye features (5 in total) the average best accuracy received is 71.5% with the highest being 93%. In order to interpret those results we must first notice that in [186], [187], the researchers attempt to solve multiclass classification problems (4 emotions and levels of arousal) while using one -but the most important- eye feature, the pupil diameter. However, the highest accuracy received lies below 58% for [186] and 54% for [187] respectively. In [188], a classification between the emotional arousal and valence is attempted while exploiting various eye features correlated with fixations, saccades as well as pupil size resulting to the relatively high accuracy of 80.00% taking into account the three class classification problem. The other two studies concern the binary depressed/non-depressed classification problem. In [189], researchers combine pupil diameter with blink rate and percentage of eyelid closure to reach a 75% accuracy while in [190], they reach the high accuracy of 93% with only exploiting blink related metrics pointing out that it is possible to decide upon the existence of depression with high precision.

For the rest of the studies (9 in total) the average accuracy percentage lies at 81.76% and the highest at 90.1%. This approximately 10% difference in average accuracy in contrast to the studies using only eye-related features is due to the use of biosignals and especially the EEG which is proven to be a very useful emotional indicator [13]. The classification problem concerning these studies is shared among multiclass between basic emotions and binary between stressed and non-stressed classes. As it is expected, the binary classification problem studies resulted in higher accuracy near and above 80.00% reaching up to 90.1%. However, the studies aiming to the quantification of basic emotions resulted in relatively high accuracy too with percentages near 80%.

By observing Table IV once again, the SVM classification system is either the one used in the majority of the studies or the one that helps researchers reach the best emotion prediction accuracy as we can see in the respective column. This is possibly due to the various advantages of the SVM classifiers especially when it comes to relatively small datasets and the clear margin of separation between classes [191].

In this subsection, we commented on the various research attempts to quantify emotional arousal and we compared them with respect to the features they used and the accuracy percentages the achieved. Conclusively, we notice that eye features and especially pupil diameter and blink related metrics play a vital role in the estimation of emotions and emotional arousal. However, although eye features help significantly decide in binary classification problems (i.e. stressed/non-stressed), they cannot by themselves help discriminating well among multiclass classification problems (low/medium/high arousal). Therefore, the combination of eye metrics along with biosignals is necessary in order to achieve higher and reliable prediction rates as we observe from the relevant studies at Table IV.

B. Cognitive workload recognition

In Table IV, we present studies related to the cognitive load levels. The nature of the task varies among the studies,

---

[176],[177], [178][179]...
### TABLE IV

**MACHINE LEARNING TECHNIQUES USED FOR EMOTIONAL AROUSAL AND COGNITIVE WORKLOAD IDENTIFICATION**

| Study | Population (subjects) | Women/Men | Age | Stimuli or Task | Feature | Classes | Classification system | Best accuracy received (best classifier) |
|-------|------------------------|-----------|-----|----------------|---------|---------|----------------------|----------------------------------------|
| Zheng et al. (2020) [186] | 10 | 0/10 | 21-28 | 360° videos presented in VR | PD | angry/happy/ sad/relaxed | SVM | 57.65% |
| Tarnowski et al. (2020) [188] | 30 | 0/30 | 21.25 ± 0.74 | Emotion evoking video fragments | NoF, FD, SA, SD, PD | high/moderate arousal, valence | SVM | 80.00% |
| Ahmad et al. (2020) [17] | 41 | 20/21 | 18-37 | word game | HR, HRV, PD, BR | low/medium/ high workload | RF, NB | 91.66% (RF) |
| Prabhakar et al. (2020) [174] | 21 | 5/16 | mean 26 | simulated and real-time driving | SK, FF, SV, PD | low/high workload | NN | 75.00% |
| Zhao et al. (2019) [192] | 16 | 10/6 | - | Emotion clips | PD, fixation details, saccade details, blink details, EEG | happy/sad/fear/disgust/neutral | SVM | 79.71% |
| Guo et al. (2019) [193] | - | - | - | Images from SEED-V Database | PD, EEG | happy/sad/fear/ disgust/neutral | SVM | 79.63% |
| Li et al. (2019) [194] | 16 | 10/6 | 19-28 | Emotion clips | PD, FD, BD, SD, EEG | happy/sad/fear/ disgust/neutral | SVM | 70.98% |
| Al-Gawwam and Benaissa (2019) [190] | 34 | - | 18-63 | Reading, answering questions | BR, BD, BA | depressed/ non-depressed | SVM, AB, NB, BT | 93.00% (SVM) |
| Wu et al. (2019) [169] | 8 | 4/4 | 26.0 ± 1.6 | simulated surgical exercises | PD, FD, Gaze entropy, PERCLOS | low/high workload | NB | 84.70% |
| Bozkir et al. (2019) [195] | 16 | - | - | simulated driving tasks | PD, driving performance measures | low/high workload | SVM, DT, RF, KNN | 80.70% (SVM) |
| Chen et al. (2019) [196] | 8 | - | 22-25 | simulated flight task | NoF, FD, SD, NoS, saccades, PERCLOS, PD | low/ medium/ high workload | KNN+SVM, SVN | 89.80% (KNN+SVM) |
| Kosch et al. (2018) [156] | 20 | 9/11 | 22-34 | n-back test | SPEM | low/ high workload | SVM | 99.50% |
| Baltaci et al. (2016) [125] | 11 | 2/9 | 33.0 ± 3.46 | IAPS pictures | PD, facial temperature | stressed/ non-stressed | DT, AB, RF | 83.80% (RF) |
| Liu et al. (2016) [197] | 30 | 15/15 | mean 21.4 | judgement trials | PD, FF, BF, Fix, PD, SV, SD, SPV, SD, SPV | low/high workload | SVM | 90.25% |
| Basterrech et al. (2015) [187] | 4 | 0/4 | 19-27 | Image (from IAPS) | PD | low/medium/ high arousal | DT | 53.60% |
| Zheng et al. (2014) [198] | 5 | 2/3 | 22-24 | Emotional film clips | PD, EEG | positive/neutral/ negative valence | SVM | 73.59% |
| Pedrotti et al. (2014) [32] | 33 | 1/16 | 41.0 ± 11.3 | Lane Change Test | PD, EEG, EDA, SKT, Facial cues | stressed/ non-stressed | NN, GA, SVM, ANN (Four-Way parallel class.) | 79.20% |
| Ren et al. (2013) [123] | 30 | 16/14 | 26.8 ± 2.56 | Stroop CWT | PD, GSR | stressed/ non-stressed | NB | 85.53% |
| Algowinem et al. (2013) [189] | 60 | - | - | Emotion evoking images | PD, BR, PERCLOS | depressed/ non-depressed | GMM, SVM | 75.00% (SVM) |
| Zhai et al. (2006) [126] | 32 | - | 21-42 | Paced Stroop Test | PD, GSR, BVP, SSTD | stressed/ non-stressed | SVM | 90.10% |
| Liao et al. (2005) [109] | 5 | - | - | Tasks on computer screen | BR, AECS, Saccades, Gaze Distribution, PD, PRV, Head Movement, Mouth Openness, Eyebrow movement | stressed/ non-stressed | DBN | 86.20% |

PD: Pupil Diameter, PRV: Pupil Ratio Variation, PERCLOS: Percentage of eyelid closure, SPM: Smooth Pursuit Eye Movements, EDA: electrodermal activity, EEG: electroencephalogram, SKT: skin temperature, GSR: Galvanic Skin Response, BVP: Blood Volume Pulse, AECs: Average Eye Closure Speed, PERCLOS: Percentage of eyelid closure, BR: Blink rate, BRV: Blink rate variability, BD: Blink Duration, BA: Blink Amplitude, NoF: Number of fixations, FF: Fixation Frequency, FD: Fixation Duration, SA: Saccade Amplitude, SD: Saccade Duration, NoS: Number of Saccades, SV: Saccade Velocity, SPV: Saccade Peak Velocity, HR: Heart Rate, HRV: Heart Rate Variability, Stroop CWT: Stroop Colour-Word Test, SVM: Support Vector Machines, GA: Genetic Algorithms, RF: Random forests, AB: AdaBoost, GMM: Gaussian Mixture Models, BT: Bagging Trees, DT: Decision Tree.
but we can identify three basic patterns: simulated real-life tasks, n-back test and reading/recalling tasks. Regarding the classification problem, all of the studies aim to solve either a binary class problem (low/high) or a three class problem (low/medium/high) with respect to the level of cognition difficulty. In all of the studies except two, the pupil pupil size is utilized as a feature, confirming that pupil diameter is very vital for the estimation of the cognitive workload with blink related metrics following right after. Aside from smooth pursuit—which is used only in [156], the rest of eye features are complementary to pupil, blinks and the biosignals feature set in order to ensure the efficient discrimination among the cognitive states. For the studies relying solely on eye features (5 in total) the average best accuracy received is 78.64% with the highest being 99.5%. In four of these five studies, the researchers aim to binary classify between low and high cognitive load with high success rates. However, in [196], the scientists aim to identify three levels of cognitive load (low/medium/high) but with low accuracy, that of 43.8%. In contrast and in [17], researchers while trying to identify between the same three categories, exploited heart rate related metrics combined with pupil diameter and blink rate and thus managed to achieve over double the accuracy percentage with 91.66%. The other study [195], uses drive measurement features along with pupil size and reaches up to 80.7% accuracy percentage for the binary classification of the level of cognitive load. So, the average accuracy percentage of the studies utilizing as features for the classification a combination of pupil metrics and biosignals is 87.55%.

By looking once again at Table IV, we can see that -likewise the previous subsection- four out of seven concentrated studies achieve their best accuracy percentages based on SVM classification system most probably for the reasons explained at the emotional arousal recognition subsection.

In this subsection, we summarized the various research attempts to classify the level of cognitive load and we compared them with respect to the features they used and the accuracy percentages the achieved. Finally, we conclude that eye features and especially pupil diameter related metrics are significant in order to reliably measure the cognitive effort especially when seeking to solve binary classification problems and determine the existence of workload. However, biosignals are required to the feature equation in order to decide upon more than two levels of cognitive workload and reach greater success rates.

VII. EYE BEHAVIOUR RELATED DATASETS

In this section, we present and discuss the merits and shortcomings of the various datasets related to emotional arousal, cognitive workload and visual attention that contain eye-tracking features and are publicly available. The relevant data sets identified are listed in Table V, which also presents the number of subjects and their age in each dataset, the stimuli used for invoking the relevant cognitive or affective state and the eye-related features available in each dataset. It should be noted that the various low level eye-related features shown in Table V are those that are provided by the dataset creators. Regarding the stimuli used, we observe that stimuli used for the creation of the datasets focusing on visual attention were video or image presentations or tasks in which focusing on a specific target was required. For the generation of the emotional arousal datasets, the stimuli used were emotional eliciting images and video clips, whilst for the creation of datasets related to the cognitive workload, the subjects were performing everyday activities or other tasks requiring increased mental effort.

As observed in Table V, the basic eye feature included in the vast majority of the datasets is the eyes’ 2D gaze coordinates. These coordinates enable the estimation of fixations and saccades related characteristics using a variety of fixation detection algorithms that are based either on velocity, i.e. Velocity-Threshold identification, Hidden Markov model identification, dispersion, i.e. Dispersion-Threshold identification and Minimum Spanning Trees identification, or the area of interest, i.e. Area of Interest identification [212].

However, as evidenced from the analysis and summary provided in Section V, pupil and blink related features play a vital role in classifying the levels of emotional arousal and in distinguishing the various levels of cognitive load. Unfortunately, only two of the datasets shown in Table V contain information of the pupil diameter, thus making it difficult to accurately estimate the level of emotional arousal or cognitive workload. Finally, only one of the datasets contains blinks related features, which are known to be highly correlated with emotional and cognitive processes.

On the other hand, the Eye Tracking Movies Database (ETMD) [200] provides the research community with a significant number of eye metrics, i.e. 2D gaze points, fixation points, and pupil position and diameter. In addition, as reported in [213], the movie clips used as stimuli were carefully selected in order to evoke different levels of emotional arousal based on the six basic emotions, i.e. happiness, sadness, disgust, fear, surprise, and anger.

Regarding cognitive workload, the EGTEA Gaze+ dataset provides researchers with a relatively large number of subject data and eye features that include 2D gaze points, fixations, saccades, blinks and pupil diameter. However, the EGTEA Gaze+ dataset was originally developed in an effort to study attention and action in first-person vision (FPV), by generating a dataset of meal preparation tasks captured in a naturalistic kitchen environment as reported in [69].

On the other hand the MAMEM Phase I dataset [201] is a dataset that combines multimodal biosignals and eye tracking information gathered under a human-computer interaction framework. The dataset was developed in the vein of the MAMEM project that focused to endow people with motor disabilities with the ability to edit and author multimedia content through mental commands and gaze activity. The dataset includes EEG, eye-tracking, and physiological signals (GSR and Heart rate) collected from 34 individuals, of which 18 were able-bodied and 16 were motor-impaired. The relevant data were collected during the interaction with a specifically designed interface for web browsing and multimedia content manipulation and during imaginary movement tasks. These were tasks that required increased mental effort and often
multitasking abilities. Therefore, despite the drawback that the MAMEM Phase I lacks blinks and pupil related features, it may be more suitable that EGTEA Gaze+ for researchers that want to study the changes in mental workload.

Finally, the datasets available for the study of visual attention are very similar both in terms of subjects present, with the exception of the Variability of eye movement when viewing dynamic natural scenes dataset, the features that they provide, usually 2D gaze points, as well as the nature of the task executed, which is usually video and/or image viewing. Worth noting is the fact that UT Multi-view and Variability of eye movements when viewing dynamic natural scenes datasets contain a relatively larger number of subjects, and that USCCRCNS includes the largest range of eye features, which include 2D gaze points, fixation number, fixation duration, saccade number, and saccade duration.

### VIII. DISCUSSION

In recent years, gaze analysis is seen as an interesting area of research regarding emotion and cognition and for revealing attentional focus and other cognitive strategies of an individual. As a result, the robust and consistent estimation of eye-related metrics through eye tracking and their interpretation for the recognition of emotional or cognitive processes is an important area of current research.

In this paper, a review on the eye features that relate to visual attention, emotional arousal and cognitive workload and their correlation with emotional and cognitive processes was pursued. In the first section, the emotional and cognitive processes are defined. Then, in Section III, the metrics related...
to the eye movements are presented and explained. Fixations, saccadic movements, pupil size, blinks, microsaccades and smooth pursuit eye movements are the most commonly used eye to describe visual attention, emotional arousal and cognitive workload.

Section IV provides a review of the current state of the art regarding eye and pupil tracking systems with special emphasis on video-based eye trackers and methods for the estimation of eye behaviour metrics. An overview of the various types of eye trackers types is presented together with their corresponding advantages and limitations. It is revealed that a popular trend relates to the use of head-mounted trackers (mainly in the form of glasses) using wearable miniaturized IR eye-cameras attached close to the eye. This setup allows to calculate both eye-related movements and pupil size variation using computer vision algorithms. Various algorithms have been developed for the gaze estimation based on 2D/3D CV models. Calibration is a critical procedure for reliable estimation of eye gaze.

An additional finding of our review is the fact that there are specific metrics of eye and pupil behaviour providing valuable information related to emotional/cognitive processes that may significantly contribute to the recognition of these states. Regarding emotional arousal and stress, the pupil size and blink rate appear to be significantly involved in most of the relevant studies. They both increase during states of increased arousal or stress. On the other hand, a clear pattern regarding all other gaze metrics reviewed did not surface, thus remaining a current research goal. Even if literature related to gaze distribution metrics provides evidence and initial research hypotheses, more studies should be conducted to validate these assumptions. Regarding visual attention, it was found that the number and duration of fixations and total fixation time are closely linked to the tendency of focusing on a specific target, whereas other metrics such as saccade amplitude, microsaccade rate and blink rate seem to form useful tools. As for the identification and quantification of cognitive workload, it was found that pupil size, number of fixations and saccadic velocity are reliable positive indicators of increase in mental workload. Blink duration, saccade rate and microsaccade amplitude are also used for cognitive load identification according to several studies.

Section VI presents studies concerning machine learning approaches for the recognition and classification of the various emotional and cognitive processes as well as makes a comparison among them. The majority of the studies exploits the features derived from pupil diameter and blink characteristics, indicating their significance for the process classification. However, both for the emotional arousal and cognitive workload cases, in order to identify among their levels and states (multiclass problem) more reliably, it is vital for many biosignals to be included in the features list. Although most of the algorithms -and especially SVM classification system- can recognise and classify the levels or states of each process with a relatively high accuracy rate, none of the studies reported focus on the discrimination among both the emotional and cognitive processes, thus providing a current research gap in the literature and prospective future research direction. Aside from that, another research matter is that the generalizability of the results is questionable and no relevant information is provided by most reported studies.

We also presented the publicly available datasets that can be used for the estimation of emotional arousal, cognitive workload and visual attention. A comparative assessment of these datasets was performed and suggestions were made regarding their suitability for the identification of emotional and cognitive processes based on the eye-related features. The fact that the available datasets differ not only in the type of affect in which the focus, but also in the number and demographics of participants, the nature of the tasks performed, and the eye-related features included creates non optimal conditions for the computational study of cognitive and affective states based solely on eye-related features.

In addition, it is important to mention that none of the datasets contained data for all - or at least two - of the emotional and cognitive processes investigated in the present manuscript. Therefore, the generation of a dataset that is focused on all of these cognitive and affective states would be extremely beneficiary for the research community.

In conclusion, our review indicates that eye metrics such as fixations, saccades, blinks and pupil size may provide valuable and reliable information for classifying emotional and cognitive processes. For improved classification results especially for classification between the various cognition levels and emotional states more biosignals except the eye features are required to be used. In addition, more research in this topic needs to be done and more datasets must be created targeting exclusively at tasks related to emotional arousal and cognitive workload leading to better discrimination among the various stages and levels.

This review can be used as a comprehensive guideline for researchers who address issues related to human emotions and cognitive processes and their reflection on eye or pupil related metrics.

ACKNOWLEDGMENT

This work was partially supported by the H2020 specific targeted research project SeeFar: Smart glasses for multifaceted visual loss mitigation and chronic disEase prevention indicator for healthier, saFer, and more productive workplAce for ageing population. (H2020-SC1-DTH-2018-1, GA No 826429) [214]. This paper reflects only the author’s view and the Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] E. Granholm and S. R. Steinhauser, “Pupillometric measures of cognitive and emotional processes,” in International Journal of Psychophysiology, vol. 52, no. 1, 2004.
[2] S. Chen and J. Epps, “Automatic classification of eye activity for cognitive load measurement with emotion interference,” Computer Methods and Programs in Biomedicine, vol. 110, no. 2, pp. 111–124, 5 2013.
[3] D. T. Burley, N. S. Gray, and R. J. Snowden, “As Far as the Eye Can See: Relationship between Psychopathic Traits and Pupil Response to Affective Stimuli,” PLOS ONE, vol. 12, no. 1, p. e0167436, 1 2017.
